THE EVOLUTION of LANGUAGE



Editors

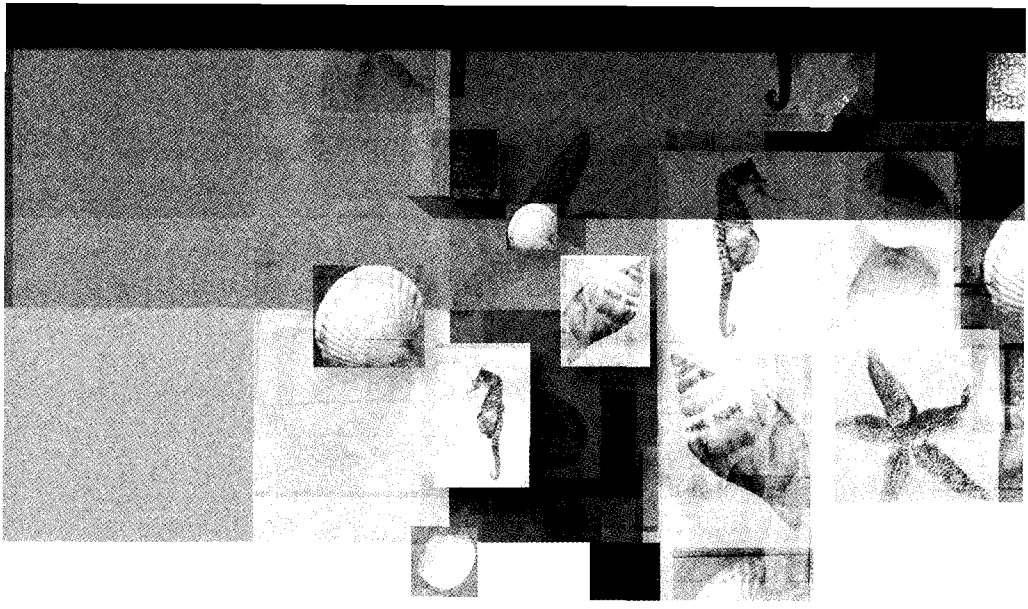
Angelo Cangelosi

Andrew D M Smith

Kenny Smith

THE EVOLUTION of LANGUAGE

This page is intentionally left blank



THE EVOLUTION of LANGUAGE

Proceedings of the
6th International Conference (EVOLANG6)

Rome, Italy 12 – 15 April 2006

Editors

Angelo Cangelosi

University of Plymouth, UK

Andrew D M Smith & Kenny Smith

University of Edinburgh, UK

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

THE EVOLUTION OF LANGUAGE

Proceedings of the 6th International Conference (EVOLANG6)

Copyright © 2006 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-256-656-2

Printed in Singapore by World Scientific Printers (S) Pte Ltd

Preface

This volume collects the refereed papers and abstracts of the 6th International Conference on the Evolution of Language (EVOLANG 6), held in Rome on 12-15 April 2006. Although EVOLANG has been running biennially since 1996, this is the first time that a full published proceedings has been produced. Submissions were solicited in two forms, papers and abstracts, and this is reflected in the structure of this volume.

The EVOLANG conference focuses on the origins and evolution of human language, and brings together researchers from many disciplines including anthropology, archaeology, artificial life, biology, cognitive science, computer science, ethology, genetics, linguistics, neuroscience, palaeontology, primatology, and psychology. The multi-disciplinary nature of the field makes the refereeing process very challenging, and we would like to thank all 85 reviewers for their conscientious and valuable efforts. Special thanks go to Jim Hurford for his encyclopaedic knowledge of researchers in the relevant fields.

Further thanks are due to:

- The EVOLANG committee: Bernard Comrie, Jean-Louis Dessalles, Tecumseh Fitch, Jim Hurford, Chris Knight, Domenico Parisi, Maggie Tallerman and Alison Wray.
- The local organising committee: Davide Marocco, Alberto Acerbi, Massimiliano Caretti, Allegra Cattani, Eduardo Coutinho, Andrea di Ferdinando, Onofrio Gigliotta, Isabella La Rovere, Gianluca Massera, Mariagiovanna Mazzapioda, Orazio Miglino, Stefano Nolfi, Angelo Rega and Vadim Tikhanoff.
- The Fondazione Rosselli, the Italian Association for Cognitive Science and the University of Rome La Sapienza, for their financial support.
- The invited plenary speakers: Tecumseh Fitch, Vittorio Gallese, Steven Mithen, Domenico Parisi, Alberto Piazza & Luigi Cavalli Sforza, Robert Seyfarth & Dorothy Cheney, Luc Steels, Leonard Talmy and Michael Tomasello.
- Finally, and most importantly, the authors of all the contributions collected here.

Angelo Cangelosi, Andrew Smith & Kenny Smith
January 2006

This page is intentionally left blank

Contents

Preface v

Part I: Papers

The Mirror System Hypothesis: From a Macaque-like Mirror System to Imitation 3
Michael A. Arbib, James Bonaiuto & Edina Rosta

Bootstrapping Communication in Language Games: Strategy, Topology and All That 11
Andrea Baronchelli, Vittorio Loreto, Luca Dall'Asta & Alain Barrat

Language Learning, Power Laws, and Sexual Selection 19
Ted Briscoe

The Baldwin Effect Works for Functional, but not Arbitrary, Features of Language 27
Morten H. Christiansen, Florencia Reali & Nick Chater

On the Emergence of Compositionality 35
Joachim de Beule & Benjamin K. Bergen

Towards a Fixed Word Order in a Society of Agents: A Data-driven Baseline Perspective 43
Guy de Pauw

Simulation Model for the Evolution of Language with Spatial Topology 51
Cecilia Di Chio & Paolo Di Chio

Mostly Out of Africa, but What Did the Others Have to Say? 59
Dan Dediu

A Comparison of the Articulatory Parameters Involved in the Production of Sound of Bonobos and Modern Humans 67
Didier Demolin & Véronique Delvaux

Generalised Signalling: A Possible Solution to the Paradox of Language 75
Jean-Louis Dessalles

Innateness and Culture in the Evolution of Language 83
Mike Dowman, Simon Kirby & Thomas L. Griffiths

Early Human Language was Isolating-Monocategorial-Associational 91
David Gil

Computational Simulation on the Coevolution of Compositionality and Regularity 99
Tao Gong, James W. Minett & William S.-Y. Wang

An Epistemological Inquiry into the "What is Language" Question and the "What Did Language Evolve For" Question 107
Nathalie Gontier

Minimalist Foundations of Language Evolution: On the Question of Why Language Is the Way It Is 115
Wolfram Hinzen

Why Has Ambiguous Syntax Emerged? 123
Stefan Hoefler

Proto-propositions 131
James R. Hurford

Convex Meanings and Evolutionary Stability 139
Gerhard Jäger

Natural-language "Cheap Talk" Enables Coordination on A Social-dilemma Game in a Culturally Homogeneous Population 145
Mark Jeffreys

Constraining the Time When Language Evolved 152
Sverker Johansson

Working Backwards from Modern Language to Proto-grammar 160
Sverker Johansson

Language Co-evolved with the Rule of Law 168
Chris Knight

A Saltationist Approach for the Evolution of Human Cognition and Language 176
Susan J. Lanyon

Interaction of Developmental and Evolutionary Processes in the Emergence of Spoken Language 184
John L. Locke

Labels Facilitate Learning of Novel Categories 190
Gary Lupyan

Emergence of Communication in Teams of Embodied and Situated Agents 198
Davide Marocco & Stefano Nolfi

A Language Emergence Model Predicts Word Order Bias 206
James W. Minett, Tao Gong & William S-Y. Wang

Talking to Oneself as a Selective Pressure for the Emergence of Language 214
Marco Mirolli & Domenico Parisi

Learning Models for Language Acquisition 222
Shashi Mittal & Harish Karnick

Simulating the Evolutionary Emergence of Language: A Research Agenda 230
Domenico Parisi

Evolving the Narrow Language Faculty: Was Recursion the Pivotal Step? 239
Anna R. Parker

From Mouth to Hand 247
Dennis Philps

Diffusion of Genes and Languages in Human Evolution 255
Alberto Piazza & Luigi Cavalli Sforza

Differences and Similarities between the Natural Gestural Communication of the Great Apes and Human Children 267
Simone Pika & Katja Liebal

The Evolution of Language as a Precursor to the Evolution of Morality 275
Joseph Poulshock

Modelling the Transition to Learned Communication: An Initial Investigation into the Ecological Conditions Favouring Cultural Transmission 283
Graham Ritchie & Simon Kirby

Towards A Spatial Language for Mobile Robots 291
Ruth Schulz, Paul Stockwell, Mark Wakabayashi & Janet Wiles

Why Talk? Speaking as Selfish Behaviour 299
Thom Scott-Phillips

Semantic Reconstructibility and the Complexification of Language 307
Andrew D. M. Smith

The Protolanguage Debate: Bridging the Gap? 315
Kenny Smith

How to do Experiments in Artificial Language Evolution and Why 323
Luc Steels

The Implications of Bilingualism and Multilingualism for Potential Evolved Language Mechanisms 333
Daniel A. Sternberg & Morten H. Christiansen

Selection Dynamics in Language Form and Language Meaning 341
Monica Tamariz

A Statistical Analysis of Language Evolution 348
Marco Turchi & Nello Cristianini

Evolutionary Games and Semantic Universals 356
Robert van Rooij

Overextensions and the Emergence of Compositionality 364
Paul Vogt

Grammaticalisation and Evolution 372
Henk Zeevat

Stages in the Evolution and Development of Sign Use (SEDSU) 379
Jordan Zlatev & The SEDSU Project

Part II: Abstracts

Alarm Calls and Organised Imperatives in Male Putty-nosed Monkeys 389
Kate Arnold & Klaus Zuberbühler

Perception Acquisition as the Causes for Transition Patterns in Phonological Evolution 391
Ching-Pong Au

The Evolution of Syntactic Capacity from Navigational Ability 393
Mark Bartlett & Dimitar Kazakov

The Subtle Interplay between Language and Category Acquisition and How it Explains the Universality of Colour Categories 395
Tony Belpaeme & Joris Bleys

The Evolution of Meaningful Combinatoriality 397
Jill Bowie

The Adaptive Advantages of Knowledge Transmission 399
Joanna J. Bryson

Determining Signaler Intentions; Use of Multiple Gestures in Captive Bornean Orangutans (*Pongo pygmaeus*) 401
Erica Cartmill & Richard Byrne

Nuclear Schizophrenic Symptoms as the Key to the Origins of Language 403
Timothy J. Crow

Articulator Constraints and the Descended Larynx 405
Bart de Boer

Evolutionary Support for a Procedural Semantics for Generalised Quantifiers 407
Samson Tikitu de Jager

The Evolution of Spoken Language: A Comparative Approach 409
W. Tecumseh Fitch

Allee Effect on Language Evolution 411
José F. Fontanari & Leonid I. Perlovksy

Rapidity of Fading and the Emergence of Duality of Patterning 413
Bruno Galantucci, Theo Rhodes & Christian Kroos

Reconsidering Kirby's Compositionality Model Towards Modelling Grammaticalisation 415
Takashi Hashimoto & Masaya Nakatsuka

The Interrelated Evolutions of Colour Vision, Colour and Colour Terms 417
David J. C. Hawkey

A Little Bit More, A Lot Better: Language Emergence from Quantitative to Qualitative Change 419
Jinyun Ke, Christophe Coupé & Tao Gong

Major Transitions in the Evolution of Language 421
Simon Kirby

Modelling Unidirectionality in Semantic Change 423
Frank Landsbergen

The Origin of Music and Its Linguistic Significance for Modern Humans 425
Steven Mithen

Co-evolution of Language and Behaviour in Autonomous Robots 428
Sara Mitri & Paul Vogt

Iconic versus Arbitrary Mappings and the Cultural Transmission of Language 430
Padraic Monaghan & Morten H. Christiansen

Mother Tongue: Concominant Replacement of Language and MtDNA in South Caspian Populations of Iran 432
Ivan Nasidze & Mark Stoneking

What can Grammaticalization Tell Us about the Origins of Language? 434
Frederick J. Newmeyer

Bootstrapping Shared Combinatorial Speech Codes from Basic Imitation: The Role of Self-organization 436
Pierre-Yves Oudeyer

How Language Can Guide Intelligence 438
Leonid I. Perlovsky & José F. Fontanari

The Roles of Segmentation Ability in Language Evolution 440
Kazutoshi Sasahara, Bjorn Merker & Kazuo Okanoya

Primate Social Cognition and the Cognitive Precursors of Language 442
Robert Seyfarth & Dorothy Cheney

Agonistic Screams in Wild Chimpanzees: Candidates for Functionally Referential Signals 443
Katie Slocombe & Klaus Zuberbühler

An Individual-based Mechanism for Adaptive Semantic Change 445
Daniel W. Smith

A Holistic Protolanguage Cannot Be Stored, Cannot Be Retrieved 447
Maggie Tallerman

Recombinance in the Evolution of Language 449
Leonard Talmy

Ape Gestures and Human Language 452
Michael Tomasello

Prehistoric Handedness: Some Hard Evidence 453
Natalie Uomini

Lateralization of Intentional Gestures in Non Human Primates: Baboons Communicate with Their Right Hand 455
Jacques Vauclair & Adrien Meguerditchian

Emergence of Grammar as Revealed by Visual Imprinting in Newly-hatched Chicks 457
Elisabetta Versace, Lucia Regolin & Giorgio Vallortigara

Beyond the Argument from Design 459
Willem Zuidema and Timothy O'Donnell

Author Index 461

Papers

This page is intentionally left blank

THE MIRROR SYSTEM HYPOTHESIS: FROM A MACAQUE-LIKE MIRROR SYSTEM TO IMITATION

MICHAEL A. ARBIB,[2,3,4] JAMES BONAIUTO[3,4] & EDINA ROSTA[1]

[1]*Chemistry,* [2]*Computer Science,* [3]*Neuroscience and* [4]*USC Brain Project*

University of Southern California, Los Angeles, CA 90089-2520

The Mirror System Hypothesis (MSH) of the evolution of brain mechanisms supporting language distinguishes a monkey-like mirror neuron system from a chimpanzee-like mirror system that supports simple imitation and a human-like mirror system that supports complex imitation and language. This paper briefly reviews the seven evolutionary stages posited by MSH and then focuses on the early stages which precede but are claimed to ground language. It introduces MNS2, a new model of action recognition learning by mirror neurons of the macaque brain to address data on audio-visual mirror neurons. In addition, the paper offers an explicit hypothesis on how to embed a macaque-like mirror system in a larger human-like circuit which has the capacity for imitation by both direct and indirect routes. Implications for the study of speech are briefly noted.

1. The Mirror System Hypothesis

Both premotor area F5 and parietal area PF of the macaque monkey brain contain *mirror neurons* each of which fires vigorously both when the monkey executes a certain limited set of actions and when the monkey observes some other perform a similar action. Imaging data show that the human brain contains *mirror regions* in both frontal and parietal lobes, namely regions that show high activation both when a human performs a manual action and when the human observes a manual action, but not when the human simply observes an object. It is widely assumed that such mirror regions contain mirror neurons, based on similarities between the human and macaque brain.

The *Mirror System Hypothesis* (MSH; Rizzolatti and Arbib, 1998) asserts that the *parity requirement* for language in humans – that what counts for the speaker must count approximately the same for the hearer – is met because Broca's area (often associated with speech production) evolved atop the mirror system for grasping with its capacity to generate and recognize a set of actions. However (Hurford, 2004), one must distinguish the mirror system for the signifier (phonological form) from the neural schema for the signified, and note the need for linkage of the two. On this view, Broca's area becomes the meeting place for phonological perception and production, but other areas are required to link phonological form to semantic form.

The crucial point is that humans have capacities denied to monkeys. Mirror regions in a human can be activated when the subject imitates an action, or even just imagines it, but there is a consensus that monkeys cannot imitate save in the most rudimentary sense. By contrast, chimpanzees exhibit "simple imitation", the ability to approximate an action after observing and attempting its repetition many times; while humans alone among the primates have the capacity for "complex imitation", being able to recognize another's performance as a combination of more-or-less familiar actions and to use this recognition to approximate the action, with increasing practice yielding increasing skill. Thus research on MSH requires not only a fuller understanding of the mirror system of the macaque, but also an understanding of how the mirror system and the circuitry with which it interacts must have changed in the course of evolution.

Arbib (2002, 2005a) modified and developed MSH by hypothesizing seven stages in the evolution of language. The first three stages are pre-hominid:

S1: Grasping.

S2: A mirror system for grasping, shared with the common ancestor of human and monkey.

S3: A system for simple imitation of grasping shared with the common ancestor of human and chimpanzee.

The next 3 stages distinguish the hominid line from that of the great apes:

S4: A complex imitation system for grasping.

S5: *Protosign*, a manual-based communication system that involves the breakthrough from employing manual actions for praxis to using them for pantomime (not just of manual actions), and then going beyond pantomime to add conventionalized gestures that can disambiguate pantomimes.

S6: *Protospeech*, resulting from linking the mechanisms for mediating the semantics of protosign to a vocal apparatus of increasing flexibility.

Arbib (2005b) argues that protosign and protospeech evolved together in an expanding spiral. The final stage is then:

S7: *Language*: the change from action-object frames to verb-argument structures to syntax and semantics.

Arbib (2005) provides arguments and counter-arguments for these various claims. The present article focuses on the earlier, rather than the later, stages in this progression. It contributes to this argument by (a) introducing a new model of action recognition learning by macaque mirror neurons which addresses data on auditory input; (b) outlining how to embed a macaque-like mirror system in a larger human-like circuit which has direct and indirect paths for "complex imitation"; and (c) noting implications for the study of speech.

2. MNS2: Recognizing Audible Actions

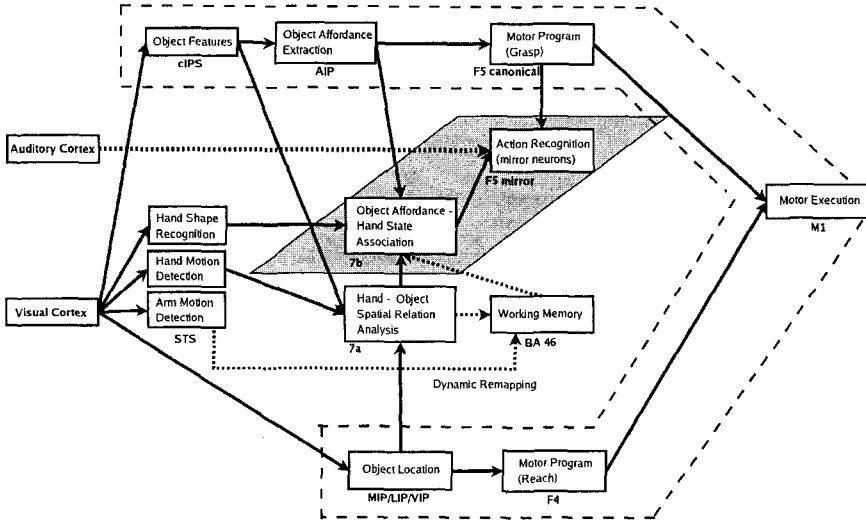


Figure 1: System diagram for the MNS2 model, updating the MNS model of Oztop & Arbib, 2002. The dashed outline shows the system for generating the reach to and grasp of an observed object. The remaining circuitry defines the mirror system and the subsystems which feed it. The encoding of the grasp motor program (F5 canonical) provides the training signal for a recurrent network which models the areas 7b and F5 mirror, shown here in the gray parallelogram, by the activity of its hidden and external output layers, respectively. The dotted arrows denote the connections unique to the MNS2 model. Auditory information about actions reaches the F5 mirror neurons via the auditory cortex. Visual data on hand-object spatial relations is input into the Object Affordance-Hand State Association schema and into working memory. When this information is not available externally, the dynamically remapped working memory trace serves in its place.

The MNS model (Oztop & Arbib, 2002) of the monkey mirror system was designed to associate activity in canonical neurons providing a premotor encoding of the type of a grasp with visual input encoding the trajectory of a hand relative to an observed object. The learning mechanism was a feed-forward backpropagation network of units with one hidden layer which required an unnatural recoding of its input. Bonaiuto et al. (2005) developed a model, MNS2, that could process the time series of hand-object relationships without such recoding, using an adaptive recurrent network to learn to classify grasps based on the temporal sequence of hand-object relations.

Umiltá et al. (2001) have shown that mirror neurons in the macaque monkey can recognize a grasp if the monkey has seen the target object which was then hidden, but cannot recognize the action lacking current or recent input on the affordances and location of the object. MNS2 incorporates working memory and dynamic remapping components (Figure 1) which allow the model to recognize

grasps even when the final stage of object contact is hidden and must be inferred. Before being hidden, the object position and its affordance information are stored in working memory. Once the hand is no longer visible, the working memory of wrist position is updated using the still-visible forearm position. If the model observes an object which is then hidden by a screen, and then observes a grasp that disappears behind that screen, the wrist trajectory will be extrapolated and the grasp will be classified accordingly.

However, the more important contribution of MNS2 within the context of MSH is that it addresses data on "audiovisual" mirror neurons which associate sounds with manual actions. Köhler et al. (2002 –see Figure 2 right) found that some of the mirror neurons in area F5 of the macaque premotor cortex responsive to the sight of actions associated with characteristic noises (such as peanut breaking) are just as responsive for the sounds of these actions.

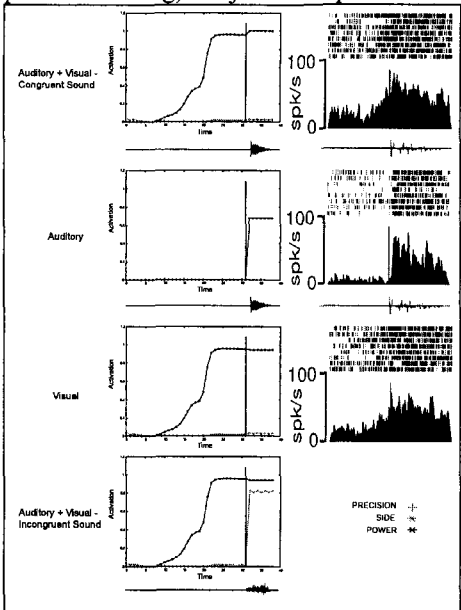


Figure 2: Left: Activation of the model's external output layer when presented with a precision grasp sequence containing (from top to bottom) visual and congruent audio, visual only, audio only, and visual and incongruent audio information. The black vertical lines indicate the time step at which the hand made contact with the object. The unit encoding the precision grasp shows the greatest level of activation, while the unit corresponding to power grasps shows a small level of transient activity. At the bottom is an oscillogram of the sound associated with the precision grasp. Right: Activation from Köhler et al. (2002) of an audiovisual mirror neuron responding to (from top to bottom) the visual and audio components, visual component alone, and audio component alone of a peanut-breaking action.

Bonaiuto et al. (2005) associate each sound with a distinct pattern of activity which is applied to audio input units which are fully connected to the output layer of the recurrent neural network, corresponding to a direct connection from auditory cortex to F5. These connection weights are modified using Hebbian learning. In this way, any sound that is consistently perceived during multiple occurrences of an executed action becomes associated with that action and incorporated into its representation. This type of audio information is inherently

actor-invariant and this allows the monkey to recognize that another individual is performing that action when the associated sound is heard.

3. A Dual Route Model of Imitation Gated by Attention

It is often suggested that mirror neurons are the substrate for imitation, matching observed actions onto motor programs producing similar or equivalent actions. However, as we saw earlier, only humans have "complex imitation", the ability to imitate sequences of behaviors and approximate novel actions as variants of known actions after one or just a few viewings of this novel behavior. As backdrop for our own work, we draw some important lessons from apraxia.

DeRenzi (1989) reports that some apraxics exhibit a *semantic deficit* – having difficulty both in classifying gestures and in performing familiar gestures on command – yet may be able copy the pattern of a movement of such a gesture without "getting the meaning" of the action of which it is part. We call this residual ability *low-level imitation* to distinguish it from imitation based on recognition and "replay" of a goal-directed action. With Rothi, Ochipa, and Heilman (1991), we thus propose a dual route imitation learning model to serve as a platform for studying apraxia. The *direct route* for imitation of meaningless and intransitive gestures converts a visual representation of limb motion into a set of intermediate limb postures or motions for subsequent execution. The *indirect route* for imitation of known transitive gestures recognizes and then reconstructs known object-directed actions. The distinction between the direct and indirect routes in praxis may be related to the well-known distinction between the dorsal and ventral streams in vision (Ungerleider & Mishkin, 1982) which also plays a crucial role in our model of the visual control of hand movements (Fagg & Arbib, 1998) and may in turn have implications for the study of language. We suggest that the interaction of these two routes underlies the human capacity for complex imitation. We hypothesize that, during sequential or complex actions, contributions from each route are encoded in a competitive queuing mechanism (Rhodes et al., 2004). The focus of attention (whether directed toward the object and limb, limb posture, or movement) determines the relative competitive weight of the movement segment encoded by each route. A modification to the competitive choice layer implements a sort of selective, n-winners-take-all mechanism, allowing non-interfering movement segments with similar weights to be executed simultaneously. In this way novel movements can be recognized as combining known actions (indirect route) with intransitive limb adjustments (direct route).

4. Complex/Goal-Directed Imitation

We have argued that humans have "complex imitation", the capacity for recognizing novel actions as combinations of (variants of) known object-directed actions, with joint adjustments to meld them together. These novel actions can then be acquired as skills through successive approximation. In addition, humans have the ability to imitate complex "meaningless" movements which are not directed towards objects – as we saw in defining the "direct route".

In their theory of goal-directed imitation, Wohlschläger et al. (2003) present the hypothesis that imitation is the result of the decomposition of the aspects of a movement and the hierarchical structuring of these goal aspects. Each of these goal aspects triggers the associated motor program for reproducing that aspect of the movement. Wohlschläger et al. (2003) attribute differences in imitative abilities across species to differences in working memory capacity. However, this is not evident from the current data, and differences in imitative ability could very well be due to differences in the mechanism(s) of hierarchical movement aspect decomposition. The fact that humans can imitate intransitive movements does not seem to be due to an increased working memory capacity, but rather the ability to decompose aspects of intransitive movements such as relative limb postures and via points. Through this process of successive approximation, complex movements can be reproduced with increasing accuracy by increased attention being paid to its subparts. This increased attention may result in a finer-scaled decomposition of the observed movement, resulting in execution of a more congruent movement.

5. Discussion

The audio properties of mirror neurons are of major interest because they may have been crucial in the transition from gesture to vocal articulation in the evolution of language. These multi-modal mirror neurons may have allowed arbitrary vocalizations to become associated with communicative gestures, facilitating the emergence of a speech-based language from a system of manual gestures. If this is indeed the case, the development of audio-visual mirror neurons may have implications for the recognition of communicative actions and ground the multi-modality of language (Fogassi & Ferrari, 2004; Arbib, 2005b).

The possible relation of the direct and indirect routes in praxis to the dorsal and ventral streams in vision may in turn have implications for the study of language. Hickok & Poeppel (2004) observe that early cortical stages of speech perception involve auditory fields in the superior temporal gyrus bilaterally

(although asymmetrically) but offer evidence that this cortical processing system then diverges into two streams:

A *dorsal stream* maps sound onto articulatory-based representations which projects dorso-posteriorly. It involves a region in the posterior Sylvian fissure at the parietal–temporal boundary, and ultimately projects to frontal regions. This network provides a mechanism for the development and maintenance of "parity" between auditory and motor representations of speech; and

A *ventral stream* maps sound onto meaning which projects ventro-laterally toward inferior posterior temporal cortex (posterior middle temporal gyrus) which serves as an interface between sound-based representations of speech in the superior temporal gyrus (again bilaterally) and widely distributed conceptual representations.

The distinction between the direct and indirect routes in praxis may also be relevant to the distinction made by Levelt (e.g., Levelt et al., 1999) between overt and internal speech. Using our normal perceptual system, we can monitor our own vocal output and discover errors, dysfluencies, or other problems of delivery in our own overt speech. However, Levelt further claims that we can monitor some internal representation – Wheeldon and Levelt (1995) offer evidence that this takes the form of a somewhat abstract phonological representation – as it is produced during speech encoding and use this internal self-monitoring ability to trace the process of phonological encoding itself. As noted by one of the reviewers, a fruitful topic for future research is to pursue the development of this dual-feedback architecture on an evolutionary scale as part of the task of elaborating the Mirror System Hypothesis.

References

Arbib, M.A. (2005a). From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics, *Behavioral and Brain Sciences, 28,* 105-167. (Supplemental commentaries and the author's "electronic response" are at *Behavioral and Brain Sciences,* http://www.bbsonline.org/Preprints/Arbib-05012002/Supplemental/Arbib.E-Response_Supplemental.pdf.)

Arbib, M.A. (2005b). Interweaving Protosign and Protospeech: Further Developments Beyond the Mirror, *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems 6,* 145–171.

Bonaiuto, B., Rosta, E., and Arbib, M.A. (2005). Recognizing Invisible Actions, *Workshop on Modeling Natural Action Selection*, Edinburgh, July, 2005. (An expanded version has been submitted for publication under the title "Extending the Mirror Neuron System Model, I: Audible Actions and Invisible Grasps".)

DeRenzi, E. (1989). Apraxia. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology*. Amsterdam: Elsevier. Vol. 2., pp. 245–263.

Fagg, A.H., Arbib, M.A. (1998). Modeling Parietal-Premotor Interactions in Primate Control of Grasping, *Neural Networks 11*, 1277-1303.

Fogassi, L., Ferrari, P.F. (2004). Mirror neurons, gestures and language evolution, Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems *5*,345-363.

Hickok, G., and Poeppel, D., 2004, Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language, *Cognition*, *92*, 67-99.

Hurford, J.R. (2004). Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution. In D.K. Oller and U. Griebel (Eds.), *Evolution of Communication Systems: A Comparative Approach*. Cambridge, MA: The MIT Press, pp. 297-313.

Köhler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V., Rizzolatti, G. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science*, 297, 846-848.

Levelt, W.J.M., Roelofs, A., Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-38.

Oztop, E., Arbib, M.A. (2002). Schema design and implementation of the grasp-related mirror neuron system. Biological Cybernetics, *87*, 116-140.

Rizzolatti, G., Arbib, M.A. (1998). Language Within Our Grasp. *Trends in Neuroscience*, *21*, 188-194.

Rothi, L.J.G., Ochipa, C., and Heilman, K.M. (1991). A cognitive neuropsychological model of limb praxis. *Cogn. Neuropsychol. 8*, 443-458.

Umiltá, M.A., Köhler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., Rizzolatti, G. (2001). I know what you are doing: a neurophysiological study. *Neuron*, *31*, 155-65.

Ungerleider, L.G., Mishkin, M. (1982) Two cortical visual systems. In D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield (Eds.), *Analysis of Visual Behavior*. Cambridge, MA: The MIT Press, pp.549-586.

Wheeldon, L.R., Levelt, W.J.M. (1995) Monitoring the time course of phonological encoding. *Journal of Memory and Language 34*, 311–34.

Wohlschläger, A., Gattis, M., Bekkering, H. (2003). Action generation and action perception in imitation: an instance of the ideomotor principle. *Phil. Trans. R. Soc. Lond.*, *358*, 501-515.

BOOTSTRAPPING COMMUNICATION IN LANGUAGE GAMES. STRATEGY, TOPOLOGY AND ALL THAT

ANDREA BARONCHELLI, VITTORIO LORETO

Dipartimento di Fisica and SMC center, Università "la Sapienza", P.le Aldo Moro 2, Roma, 00187, Italy

andrea.baronchelli@roma1.infn.it, vittorio.loreto@roma1.infn.it

LUCA DALL'ASTA, ALAIN BARRAT

Laboratoire de Physique Théorique (UMR du CNRS 8627), Université de Paris-Sud, Bâtiment 210, Orsay Cedex, 91405, France

luca.dallasta@th.u-psud.fr, Alain.Barrat@th.u-psud.fr

Semiotic dynamics is a fast growing field according to which language can be seen as an evolving and self-organizing system. In this paper we present a simple multi-agent framework able to account for the emergence of shared conventions in a population. Agents perform pairwise games and final consensus is reached without any outside control nor any global knowledge of the system. In particular we discuss how embedding the population in a non trivial interaction topology affects the behavior of the system and forces to carefully consider agents selection strategies. These results cast an interesting framework to address and study more complex issues in semiotic dynamics.

1. The Naming Game

In recent times, the view of language as a complex dynamical system that evolves and self-organizes has gained ground in the scientific community (Steels, 2000). In this new perspective, complex systems science turns out to be a natural allied in the quest for the general mechanisms underlying the emergence of a shared set of conventions in a population of individuals.

The issue is of the outmost topicality since, for the first time, the web allows for the spreading and the study of global bottom up created semiotic systems. Recently, for instance, new web tools (such as del.icio.us or www.flickr.com) enable users to self organize systems of tags and in that way build up and maintain social networks and share information. On the other hand, many technological systems are nowadays composed of single communicating entities. The capability of developing ontologies or proto languages without any intervention from the outside would be of great importance for instance in those cases in which teams of ar-

tificial embodied agents should explore highly unknown environments, such as distant planets or deep seas.

A possible approach to the understanding of language self-organization is that of modeling artificial population of agents and studying their evolution. The choice is then between endowing agents with simple properties, so that one can hope to fully understand what happens in simulations, or with more complicated and realistic structures that yet risk to confuse experiments outputs. We choose to follow the first possibility since we are more interested in the global behavior of the population. In this perspective we do not seek answers to specific issues in the evolution of language, but rather we aim at analyzing deeply basic models that can constitute valuable starting points for more sophisticated investigations. Nevertheless, as we shall see, also extremely transparent agents and interaction rules can give rise to very complex and rich global behaviors and the study of simple models can help to shed light on general properties - a well known lesson in statistical physics.

We discuss here a recently introduced Naming Name model (Baronchelli, Felici, Caglioti, Loreto, & Steels, 2005), inspired to the one proposed by Steels (1995), in which agents play pairwise interactions in order to negotiate conventions, i.e. associations between forms and meanings. The population reaches a final convergence state without any external or global control. This is a central point, since, of course, no such control has been present in the development of natural language, and, as mentioned above, its absence is becoming a desirable feature also for many technological systems. Also, it is worth noting that this model accounts for the emergence of a shared set of conventions (a vocabulary, in our case) from the point of view of cultural transmission (Hutchins & Hazlehurst, 1995; Steels, 1995), without resorting to any evolutionary issue (Hurford, 1989; Nowak, Plotkin, & Krakauer, 1999).

The game is played by a population of N agents. Each agent is characterized by its *inventory*, i.e. a list of form-meaning associations that evolve dynamically during the process. For the sake of simplicity we do not take into account the possibility of homonymy, so that all meanings are independent and we can work with only one of them, without loss of generality. Agents aim to converge to a unique shared form (or word) to associate with the meaning (or object). Agents have empty inventories at time $t = 0$ and at each time step ($t = 1, 2, ..$) two players are picked at random to play an interaction: one of them plays as *speaker* and the other as *hearer*. Their interaction obeys the following rules:

- The speaker randomly extracts a word from its inventory, or, if its inventory is empty, invents a new word.

- If the hearer has the word selected by the speaker in its inventory, the interaction is a success and both players maintain in their inventories only the winning word, deleting all the others.

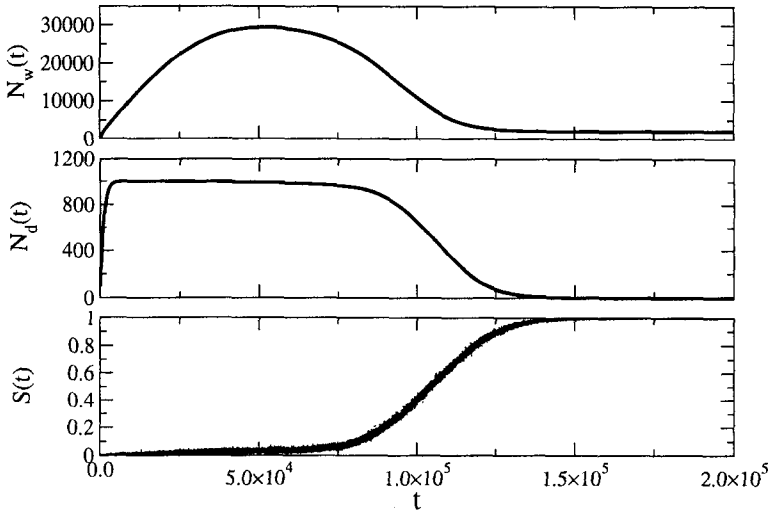


Figure 1. Time evolution of the most relevant global properties of the Naming Game. From up to down: the total number of words, $N_w(t)$, the number of different words known by the agents, $N_d(t)$, and the probability of a successful interaction at a give time, $S(t)$. Convergence is reached with a quite abrupt disorder/order transition that starts approximately just after the peak of the $N_w(t)$ curve has disappeared. Data are relative to a population of $N = 2000$ agents and averaged over 300 simulation runs.

- If the hearer does not have the word selected by the speaker in its inventory, the interaction is a failure and the hearer updates its inventory adding the new word.

The most relevant quantities to describe the evolution of the population are: the total number of words stored by the system at each time step, $N_w(t)$, the number of different words known by the agents, $N_d(t)$, and the probability of a successful interaction at a given time, $S(t)$. In Figure 1 we report the time evolution of a population of $N = 2000$ agents. It is immediately clear that the population reaches a final coherent state in which there is only one word ($N_d = 1$) and all interactions are successful (Baronchelli et al., 2005). As we mentioned above this is a remarkable fact, considering the simplicity of the rules that govern our process, and makes it worthy a more detailed analysis of the model.

The process starts with a trivial phase in which agents invent new words. It follows a longer period of time where the $N/2$ (on average) different words are exchanged after unsuccessful interactions. The probability of a success taking place at this time is indeed very small since each agent knows only few different words. As a consequence, the total number of words grows while the number of different words remains constant. However, agents keep correlating their inventories so that at a certain point the probability of a successful interaction ceases to

be negligible. As fruitful interactions become more frequent the total number of words at first reduces its growth and then start decreasing. Moreover, after a while some words start disappearing from the system. The new virtuous correlations among inventories make the process evolve with an abrupt increase in the number of successes and a further reduction in the numbers of both total and different words. Finally, the dynamics ends when all agents have the same unique word and the system is in the attractive convergence state. It is worth noting that the developed communication system is not only effective (each agent understands all the others), but also efficient (no memory is wasted in the final state).

2. Interplay with topology

In the Naming Game model described above at each time step two agents are *randomly* chosen, thus implying that we deal with a completely unstructured population (i.e. we are in the mean-field case). However, the hypothesis that each agent can in principle talk to anybody else is strongly unrealistic when we deal with large numbers. The remedy is to embed agents in a quenched spatial structure, typically a regular lattice. More realistic alternatives to regular structures are given by complex networks. A network is, roughly speaking, an ensemble of nodes connected by links (or edges). Examples of such structures are common, Internet and the World Wide Web being the most obvious. Moreover, recently, it has been found out that many more systems can be described as networks (Albert & Barabási, 2002; Pastor-Satorras & Vespignani, 2004). As examples we can mention social networks in which people are the nodes and their social relations are the links (Wasserman & Faust, 1994), scientific collaboration networks, where two scientist are connected by a link if they have co-authored at least an article (Newman, 2004), metabolic networks in which nodes are the substrates and edges are chemical reactions in which the substrates participates (Jeong, Tombor, Albert, Oltvai, & Barabási, 2000) and food webs in which the nodes are species and the links represent predator-prey relationships (Garlaschelli, Caldarelli, & Pietronero, 2003). Among the most peculiar features shared by most natural or artificial networks there are the "small world" property (Watts & Strogatz, 1998) and the scale free degree distribution (Barabási & Albert, 1999). The first is the name attributed to the evidence that the minimal hop distance between each pair of nodes scales logarithmically with the network's size instead of algebraically as in usual regular lattices. The second is the fact that, said degree k of a node the number of links which connect it to other nodes, the degree distribution $P(k)$ follows a power law $P(k) \sim k^{-\gamma}$, thus allowing for the presence of very few nodes with very high connectivity that in general play a central role in the structural and dynamical properties of the system.

In our simulations we have chosen to place agents on the nodes of Barabasi-Albert network (Barabási & Albert, 1999). This is an artificial network which displays a power law degree distribution $P(k) \sim k^{-3}$. It is built starting from

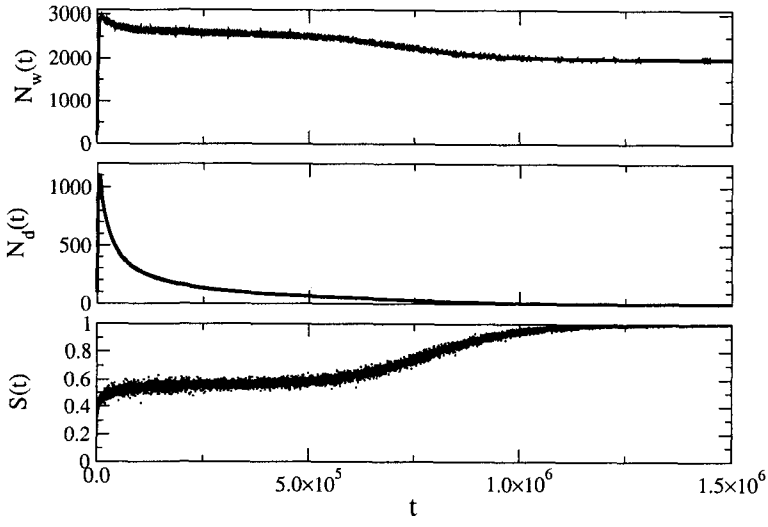


Figure 2. Time evolution of the Naming Game played by a population of $N = 2000$ agents embedded in a Barabasi-Albert network with $m = 2$. Convergence is considerably slower and the maximum number of words is smaller than in the mean field case.

a core of fully connected nodes and adding sequentially new nodes with m links each. The existing nodes to be linked with the newcomer are chosen with a probability proportional to their degree (the well-known "preferential attachment" rule), so that new links are likely to be added to well connected nodes (Note that for each node, the initial degree is m, but it can subsequently grow when newcomers attach to the node; the average degree in the network is $\langle k \rangle = 2m$). In Figure 2 we report data relative to a Naming Game played by a population of 2000 agents placed on the nodes of a Barabasi-Network graph with $m = 2$. The radical difference with Figure 1 is manifest. When the network is present, the convergence time is extremely longer (than for the mean-field case), while the maximum number of total words is smaller. Moreover the curves of both total and different words are qualitatively modified, since in the latter the plateau region disappears, while a flat region is now present in the first. This can be understood observing the success rate curve, that at first grows very rapidly, but then remains stuck in a very long plateau. The first growth is due to the consensus reached locally among small clusters of agents, while global ordering takes a much longer time and is responsible for the quasi constant success rate. However, to gain a clearer picture of what goes on, it is crucial describing how the interacting agents selection takes place (Castellano, 2005). In the simulations we have just discussed, the first extracted agent played as speaker. This is a relevant information since, due to the scale-free degree distribution, a randomly selected node is, with high probability,

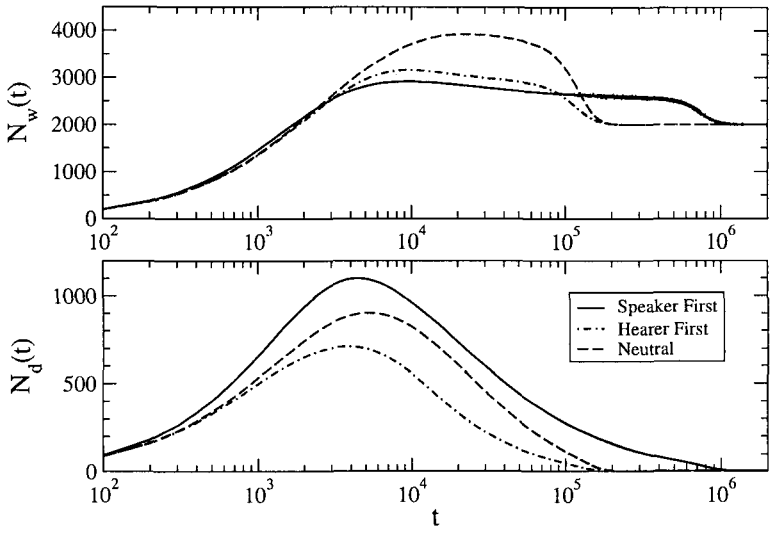


Figure 3. Interacting pairs selection strategies; in "Speaker First" the speaker is at first selected and one of its neighbors plays as hearer, while the opposite happens in the "Hearer First" approach. These strategies lead to very different outcomes since, due to the scale free nature of the network, the first selected node is usually low connected, while the second has great chances to be a hub. The neutral method, finally, consists in selecting a link and tossing a coin to attribute roles to its extremes; it gives rise to hybrid behavior. Curves are the results of 300 simulation runs performed for a population of $N = 2000$ agents embedded in a Barabasi-Albert graph with $m = 2$. Note logarithmic scale on the abscissa.

one with low degree. Such nodes form, in fact, a vast majority. The hearer, then, being selected among the neighbors of the speaker, is likely to be a high degree node, i.e. to be directly reachable from many nodes. So, low degree nodes invent a large number of different words at the beginning and pass them to the hubs, which tend to store on average a larger amount of words. Highly connected nodes are thus obviously the necessary go-between of successes, but their passive role slows down the dynamics. On the other hand, their low number allows the total number of words to stay low.

Given that the rules chosen to assemble interacting pairs are likely to be determinant, we have performed experiments following two different gathering strategies. The first consists in picking up at first the hearer and then one of its neighbors as the speaker. The last one is a "neutral" strategy in which a link is randomly selected and then a coin toss assigns the speaker/hearer roles. Results relative to numbers of words are shown in Figure 3. The curve relative to the "hearer first" strategy confirms our picture. Noticeably, here the convergence is much faster than when the speaker is selected at first. The reason is due to the active role played by hubs which, being most frequently speakers, tend to propagate success-

ful words to low connected nodes. Their active role in the initial invention process also keeps the number of different words low. The other side of the coin is the larger number of words the population has to store during the process, due to the fact that low connected agents need to store more words than in the "speaker first" case. Finally, neutral strategy determines an hybrid behavior between those of the "speaker/hearer first", the only relevant feature being a peak of the $N_w(t)$ curve higher than those recorded for the other strategies.

Before concluding two remarks are in order. First of all, it must be noted that pairs selection strategies are so crucial due to the non trivial topology on which the games are performed. Both for complete graphs and regular lattices, in fact, the three strategies mentioned above are completely equivalent and give rise to identical results. Secondly, we stress that, while in our case we have embedded the population on a static interaction pattern, it would also be desirable to consider a dynamically evolving topology. In this respect, an interesting study of co-evolution of language and social structure has been done by Gong, Ke, Minett and Wang (2004), for a more complex language game.

3. Conclusions

In this paper we have presented a very simple model able to account for the emergence of shared conventions in a population of agents. In our case individuals agree on the name to assign to an object, but the model can be straightforwardly enriched to study the genesis of more complex language structures. In this perspective, the study of the Naming Game model provides a fundamental first step towards more realistic modeling.

After studying the most important global properties of the mean-field case, we have performed simulations in which the population is put on a non-trivial quenched interaction topology. In particular, we have shown that the topology strongly affects the way in which final convergence is reached. Moreover, due to the heterogeneity of the underlying network, pair selection strategies become crucial in determining the behavior of the system. This is due to the role of highly connected nodes, the hubs, which considerably speed up the convergence when are able to distribute conventions actively and slow it down if used as passive agents connectors.

These findings offer useful hints to understand real systems. For different reasons, as a matter of fact, in social networks not all the individuals play the same role. Indeed, the structure of such networks is scale free, but the ways in which a node can become a hub can be very different. In web based communities, for instance, a node can attract connections providing passively useful or interesting material to the community, or on the contrary can increase its social influence being very active in establishing relations with other users. In this perspective, our work contributes to shed light on how such different human-based mechanisms underlying the emergence of social interaction structures can heavily affect

the semiotic dynamics happening upon them. Finally, an important consequence concerning the *role* of the hubs is that, when we are interested in the semiotic processes taking place on a non trivial social structure, the mere knowledge of its topology might be insufficient to make predictions.

References

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics, 74*, 47.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509.

Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., & Steels, L. (2005). Sharp transition towards shared vocabularies in multi-agent systems. *arxiv:physics/0509075, submitted for publication.*

Castellano, C. (2005). Effect of network topology on the ordering dynamics of voter models. *AIP Conference Proceeding, 779*, 114.

Garlaschelli, D., Caldarelli, G., & Pietronero, L. (2003). Universal scaling relations in food webs. *Nature, 423*, 165.

Gong, T., Ke, J., Minett, J. W., & Wang, W. S.-Y. (2004). A computational framework to simulate the co-evolution of language and social structure. In *Alife 9.* Boston, MA, U.S.A.

Hurford, J. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua, 77*, 187.

Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: the development of shared symbols in interaction. In G. N. Gilbert & R. Conte (Eds.), *Artificial societies: The computer simulation of social life.* London: UCL Press.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. *Nature, 407*, 651–654.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proc. Natl. Acad. Sci. USA*, 5200-5.

Nowak, M. A., Plotkin, J. B., & Krakauer, J. D. (1999). The evolutionary language game. *Journal Theoretical Biology, 200*, 147.

Pastor-Satorras, R., & Vespignani, A. (2004). *Statistical mechanics of complex networks.* Cambridge: Cambridge University Press.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life, 2*(3), 319–332.

Steels, L. (2000). Language as a complex adaptive system. In M. Schoenauer (Ed.), *Proceedings of ppsn vi.* Berlin, Germany: Springer-Verlag.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* New York and Cambridge, ENG: Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small world' networks. *Nature, 393*, 440.

LANGUAGE LEARNING, POWER LAWS, AND SEXUAL SELECTION

TED BRISCOE

Computer Laboratory
University of Cambridge
JJ Thomson Ave
Cambridge CB3 OFD, UK
ejb@cl.cam.ac.uk

I discuss the ubiquity of power law distributions in language organisation (and elsewhere), and argue against Miller's (2000) argument that large vocabulary size is a consequence of sexual selection. Instead I argue that power law distributions are evidence that languages are best modelled as dynamical systems but raise some issues for models of iterated language learning.

1. Introduction

A diagnostic of a power law distribution is that a log-log plot of frequency against rank yields a (nearly) straight line. For instance, Zipf (1935) plotted word token counts in a variety of texts against the inverse rank of each distinct word type and showed that typically such plots approximate a straight line. The characteristic 'Zipf curve' of word frequency against rank deviates from this line because the relative frequency of very common word types, such as the English determiners *the* and *a*, tend to be more similar than the power law predicts, as also does the relative frequency of very rare words in the tail of the distribution. Zipf's 'law' is often expressed as:

$$c(w) \propto \frac{1}{r(w)^B} \tag{1}$$

where $B > 1$, the exponent, defines the slope of the plot, frequency $c(w)$ is the token count of word type w in text, and rank $r(w)$ is the position of word type w in the list of word types sorted in descending order of frequency, $c(w)$. Guiraud's (1954) related law states that the number of word types V in a text is proportional to the length of that text N:

$$V \propto N^A \tag{2}$$

Although the models of power law distributions of which I am aware have a dynamical component, they have received little attention from evolutionary linguists.

I know of only one argument, due to an evolutionary psychologist (Miller, 2000), which utilises Zipf's observation about word frequencies in attempting to explain large, redundant vocabularies in terms of sexual selection. I argue against this explanation in §4, but, before doing so, I discuss the ubiquity of power law distributions in §2, some relevant models of them in §3, return to Miller's argument in §4, and then discuss some issues power law distributions raise for evolutionary models of iterated language learning in §5.

2. Manifestations of Power Law Distributions

Power law distributions are very different from normally distributed phenomena, such as height, which yield the characteristic 'Bell Curve'. The factors that influence a person's height, such as nutrition and genetic inheritance, combine in a more linear manner so that (relatively minor) variation in height is normally distributed around a mean that can be accurately estimated from a representative sample of the population. Zipf (1949) noted that Pareto's observations about the distribution of wealth in the population could also be modelled using a version of his 'law'. What makes wealth different from height intuitively is that the factors that influence the amount of money we have combine non-linearly and there are strong (positive) feedback effects (i.e. 'the rich get richer'). We now know that power law distributions are good approximations of many other non-linguistic phenomena, such as the distribution of people within cities, citations amongst scientists, accesses of web pages, species within habitats, authors amongst scientific articles, actors within films, links between web pages, activation of genes, size of earthquakes, number of sexual partners, and many more (e.g. Albert & Barabasi,2002).

There are similar results regarding extrinsic properties of languages: for instance, the distribution of languages within language families approximates a power law (Wichmann, 2005). In terms of inherent properties of language, Zipf also showed that plotting the length or the number of meanings of word types against their frequency also yields similar distributions. With the increasing availability of annotated electronic corpora, these observations have been extended to many other areas of language organisation, such as the frequency of contiguous sequences of words (bigrams, trigrams, and more generally ngrams), of grammatical rules, of construction types, of lexical relations between word types, as well as the length of constituents, and the association of verbs with constructions (e.g. Sharman, 1989; Manning & Schutze, 1999; Korhonen 2002; Yook *et al.*, 2001).

3. Models of Power Law Distributions

So far I have used the term 'distribution' ambiguously between the linguistic and probabilistic sense. The most important insight about such distributions with large numbers of rare events (e.g. Baayen, 2001) is that converting a frequency-rank plot into a probability-rank plot via maximum likelihood (i.e. relative frequency)

estimation, and treating the result as a probability distribution is unwise. Since the counts of the tail are very low, statistical estimation theory tells us that they will be unreliable. A rare word, for instance, may suddenly become fashionable (e.g. *eggregious*, *serendipity*) and thus increase in relative frequency over a given time period. Since, we always see a long tail of rare events no matter how much (more) text we sample, and the number of types grows in proportion to the size of this sample (Guiraud's law), power law distributions are often described as 'scale-free'. In statistical terms, power law distributions which remain invariant over different sample sizes are a strong indication that we may be sampling from a statistically unrepresentive non-stationary (i.e. dynamical) system.

Baayen (1991), following in the tradition of Mandelbrot (1953) and Simon (1955), develops a stochastic Markovian model of phonotactically legal Dutch word strings and relates it to empirical data on similarities between words by phonological form and by relative frequency. He finds that to model these effects accurately, it is necessary to add a second 'dynamical' stochastic model which introduces or removes word types with probability proportional to their token frequency. This has the effect of increasing overall frequency-based and decreasing form-based similarity. For present purposes, it is indicative that the second dynamical word 'birth-death' processs is required even though it says nothing directly about the relationships between word types.

Albert & Barabasi (2002) provide a recent survey of work on 'small world' networks in which most nodes of a network can be reached by any other in a small number of (node) steps, though the overall number of nodes can be arbitrarily high. They define a dynamical algorithm for generating such networks, by continuously adding new nodes and attaching them to old nodes with probability proportional to their number of existing links. They prove that such networks evolve to a scale-free organisation obeying a power law distribution in which there is a long tail of nodes with low numbers of links and a small number of 'popular' nodes with many links. They also prove that both 'growth', the dynamical component, and 'preferential attachment' are necessary for this pattern to emerge. Such networks have been applied to models like that of Baayen (1991), described above (e.g. Bornholdt and Ebel, 2001), and to lexical semantic organisation (e.g. Yook *et al.*, 2001).

4. Power Laws and Sexual Selection

Miller (2000:369f), in the context of a more general argument that human language evolved by sexual selection, argues that large vocabulary size, in comparison with those of other (artificial and natural) animal communication systems, evolved through sexual selection. Women preferred men with large active vocabularies but needed to acquire large passive vocabularies themselves to assess the trait. Miller offers, as evidence for the non-functional nature of much of this vocabulary, Zipf's observation that vocabulary distributes like a power law and

contains many near synonyms:

> ...any of the words we know is likely to be used on average about once in every million words we speak... Why do we bother to learn so many rare words that have practically the same meanings as common words, if language evolved to be practical? (Miller, 2000:370)

He argues that human variation in vocabulary acquisition correlates with intelligence and has a heritable component, and thus is an (indirect) fitness indicator, triggering an 'arms race' in which advertising excessive vocabulary size is a 'display' of fitness akin to the peacock's tail, precisely because it does not contribute usefully to communication.

In §2 we saw that power law distributions manifest themselves in many areas of linguistic organisation. For instance, there is a tail of rare long constituents in text samples (Sharman, 1989). However, there is no evidence that 'display' of such forms is a particular feature of courtship, nor that such forms are non-functional. As we saw in §3 models predicting such distributions need only a dynamical component and no element of natural or sexual selection whatever. Evidence of power law distributions in both idiolects and language forces us to conclude that both are best modelled as dynamical systems – rather than well-formed sets, as in generative linguistics (e.g. Sampson, 2001:165f) – but nothing more.

If vocabulary size were non-functional, we might expect there to be many truly synonymous words. What we find in the organisation of vocabulary is that partially synonymous words have different distributions in terms of specificity of reference, syntactic potential, or genre and register. There is, in fact, considerable evidence that children avoid hypothesising synonyms in language acquisition (e.g. Clark, 2003) and that language users adhere to the convention of preemption by synonymy, except where discourse or syntactic context triggers a non-synonymous reading (e.g. Briscoe *et al.*, 1995; Copestake & Briscoe, 1995). For instance, *cow*, unlike *chicken*, is not generally used to refer to the meat because of the existence of *beef*. However, in an appropriate context *cow* can be used this way and triggers an implicature of 'disgust':

> There were five thousand extremely loud people on the floor eager to tear into roast cow with both hands and wash it down with bourbon whiskey. (Tom Wolfe, 1979. *The Right Stuff*, Farrar, Straus and Giroux, New York (p. 298, Picador edition, 1991))

Similarly, the word *stealer*, formed by the fairly productive derivational rule of agentive *+er* nominalisation, is blocked by *thief*, except in syntactic contexts where the specificity of reference is narrowed:

> He is an inveterate *stealer / thief / stealer of Porsche 911s

These and many similar observations suggest that partial synonymy is communicatively useful and actively exploited to convey meaning.

To understand why we have so many words and how the cognitive ability to cope with them (co-)evolved, consider the likely environment of adaptation for language. In a foraging, scavenging or hunter society, the ability to discriminate – and thus name more and more species, according to nutritional value, location, method of capture or harvesting, and so forth – would be of value for survival because it would allow efficent transmission of these skills to kin as well as survival over larger and more varied habitats. Modern hunter-gatherers are known to have large vocabularies specialised in this way (Diamond, 1997). This may not have been the sole driver for increasing vocabulary size, but it has the advantage that it predicts that vocabulary will be to a large extent organised by specificity of reference. It is useful not only to be able to talk about plants in general but also species and subgroups (e.g. by location or edible part) in order to discriminate the edible, find the source, and harvest effectively. Once we accept such a pressure to name in an increasingly complex and multifaceted environment, then the tendency for there to be smaller numbers of high frequency words of generic reference and a larger number of rarer words with highly specific denotations is just a case of the structure of vocabulary mirroring (our perception of) this environment.

5. The Real Challenge – Iterated Learning

One achievement of recent evolutionary models of language is the demonstration that treating languages as complex adaptive systems responding to conflicting selection pressures (e.g. Briscoe, 2000) leads to insightful acccounts of typological and other linguistic universals without the need to invoke innateness. These accounts rely heavily on the iterated learning model (ILM, e.g. Kirby, 2001) in which linguistic traits must undergo repeated relearning by successive generations of language learners acquiring their language from that of the previous generation. For instance, Kirby (2001) demonstrates that languages in the ILM evolve to have compositional structure in which only high frequency irregular form-meaning mappings are stable, given the following assumptions:

1. an *invention strategy* for form-meaning pairs,
2. a *production bias* to express meanings using short forms,
3. an *inductive bias* to learn small grammars and lexicons,
4. a *learning period* in which not all form-meaning pairs appear
5. and *environmental structure* which favours some meanings

In the simulation, initial (proto)languages are holistic and non-compositional but chance regularities which emerge in form-meaning mappings are acquired by

learners, who then reliably exemplify them for the next generation of learners, because regularities are, by definition, more frequent in data. Thus, over time the language evolves to be mostly compositional and regular. However, (short) irregular mappings can survive provided they are associated with meanings which are expressed frequently and, therefore, also occur reliably during the learning period.

This instantiation of the ILM neatly explains the observation that irregularity correlates with high frequency in attested languages: children would continue to say *goed* into adulthood if *went* were not a high frequency form. The corollary, however, is that rare unpredictable properties of language which do not follow from some regularity manifest during the learning period should be unstable and, therefore, rarely observed.

Rare word-meaning associations are unpredictable and may also influence lexico-grammatical behaviour. For example, the verb *obsess* is a stable lexeme of English, but does not appear in any of the 40 or so case studies of child-directed speech in CHILDES[a]. It is transitive but usually appears in the passive in adult speech accompanied by a PP headed by *by*, *with* or *over*. However, vocabulary acquisition continues through adulthood, so the ILM (and other models) simply predict that such vocabulary will be acquired later (and less universally).

Marked but predictable constructions, such as multiple centre-embeddings, which Sampson (2001:21) estimates occur once in every 250K words on average, are also not counter-examples if one believes that they are a consequence of learners acquiring, in the basis of more frequent constructions, grammatical rules which correctly predict the appropriate form-meaning mapping for these constructions.

A more challenging case for the ILM is diathesis alternation, in which verbs of certain semantic classes semi-predictably occur in alternant constructions often with predictable meaning changes. For instance, *eat* can appear in intransitive and transitive constructions but when it occurs intransitively the theme of the action is 'understood'. However, verbs with similar senses, such as *devour* or *consume* do not undergo this alternation. There is evidence that children learn at least some of these alternation rules by 3 years because they produce errors, such as *Don't fall my dolly down* – the causative-inchoative alternation. However, the rate at which such errors occur also suggest that alternation rules are learnt conservatively and only rarely overapplied. There are on the order of 100 such alternation rule types in English, when productive meaning change is taken into account.

Figure 1 shows log-log plots of the unconditional probability of over 150 verb-headed constructions against their inverse rank on the left and of the conditional probability of these constructions when headed by any form of the verb *believe* on the right, calculated from 30M words of automatically parsed text along with the closest fit straight line derived using (1) above with B set appropriately. Both distributions loosely approximate power laws with long tails of rare events, but

[a]http://childes.psy.cmu.edu/data/

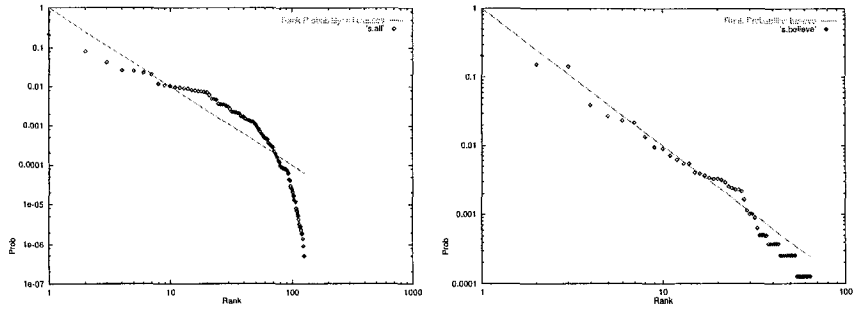


Figure 1. AllVerb/Believe Constructions

critically the correlation between the unconditional distribution and conditional ones for individual verbs is low (0.47 Spearman-Rank Coefficient for 14 verbs). This means that it is not possible to predict the individual association of verbs and constructions on the basis of the unconditional distribution. For instance, sentential complements are rare overall but the most common construction with *believe*, accounting for over 90% of occurrences. This lack of correlation, taken together with the fact that analysis of CHILDES shows that child-directed speech only exemplifies common verb-construction associations (e.g. Buttery & Korhonen, 2005), suggests that children do not have reliable evidence for the existence of most alternation *rules* – assuming that evidence would be several exemplars of the same alternation involving several different verbs.

It may be that such semi-productive alternations are also acquired later in life, despite the occasional errors in children's speech. This is a general strategy that proponents of ILM-style explanations can take. But on the other hand, there must also be some learning 'bottleneck', caused by limited exposure to data during the learning period, for ILM accounts of linguistic evolution to work. Cases like this pose interesting challenges for the approach because they suggest that linguistic data is distributed in such a fashion that there may still be a 'poverty of stimulus' issue during the sensitive period for acquisition. More empirical work on language acquisition is needed to determine whether the ILM's predictions hold up for such specific cases.

References

Albert, R. & A. Barabasi (2002) 'Statistical machanics of complex networks', *Reviews of Modern Physics, vol.74,* 47–97.

Baayen, H. (1991) 'A stochastic process for word frequency distributions', *Proceedings of the Assoc. for Computational Linguistics,* Morgan Kaufmann, Menlo Park, CA, pp. 271–278.

Baayen, H. (2001) *Word Frequency Distributions,* Kluwer, Dordrecht.

Bornholdt, S. & Ebel,H. (2001) 'World Wide Web scaling exponent from Simon's 1955 model', *Physical Review, vol.64,* 035104.

Briscoe, E.J. (2000) 'Evolutionary perspectives on diachronic syntax' in (eds) Pintzuk, S., Tsoulas, G. and Warner, A. (eds.), *Diachronic Syntax: Models and Mechanisms,* Oxford University Press, Oxford, pp. 75–108.

Copestake, A.A. and E.J. Briscoe (1995) 'Regular polysemy and semi-productive sense extension', *Journal of Semantics, vol.12,* 15–67.

Briscoe, E.J., A.A Copestake and A. Lascarides (1995) 'Blocking' in St. Dizier, P. and Viegas, E. (eds.), *Computational Lexical Semantics,* Cambridge University Press, Cambridge, pp. 273–302.

Buttery, P. & A. Korhonen (2005) 'Large-scale analysis of verb subcategorization differences between child directed speech and adult speech', *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes,* Saarland University.

Clark, E. (2003) *First Language Acquisition,* Cambridge University Press, Cambridge.

Diamond, J. (1997) *Guns, Germs and Steel: The Fate of Human Societies,* Random House, New York.

Guiraud, H. (1954) *Les Charact'eres Statistiques du Vocabulaire,* Press Universitaires de France, Paris.

Kirby, S. (2001) 'Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity', *IEEE Transactions on Evolutionary Computation, vol.5(2),* 102–110.

Korhonen, A. (Computer Laboratory, University of Cambridge) *Subcategorization Acquisition,* Techical Report UCAM-CL-TR-530. 2002

Mandelbrot, B. (1953) 'An informational theory of the statistical structure of language' in W. Jackson (eds.), *Communication Theory,* Butterworths, London.

Manning, C. & H. Schutze (1999) *Foundations of Statistical Natural Language Processing,* MIT Press, Cambridge MA.

Miller, G. (2000) *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature,* William Heinemann, London.

Sampson, G. (2001) *Empirical Linguistics,* Continuum, London.

Sharman, R. (1989) *Observational Evidence for a Statistical Model of Language,* IBM UKSC Report 205.

Simon, H. (1955) 'On a class of skew distribution functions', *Biometrika, vol.42,* 435–440.

Wichmann, S. (2005) 'On the power law distribution of language family sizes', *Journal of Linguistics, vol.41,* 117–131.

Yook S., Jeong, H., Barabasi, A-L & Tu, Y. (2001) 'Weighted evolving networks', *Physical Review Letters, vol.86,* 5835–5838.

Zipf, G. (1935) *The Psycho-Biology of Language: An Introduction to Dynamic Philology,* Houghton-Miflin, New York.

Zipf, G. (1949) *Human Behavior and the Principle of Least Effort,* Addison-Wesley, Cambridge, MA.

THE BALDWIN EFFECT WORKS FOR FUNCTIONAL, BUT NOT ARBITRARY, FEATURES OF LANGUAGE

MORTEN H. CHRISTIANSEN & FLORENCIA REALI
Department of Psychology, Cornell University, Uris Hall, Ithaca, NY 14853, USA

NICK CHATER
Department of Psychology, University College London, 26 Bedford Way London, WC1H 0AP, UK

Human languages are characterized by a number of universal patterns of structure and use. Theories differ on whether such linguistic universals are best understood as arbitrary features of an innate language acquisition device or functional features deriving from cognitive and communicative constraints. From the viewpoint of language evolution, it is important to explain how such features may have originated. We use computational simulations to investigate the circumstances under which universal linguistic constraints might get genetically fixed in a population of language learning agents. Specifically, we focus on the Baldwin effect as an evolutionary mechanism by which previously learned linguistic features might become innate through natural selection across many generations of language learners. The results indicate that under assumptions of linguistic change, only functional, but not arbitrary, features of language can become genetically fixed.

1. Introduction

Although the world's languages differ considerably from one another, they nonetheless share many systematic constraints on how they are structured and used. Explaining how such universal linguistic constraints evolved in the hominid lineage is the focus of much debate in language evolution research. One view suggests that linguistic universals are best viewed as *arbitrary* features of language with no functional explanation, but instead deriving from an innate Universal Grammar (UG; Chomsky, 1965). This abstract body of linguistic knowledge is proposed, by some theorists, to have evolved gradually through biological adaptations complex grammars (e.g., Briscoe, 2003; Pinker & Bloom, 1990). An alternative view seeks to explain linguistic universals as *functional* features of language, emerging due to communicative and cognitive factors outside of grammatical knowledge (e.g., Bybee, 1998). These features are seen as by-products of linguistic adaptation, in which language itself has been adapted through cultural transmission across many generations of language learners (e.g., Tomasello, 2003).

The Baldwin effect (1896) is the primary evolutionary mechanism by which the arbitrary features of UG are envisioned to have been genetically fixed in the human population. Although a Darwinian mechanism, the Baldwin effect resembles Lamarckian inheritance of acquired characteristics in that traits that are learned or developed over the life span of an individual become gradually encoded in the genome over many generations (see Weber & Depew, 2003). That is, if a trait increases fitness, then individuals that, due to random genetic variation, require less exposure to the environment to develop that trait will have a selective advantage. Over generations, the amount of environmental exposure needed to develop this trait decreases, as individuals evolve increasingly better initial conditions for its rapid development. Eventually, no environmental exposure may be needed; the trait has become genetically encoded. A frequently cited example of the Baldwin effect (e.g., Briscoe, 2003) is the ability to develop hard skin on certain areas of the body with relatively little environmental exposure. Over time, natural selection would have favored individuals that could develop hard skin more rapidly (because it aids in mobility, prevents infection, etc.) until it became fixed in the genome, requiring little environmental stimulation to develop. Similarly, it has been suggested that arbitrary linguistic features, which would originally have had to be learned, gradually became genetically fixed in UG via the Baldwin Effect (Pinker & Bloom, 1990).

In this paper, we use computer simulations[a] to investigate the circumstances under which the Baldwin effect may operate, for arbitrary and functional features of language. Building on previous work (Chater, Christiansen & Reali, 2004), Simulation 1 indicates that arbitrary linguistic features *cannot* be genetically fixed via the Baldwin effect when linguistic change is incorporated — even when this change is driven in part by the genes themselves. In Simulation 2, we show how functional features of language can come to be genetically fixed in the population when they promote better communicative abilities. Finally, we discuss the implications of the simulations for theories of language evolution.

2. Simulation 1: Arbitrary Language Features

Following recent work on the possible evolution of UG (e.g., Briscoe, 2003; Nowak, Komarova & Nyogi, 2001), we model language and learners as a set of binary vectors. Specifically, we adopt the framework of the pioneering

[a] All simulations were replicated several times due to their stochastic nature.

simulations of Hinton & Nowlan (1987), used by Pinker & Bloom (1990) to support their suggestion that the Baldwin effect underlies the gradual genetic fixing of arbitrary grammatical features in UG. Our previous work indicated that although the Baldwin effect can occur within this framework in the context of arbitrary linguistic features, the effect disappears when language is allowed to change (Chater et al., 2004). However, these simulations were limited in scope; we therefore conducted a new series of simulations to determine whether our original results would replicate after addressing the limitations.

In our earlier simulations, a language was defined as a set of arbitrary binary features, $F_1 \ldots F_n$, taking the values 0 or 1. The n "genes" of the learners correspond to each of the n features of the language. The genes can take three values, representing an innate bias (0, 1) for a feature being 0 or 1 in the language; or neutrality (represented as '?'). For example, if $n = 3$ the language may correspond to [0, 1, 1] and the genes of a random agent to [?, 1, 0]. At the beginning of each generation, an initial language (phenotype) is expressed for each agent based on its genes (genotype). The innate bias toward a particular feature value will in most cases result in that value being expressed in the phenotype (in most of the simulations the 'stickiness' of the bias is 95% in the direction of the designated value), but on occasion it will be expressed in the opposite direction. For the neutral (learning) genes there is a 50% change of either setting (1 or 0). Thus, in our previous example, the initial language of the agent could be [1, 1, 1]. If the initial language does not match the target language, the agent begins a process of trial and error learning, in which learners randomly sample features using the biases in their genes. Once a feature is 'guessed' correctly, it is not changed. The learner keeps guessing until all the features in its language match those of the target language, with the fastest learners being selected to form the basis for the next generation. Some mutations would occur across generations, with an equal probability of randomly reassigning a gene to 0, 1, or ? (mutation rate varied between simulations). Although the neutral bits initially speeds learning, agents that are genetically biased toward a feature F_i will guess it faster. Thus the Baldwin effect should gradually ensure that all the arbitrary features of the language become genetically encoded.

Chater et al. (2004) found a Baldwin effect for arbitrary linguistic features, for the case where the language is fixed. In these simulations, reproduction was implemented as simple duplications of the top 50% of the learners subject to a 1% mutation rate. Does the same result hold, given a more realistic model of genetic transmission? To better approximate hominid evolutionary dynamics,

Table 1. Number of generations needed to reach the success criterion for the Baldwin effect (parameter value : number of generations)

Genome Size	Population Size	% Initial Neutral Bits	Stickiness of Innate Bias	% Survivors	% Mutation Rate
10 : 25	24 : 369	0 : 23	100 : 152	26 : 52	0.1 : 232
20 : 51	100 : 51	25 : 69	95 : 51	50 : 51	1 : 51
50 : 201	250 : 47	75 : 137	90 : 85	74 : 195	2.5 : 104
80 : 1045		100 : 147	80 : 88		

the current simulations use a simple model of sexual reproduction, instantiated as random cross-over between two sets of learner genes.

We first replicated our original results in which the language/genome size was set to 20, the population size to 100, the number of initially neutral bits to 50%, the 'stickiness' of the innate genetic bias to 95%, the number of surviving agents to the top 50%, and a 1% mutation rate. Using a success criterion that more than 95% of the initial bits in the top 50% of the learners' genomes should correctly match the target language, we found that a robust Baldwin effect occurred after 51 generations. We then varied the simulation parameters and found that a robust Baldwin effect occurred in all circumstances, with parameter variations only affecting the speed with which it emerged (see Table 1). These results show that our earlier results generalize to sexual reproduction, and show that the Baldwin effect is highly robust, with a fixed language. If such a robust effect disappears under when the language is allowed to change, this cannot easily be dismissed.

An important limitation of our original simulations is that language change was completely independent of the genes. It seems reasonable to assume that if the genes control language learnability then they should also influence the direction of language change in a process similar to Baldwinian niche construction (e.g., Odling-Smee, Laland & Feldman, 2003). To explore this, we carried out a set of simulations in which language at time $t+1$ was determined by a combination of genes and language at time t. Specifically, p percent of the change would be determined by the most frequent gene values in the previous population and the remaining $1-p$ percent of change by the previous language. Given that other pressures than learnability also affects language change (such as cognitive/communicative constraints, parsability, language contact, linguistic drift, etc.), we also incorporated random language change at a rate of ten times faster than the mutation rate (i.e., 10%). The faster rate of linguistic change reflects the fact that cultural evolution is much faster than biological evolution (Dawkins, 1976). Whereas linguistic change is measured in thousands of years, biological evolution is measured in hundreds of thousands of years. Other

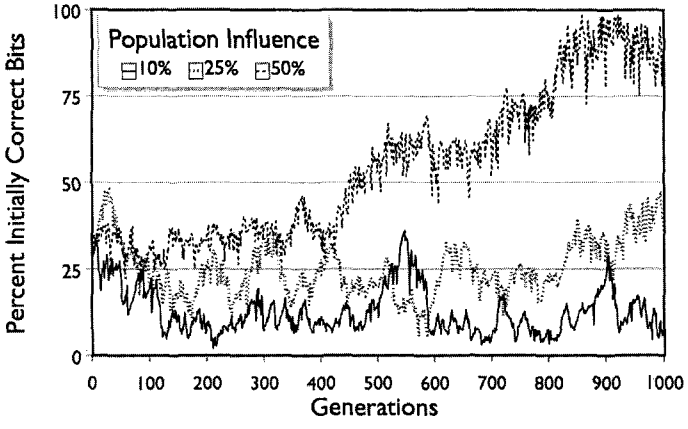


Figure 1. The effect of population influence on the emergence of the Baldwin effect.

simulation parameters were the same as in our initial replication above.

The results of these simulations (Figure 1) show that only when there is a very high degree of population influence does the Baldwin effect emerge. Only when the direction of linguistic change is at least 50% determined by the previous generations genes do we observe a robust Baldwin effect after 835 generations. This suggests that arbitrary features of language would have to be predetermined strongly by the genes from the very beginning, thus leaving little room for subsequent evolution of the kind envisioned by Pinker & Bloom (1990). This corroborates our previous findings that under reasonable assumptions about language change, the Baldwin effect does not occur for arbitrary linguistic features. Unlike the example of hard skin, where the environment provides a stable target for the Baldwin effect, language change is too fast for genetic commitments to arbitrary features to be worthwhile. However, it is possible that non-arbitrary features of language could become genetically fixed in the population if they facilitated communication in some manner; e.g., improved abilities for word learning, increased working memory capacity for language, vocal apparatus optimizations for speech, and so on.

3. Simulation 2: Functional Language Features

Because the arbitrary features of language by definition do not affect communicative function (e.g., Pinker & Bloom, 1990), Simulation 1 did not need to incorporate communication between agents. However, to explore the degree to which functional features of language could have become genetically

fixed via the Baldwin effect, it is necessary to take communication into account to provide a context within which the non-arbitrary features can be functional.

We used the same representation of language and genes as before, with the initial language expressed in the same way. However, learning was implemented differently, now mediated by communicative interactions. Communication was only possible between agents who had a majority of the same kinds on language features (either 0 or 1). Thus, an agent, a_1, whose language is [0, 0, 0, 0, 1], would be able to communicate with an agent, a_2, with a [0, 0, 0, 0, 0] language but not with agent a_3 that has a [0, 1, 1, 1, 0] language. Agents benefit mutually from successful communication in proportion to the overlap in their features. The successful *two-way* interaction between a_1 and a_2 would result in an increase in both agents' communication scores by 9 (the combined number of 0s in their two languages). The simulations also integrate the developmental trend that comprehension precedes production: even though a_1 can only "produce" four 0s, it can "comprehend" a_2's five 0s. However, if the difference between the productive abilities of two agents is more than one unit, then lesser competent "speaker" will not be able to understand its more proficient communication partner, resulting in a *one-way* interaction. In this case, the proficient speaker received the combined communication score (as before), whereas the less competent agent would only receive its own contribution to that score. Hence, if a_2 interacted with a_4, whose language is [0, 1, 0, 1, 0], a_2 would increase its communication score by 8 while a_4's score would only increase by 3.

In this framework, less competent agents are able to learn from more competent agents (with stronger bias towards 0s or 1s); this is meant to reflect the tendency for children to learn much of their language from others with greater language skills than themselves (e.g., adults or older children). Learning can only happen when two-way communication is possible (as described above), and consists in a process in which the less competent agent, based on the biases in its genome, re-samples the first bit in its language that differs from the more competent agent's language. For example, in a communicative interaction between a_1 and a_4, the latter would resample its second language bit. If a_4's genome encoded an innate bias (0 or 1), then there would be a 95% chance of getting this bit expressed; but if the genome encoded a neutral bit, the chance of either value would be 50%. Thus, genes constrain learning as in Simulation 1.

To further mirror the learning conditions from the previous simulations, we introduced noise into the learning process at a rate ten times higher than the mutation rate. During 10% of the learning opportunities a random bit in the

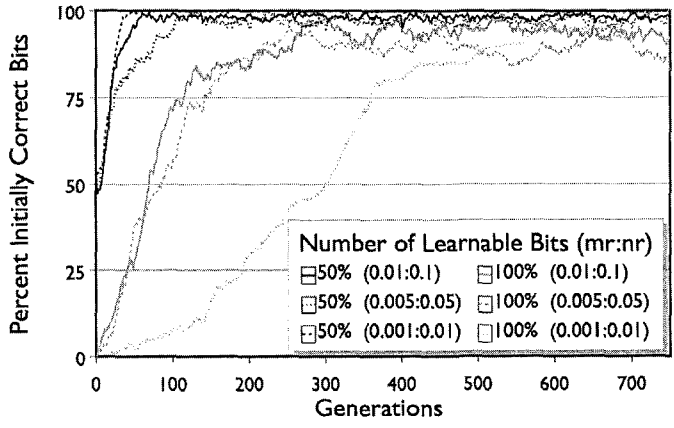


Figure 2. The influence of variations in the number of initial learnable bits on the Baldwin effect for different mutation rates (mr) and noises rates (nr).

learner's language would be chosen for potential reassignment (given the learner's genetic bias for that bit) instead of the first bit that deviated from the competent speaker's language. This paralleled the 10% random change in the target language in Simulation 1.

From each generation of 100 learners, pairs of agents were randomly picked for 500 interactions. The 50 agents with the highest communication scores were selected, and cross-over sexual reproduction used to create the next generation (combined with a 1% mutation rate). The results (Figure 2) show that a robust Baldwin effect emerges across several different variations in mutation rate and number of neutral bits in the first generation. Even when the first generation has all neutral (learnable) bits, a robust Baldwin effect emerges after 33-269 generations. Thus, functional features that improve communication abilities may become genetically fixed in the population. For example, vocabulary learning is likely to rely on innate domain-general abilities for establishing reliable mappings between forms and meanings (e.g., Bloom, 2002). As such, the ability to acquire a large vocabulary may have become gradually innate by way of the Baldwin effect because it would have increased communicative abilities.

4. General Discussion

These results indicate that the Baldwin effect may not provide a suitable evolutionary mechanism for explaining the emergence of arbitrary features of language. Rather, the results suggest that functional features that facilitate communication may be a better candidate for aspects of language that have come

to be genetically fixed over evolutionary time. For a trait to be amenable to the Baldwin effect, it needs to be stable over a period of many generations. Functional features are stable in that they facilitate communication on a continuous basis and thus are likely to become 'Baldwinized' when communicative abilities affect selective fitness in a population. In contrast, abstract linguistic features are free to change randomly exactly because they are non-functional and not subject to direct selective pressures. More generally, the simulations raise doubts about the gradual evolutionary emergence of a UG, as proposed by Pinker & Bloom (1990), and instead support a cultural transmission model of language evolution in which the Baldwin effect has enabled certain cognitive/functional features to become genetically encoded.

References

Baldwin, J.M. (1896). A new factor in evolution. *American Naturalist, 30,* 441-451.

Bloom, P. (2002). *How children learn the meanings of words.* New York: OUP.

Briscoe, T. (2003). Grammatical assimilation. In M.H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 295-316). New York: OUP.

Chater, N., Christiansen, M.H. & Reali, F. (2004). *Is coevolution of language and language genes possible?* Paper presented at the Fifth International Conference on the Evolution of Language, Leipzig, Germany.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Bybee, Joan (1998). A functionalist approach to grammar and its evolution. *Evolution of Communication 2.* 249-278.

Dawkins, R. (1976). *The selfish gene.* New York: Oxford University Press.

Hinton, G.E. & Nowlan, S.J. (1987). How learning can guide evolution. *Complex Systems, 1,* 495-502.

Nowak, M.A., Komarova, N.L. & Nyogi, P. (2001). Evolution of universal grammar. *Science, 291,* 114-118.

Odling-Smee, F.J., Laland, K.N. & Feldman, M.W. (2003). *Niche construction: The neglected process in evolution.* Princeton, NJ: Princeton University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707-784.

Tomasello, M. (2003). On the different origin of symbols and grammar. In M.H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 94-110). New York: OUP.

Weber, B.H. & Depew, D.J. (Eds.) (2003). *Evolution and learning: The Baldwin effect reconsidered.* Cambridge, MA: MIT Press.

ON THE EMERGENCE OF COMPOSITIONALITY

JOACHIM DE BEULE

ARTI, Free University of Brussels, Pleinlaan 2, Brussel, 1050, Belgium
joachim@arti.vub.ac.be

BENJAMIN K. BERGEN

Linguistics Department, University of Hawaii, Honolulu, HI 96822, U.S.A.
bergen@hawaii.edu

Compositionality is a hallmark of human language - words and morphemes can be factorially combined to produce a seemingly limitless number of viable strings. This contrasts with non-human communication systems, which for the most part are holistic - encoding a whole message through a single, gestalt form. We show that compositional language can arise automatically through grounded communication among populations of communicators. The proposed mechanism is the following: if a holistic and a compositional approach are in competition and if both structured (compositional) and atomic meanings need to be communicated, the holistic strategy becomes less successful as it does not recruit already acquired bits of language. We demonstrate the viability of this explanation through computer simulations in which artificial agents perform a communicative task. It is shown that simple reinforcement mechanisms applied during communicative interactions can account for the emergence of linguistic compositionality.

1. Introduction

Compositionality is a universal feature of human language. There are varied accounts of where this feature comes from. On the one hand, it could be a requisite characteristic of human language, dictated by an innate universal language capacity endowed upon us by evolution (Chomsky, 1975; Pinker & Bloom, 1990). Alternatively, it could be a cultural innovation, which different language communities composed of generally intelligent humans have consistently converged on because of its tremendous utility. It is the latter explanation that we explore below. We investigate the conditions under which a language community driven only by the success of individual communicative interactions will come to preferentially adopt a globally compositional, rather than a holistic language strategy.

The literature on cultural causes for the emergence of compositional language (Nowak & Krakauer, 1999; Kirby, 2000; Brighton, 2002; Smith, Brighton, & Kirby, 2003) proposes a variety of factors, including iterated learning (IL), learning bottlenecks, expressibility, and the presence of noise in transmission. While all these explanations provide important insights into potential sources of language

emergence and change, they leave out three critical considerations.

First, they often address only the emergence of the form and not the meaning facet of compositionality. When linguistic units are brought together to form a composite structure, their forms are assembled (one word might precede the other, for example), but so are their meanings. The Enlgish phrase "Jack kisses Mary" not only describes a kissing event involving Jack and Mary, but additionally specifies what roles Jack and Mary play relative to the event. The compositional grammars of Nowak and Krakauer (1999), Brighton (2002) and Smith et al. (2003) are unable to distinguish between "Jack kisses Mary" and "Mary kisses Jack" unless e.g. separate words are used for Jack depending on whether he fulfills the agent or patient role in the kissing event, and the same holds for Mary. Without semantic compositionality, formal compositionality is of questionable utility to language users, and in any case does not approach the expressive power of compositionality in human languages.[a]

Second, in previous work, compositionality emerges at least partially on the basis of experimenter-imposed principles, like an implemented drive to search for generalizations in the learning data and reduce language inventory size (i.e. hypothesize compositional rules). The concern we have with such solutions is that while a trend for decreased inventory size over time (in language learners or populations) might correlate with increased compositionality, implementing this as a causal mechanism effectively causes compositionality to evolve in conformity with the experimenters beliefs about preferred properties of a language system. A more causally satisfying solution would be one in which the language evolves through fundamental principles of communicative interaction, and where changes in inventory size and learnability are byproducts of lower-level causal factors.

Third, IL models typically require many hundreds or thousands of generations for a compositional language to evolve. Moreover, they do not explain how the resulting language could become shared among the members of a larger community, as most IL studies consider only one teacher and one learner. We take issue with such models because of the wealth of evidence suggesting that a population of communicators can arrive at a successful, compositional communication system within one or two generations, as evidenced by the emergence of Pidgins and Creole languages and new signed languages, among others.

These properties of previous models all appear to derive from the assumption that the function of language is irrelevant to its form. Because their focus is on how IL can account for the structure and evolution of language, they (often explicitly)

[a]This criticism does not apply to Kirby (2000) who simulates a population of agents negotiating about how to express 5 '*event*(?ev,?agent,?patient)' type events.(We adopt a logic-based representation of meaning. Symbols starting with a question mark are variables.) He shows that if agents hypothesize generalizations (rules of grammar), then because more general rules have a better chance of being replicated to the next generation, a shared and compositional language emerges. Despite this quite interesting result his model still suffers from two of the concerns mentioned in the text.

disregard linguistic function in their causal models. We present an alternative view on why and how compositional language might emerge, which accounts for both formal and semantic compositionality without relying on endstate-oriented learning mechanisms, and demonstrates convergence on compositional language strategies within a single generation. This view starts from the assumption that the primary function of language is communication. In the experiments we report on, a population of agents is iteratively faced with a communication task in which they talk about a set of observed scenes. The next section spells out the nature of the task and the world in which it is performed. In the following section, a simulation shows how low-level interactive mechanisms - communicative success and a variant of classical conditioning - lead to the emergence of a compositional language when the world obeys certain characteristics. Following (Elman, 1993), we then show how the same effect can be achieved with a wider range of world configurations if instead the agents go through developmental stages.

2. Experimental Setup

Negotiation Model. In our experiments, a speaker and hearer are randomly selected from a population of agents to perform a communication task. They observe a set of scenes, which may differ in terms of their entities or events, the roles the entities play in the events, or combinations of these. The speaker is given one out of the set of events to verbalize. He will use existing language if appropriate, or propose new elements of language if needed, as described below. The hearer decodes the speaker's utterance and tries to identify the topic. If he does so correctly, the game is a success; otherwise it is a failure and the speaker points to the topic. In both cases the agents learn from the outcome.

Other negotiation models could of course be envisioned, and the one used is a radical simplification of the human equivalent. But it is a not unreasonable facsimile of something humans do very frequently both as language learners and language users - namely, observing the world, picking a subset of observations to talk about, and describing that subset using whatever linguistic tools are available. This negotiation framework also importantly allows us to test our main thesis that compositional language can arise automatically in communicative interactions and it does not bias agents towards a holistic or compositional type of solution, only towards a solution that yields communicative success. The same will hold for the language and learning models as explained below.

World Model. Similar to Kirby (2000), the set of potential topics to be verbalized contain partial instantiations of events involving agents or patients of the form '*event*(?ev,?agent,?patient), *person*(?agent)' or '*event*(?ev,?agent,?patient), *person*(?patient)'. Crucially the fact that some entity is playing a particular role in the event, that is, the link (equality) between the ?agent (or ?patient) variable in the event predicate and the corresponding one in the person predicate, is part of the meaning that has to be expressed. If there are N_e different event predicates

(e.g. kiss, kick, etc.) and N_p different person predicates (e.g. John, Mary, etc.), this yields $2N_eN_p$ *structured* topics. The *atomic* topics are also included, i.e. one of the $N_e + N_p$ atomic events or people. The fraction of structured topics with which the agents are presented is called the *task complexity*. When it is 0, agents only have to verbalize the atomic topics, and when it is 1, they verbalize only the structured ones. With intermediate values, they eventually get to see them all.

Language Model. The agents are implemented using the Fluid Construction Grammar (FCG) formalism (De Beule & Steels, 2005; Steels, De Beule, & Neubauer, 2005) which is a general unification-based inference engine, designed to support experiments in the self-organization of language. For the current purposes it suffices to say that in the experiments reported on, an agent's language inventory consists of a lexicon and a set of linking constructions which pair word-order with agent and patient role bindings. Agents start with an empty inventory. Whenever a speaker's lexicon does not cover some meaning to be expressed, he creates a new entry, associating a new form with the uncovered meaning. Crucially, the uncovered meaning can be the complete meaning to be expressed or only part of it. For example, if the meaning is '*kiss*(?ev,?agent,?patient), *Mary*(?agent)', and the speaker does not have a word for *kiss* or for *Mary*, he introduces a new word for the entire meaning, which will be holistic. If, however, he already knows a word for *kiss* but not for *Mary*, he will only introduce a new word for *Mary*, which might lead to compositional language. In this case he also needs a linking construction to express the fact that Mary plays the agent role in the kissing event. The speaker has two choices: he can put the predicate either before or after the agent. For successful communication a different word order needs to be used for encoding the agent and patient roles, but whether this solution is found is left to the learning dynamics, based on communicative success. So it may be that a speaker initially uses the same order to encode both the agent and patient cases, or that different agents use different conventions. Newly introduced words or constructions need to be adopted by the hearer. In our model, after pointing a hearer has all the necessary ingredients to adopt any newly introduced words or constructions if at most one word in the utterance is unknown to him.

A strength (a number between 0 and 1) is associated with all lexical and constructional entries. Because all agents introduce new elements of language, an agent's inventory will quickly contain competing entries, e.g. (partial) synonyms and incompatible word order conventions. When verbalizing a topic, a speaker chooses those entries that have the greatest combined strength.

Learning Model. Agents learn by increasing the likelihood that they will reuse successful language strategies, and decreasing the likelihood of reusing unsuccessful ones. Whenever a hearer successfully understands an utterance, he increases the strength s_u of the linguistic elements used according to $s_u \leftarrow \alpha + (1 - \alpha)s_u$, where the parameter α is the learning rate, fixed to 0.2. In addition he

decreases the strength s_c of all competing elements according to $s_c \leftarrow (1-\alpha)s_c$. This decrease in the strengths of competing entries, a sort of lateral inhibition, is needed to reduce the size of the inventories (i.e. to 'forget' synonyms and reach a coherent language) and introduces a competition between the holistic and compositional strategies. In case of a failed game, both the speaker and the hearer decrease the score of all used entries. This updating scheme can be implemented as an associative memory updated according to the Rescorla-Wagner/Widrow-Hoff learning rule as described in Sutton and Barto (1981). It typically gives rise to the following dynamics. Initially, because agents start with empty inventories, many new elements of language are introduced and the size of language inventories grows. Negotiation then gradually strengthens some of the entries while weakening others and the overall communicative success rises. Entries for which the strength becomes too small are forgotten and the inventory sizes decrease again.

3. The Emergence of Compositionality

Simulations using the framework described above show that a compositional language can only emerge when the task complexity is different from 0 or 1. If it is 0, that is if only the $N_e + N_p$ atomic topics are considered, then the agents never have to express any structured meaning. If it is 1, that is if only the $2N_eN_p$ structured topics are considered, then they never get a chance to introduce words covering only the atomic predications and because the agents do not generalize they again will evolve a stable holistic language. This is also shown in Fig. 1. The simulations also show that a successful communication system only emerges when the complexity is below some threshold (about 0.2 for the simulations shown, or some higher value when either the learning parameter, the population size, or the world size is decreased).[b] These languages are compositional - the bottom figure shows that they require 10 lexical items to communicate the 60 possible topics.

The mechanism at work is the following. Whenever an agent has to verbalize structured meaning, his structured and atomic lexical entries compete, and whichever strategy wins will weaken the other. On the one hand, the compositional strategy needs three successful elements of language to be successful itself: two atomic lexical items and one linking construction. In contrast, the holistic approach only needs one. This favors the holistic strategy. On the other hand, this advantage of holistic strategies is counteracted by the fact that a holistic lexical item has fewer opportunities for application, and hence its chances of spreading through the population are smaller, comparable to the replicator dynamics governing Kirby's (2000) IL scheme. The lower the task complexity (as long as it is

[b]Communicative success also goes to 1 when the task complexity is 1.0 because in this case the game is equivalent to a standard naming game about $2N_eN_p$ topics in which there is obviously no competition between the holistic and compositional strategies. However for intermediate task complexities this built-in competition seems to prevent the agents to reach a successful communication system.

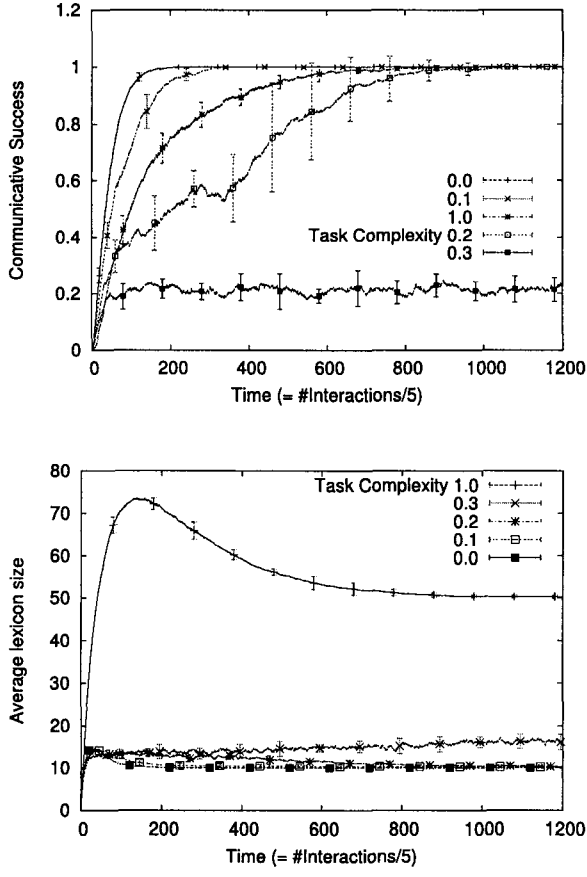


Figure 1. Influence of the task complexity on the evolution of the communicative success and the average lexicon size for several values of the task complexity. The curves were obtained for 5 agents, $N_e = 5$ and $N_p = 5$, averaged over 10 independent runs.

positive), the more of an advantage the compositional approach displays, since for lower complexities the unstructured (atomic) meanings need to be verbalized relatively more often, allowing the agents to evolve successful simple lexical items, which can then be recruited for communicating structured meanings in a compositional fashion. In sum, then, if the initial task complexity is low enough, the population invariably adopts a successful compositional strategy. (Once successful communication is established, the task complexity is of no further importance.)

One potential drawback of this demonstration of the emergence of compositionality is the constraint that initial task complexity must be low. As explained,

this is due to the built-in competition between structured and atomic lexical entries together with the fact that a successful language always requires atomic entries for expressing unstructured meanings. Removing this competition invariably leads to a successful mixed holistic/compositional language, with the degree of compositionality depending on the task complexity. However, as it turns out, there is another configuration that leads to the emergence of a fully compositional language with a broader range of task complexity settings. Following Elman (1993), we can consider whether instead of manipulating the world (the task complexity), we can let the agents themselves go through developmental stages. Elman showed that a gradual increase of attention span or, equivalently, a gradual increase of memory size allowed his neural networks to solve tasks that were unsolvable when starting with a 'full-grown' network. Implementing attention and memory limitations in our agents can be achieved by letting the agents ignore complex situations, i.e. structured topics. To test this, we conducted experiments in which the agents were presented with an equal number of structured and unstructured meanings (i,e, the task complexity was 0.5), but in which a developing learning scheme is adopted. Namely, as long as an agent's communicative success is below 0.7 it ignores all structured topics. Subsequently, as the success increases, gradually more of the structured topics are considered. When 95% success is reached all topics are considered. Not surprisingly, the result is that successful compositional language indeed emerges, even with this higher task complexity. Although more elaborate experiments are needed, the time scale at which this happens seems to be a linear function of the number of rules of language required – i.e. of $N_e + N_p + 2$ in this case: the atomic lexical items plus two constructions.

Such a developmental mechanism has at least two clear correlates in human language use. Perhaps most obviously, children's ability to perceive and conceptualize events is known to increase as part of normal cognitive development. Alternatively, or perhaps in complementary fashion, the development of language about atomic meanings could necessarily precede language about structured meanings; as expertise with language increases, so does the complexity of possible topics. Regardless, the proposed developmental mechanism is a feasible candidate for explaining how humans can arrive at compositional language within one generation, and even more so when a relatively stable simple lexicon is already at hand, which is the case in the development of Creole languages from Pidgins.

4. Conclusion

Compositional language can arise automatically through grounded communication within populations of communicators. This language is compositional both in terms of form and in terms of meaning, and arises within a generation, over the course of hundreds of communicative interactions. Crucially, this is accomplished as the emergent product of simple communication and learning mechanisms. Nothing in the agents' architecture biases them towards compositionality.

Instead, the agents are implicitly driven to re-use already established features of language in order to be successful, and are thereby guided towards adopting a compositional strategy.[c] Possibly due to the simplifications made in our agent and learning models, this mechanism only works either when the agents more frequently need to express unstructured meaning or else when they go through developmental stages, which could be explained either cognitively or else in terms of expertise. The emergence of compositionality can be most parsimoniously explained not as an innate universal language capacity endowed upon us by evolution, but through the communicative benefit it brings to individual language users.

References

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*(1).

Chomsky, N. (1975). *Reflections on language.* New York: Pantheon Books.

De Beule, J., & Steels, L. (2005). Hierarchy in fluid construction grammar. In F. U. (Ed.), *Proceedings of ki-2005* (p. 1-15). Berlin: Springer-Verlag.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71-99.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, J. Hurford, & M. Studdert-Kennedy (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form.* Cambridge University Press.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proc. Nat. Acad. Sci. USA*, *96*, 8028-8033.

Pinker, S., & Bloom, P. (1990). Natural languages and natural selection. *Behavioral and Brain Sciences*, *13*, 707-784.

Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Artificial Life*, *9*(4), 371-386.

Steels, L., De Beule, J., & Neubauer, N. (2005). Linking in fluid construction grammar. In *Proceedings of bnaic-05.* Brussels, Belgium: Royal Flemish Academy for Science and Art.

Sutton, R., & Barto, A. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*(88), 135-170.

Vogt, P. (2005). On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. *Adaptive Behavior (Special issue on Evolution and Acquisition of Language)*, *13*(4).

[c]This mechanism might be related to what is called the 'implicit bottleneck' (see e.g. Vogt (2005)) that is introduced because agents are by definition only partially exposed to the emerging language as the language is still under development

TOWARDS A FIXED WORD ORDER IN A SOCIETY OF AGENTS: A DATA-DRIVEN BASELINE PERSPECTIVE

GUY DE PAUW

CNTS - Language Technology Group
University of Antwerp - CDE
Antwerp, B-2610, Belgium
guy.depauw@ua.ac.be

In this paper we present a computational model for the emergence of a fixed word order in the language of a society of babbling agents. In contrast to the majority of research modeling the emergence of syntactic principles, in which a very strong governing role is set aside for the semantic aspect of communication, this paper touches on the possibility that one of the most basic grammatical principles, i.e. fixed word order, can develop without direct reference to concepts of shared meaning. With its rigid data-driven approach, the system described in this paper introduces a hitherto unexplored baseline perspective to the research efforts modeling the emergence of grammatical properties in language.

1. Introduction

The last decade has seen many researchers propose computational models as a way to provide some insight into how early hominids may have evolved into language users. Very interesting results have been achieved in the computational modeling of the emergence of structural properties and compositionality in language (Hashimoto & Ikegami, 1996; Kirby & Hurford, 2001; Batali, 2002; Steels, 2004). While sharing the same general research objective, these models differ from one another in the incentives they define for raising the complexity of communication into the realm of compositional language and what kind of cognitive capacities are assumed to be able to do so.

Even though computational modeling provides an interesting account of how different properties of compositional language may have emerged, some fundamental issues are apparent, as agents often find themselves equipped with a very intricate set of linguistic and cognitive principles that are inherently biased towards compositionality. While there is some evidence that these kinds of cognitive capacities might have been available to early hominids, a more problematic aspect underlying many of these computational models, is the concept of explicit meaning sharing in inter-agent communication (Kirby & Hurford, 2001; Batali, 2002). Following the 'gavagai' argument in Quine (1960), the referent of a signal

should indeed not be considered as something that is unambiguously available to participants in a conversation.

In this paper, we describe a computational model that approaches the problem from an entirely different angle. It minimizes the semantic value of the agents' communicative attempts and avoids the problematic concept of shared meaning altogether. Experiments show that even in this overly restrictive setup, convergence towards fixed word order can emerge out of initial chaotic attempts. With its rigid data-driven perspective to language evolution, the baseline model presented in this paper raises some interesting questions as to the role of distributional properties in (i) the emergence of governing principles in compositional language and (ii) the previously unexplored role of statistical conventionalization effects in the evolutionary computational models of language.

2. Stripping Down Data-Driven Grammar Optimization

The system described in this paper is essentially a stripped down version of an agent-based system called GRAEL that is geared towards data-driven optimization and induction of grammars for natural language (De Pauw, 2002). The instantiation of GRAEL described in this paper discards the notion of a pre-defined language. The agents are equipped with the ability to create mental constructs and process agent-patient type relationships. They also have the capacity to record the frequency of linguistic events, a psycholinguistically realistic cognitive concept (Juliano & Tanenhaus, 1993).

We also assume the agents have acquired the naming insight. This translates in the following limited lexical knowledge attributed to the agents: a lexicon of 40 single-entity names, including the 10 agents in the society and 30 objects with an associated [±animate] subcategorization classification. They also have a repertoire of 4 attributes expressing singular properties for [+animate] entities (e.g. {**hungry** `tiger`}) and 20 terms that express relationships between a [+animate] agent of an action and its patient (e.g. {**eat** {`hungry tiger`},`agent8`}).

In accordance with Wray (1998) who proposes that the naming insight occurs at the cross section between compositional and holistic language, we simultaneously provide the agents with a set of 200 randomly generated holistic utterances expressing entire SVO relationships in one single term, but allowing for wildcards (e.g. {**eat** `[+animate],any`}) to express general concepts.

3. Something to talk about

The experiments described in this paper take place in a closed world of 10 agents and the 30 objects named in the lexicon. We assume the conversations to be a form of gossip, as suggested by (Dunbar, 1996), and suggest that the referents for these utterances are not visually available to the listener. This entails that the agents only have language to convey meaning without the help of gesturing or gaze following. When the speaker tries to convey a situation to the listener, he

can only do so by uttering words in a specific word order. For the hearer, the only extra-linguistic tool for disambiguation of the utterances is the knowledge that the internal structure of the state-of-affairs expressed by the speaker does not change over time and the internal agent-patient relationships remain static.

Each agent in the society is provided with a single, static state-of-affairs to talk about, represented as a 9cell situation matrix (Figure 1, left). There are 8 participants in each state-of-affairs and the [+animate] participants have some attribute assigned to them (represented as circles in Figure 1). Relationships are defined between neighboring cells of the table, which can hold between the atomic entities in the cells. The concept *Agent6 and Agent7* for example could be expressed by `{and 4,1}`. Relationships can also be used as entities in a relationship themselves. This allows the situation *Agent6 and Agent7 chase a hungry tiger eating a pig* to be expressed by `{chase {and 4,1},{eat {hungry 5},8}}`.

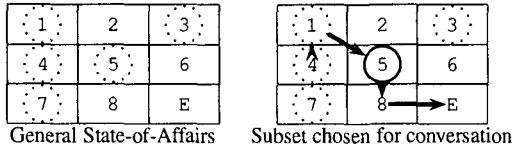


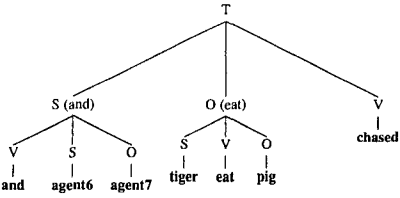
Figure 1. State-of-Affairs: Situation Matrix

In each communicative attempt a subset of the state-of-affairs is discussed. An agent will choose a cell from the table and follow a random path to the end-state `E`. The path through the situation matrix on Figure 1 (right) for example expresses the above example *Agent6 and Agent7 chase a hungry tiger eating a pig*, starting at cell `4` and ending in cell `E`. For a more detailed overview of how these state-of-affairs are internally represented, we refer you to De Pauw (2002).

4. Processing Utterances

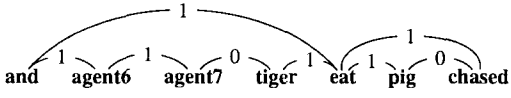
The concept of the situation matrix implies that the agents have the cognitive capacity to process structured mental representations. We assume that the same mental capacity also allows the agents to make a linguistic distinction between the agent and patient of an event. We define six different possible ways to order the agent (**S**ubject), action (**V**erb) and patient (**O**bject) of the relationship: **SVO**, **SOV**, **VSO**, **VOS**, **OSV** or **OVS**. Attributes are represented as either **VS** or **SV**.

Different relationships can be expressed using different orderings. Let us suppose the agent decides to generate the fully compositional (i.e. not using holistic tokens) utterance expression for the event `{chase {and agent6,7},{eat tiger,pig}}`. He therefore needs to express three different constructs, which each can be produced using a different ordering, as exemplified in the following structure:



The utterance the agent uses to express the complex meaning is consequently *And agent6 agent7 tiger eat pig chased.* Although this method does not establish totally free word-order, there are still many different possible word orders available to express the mental structure.

After generating the sentence, the speaker will store the frequencies of the component bigrams of the utterance in memory. The basic frequency of a bigram '*a b*' is defined as its total number of occurrences, divided by the total observed number of occurrences of '*a*'. The speaker however is allowed to refine the notion of a bigram by referring to the structural properties of his utterance. Adjacent bigrams are not counted if they belong to different nodes in the structure, while bigrams of non-adjacent head-nodes in a structure are. This is illustrated in the following example:



In subsequent productions, the agent will still consider the six aforementioned possibilities, but will calculate for each ordering the probability by multiplying the frequency values of its component bigrams. The ordering with the highest probability will consequently be the one the speaker prefers to use in his utterance.

Let us now turn to the **hearer** agent. He observes this sentence, but has no access to its meaning, nor its underlying structure. He simply records the frequency of the bigrams, so that he can use these frequencies later to apply a preferred ordering to new utterances. Structural information is not available to the hearer, so the bigrams are simply recorded as they are sequentially observed, without structural properties influencing the bigram count.

After processing the sentence, the hearer will have some vague notion of what exactly the speaker is talking about. The hearer agent will then pick one of the objects from the utterance and repeat it to the speaker, who will then turn back to his situation matrix, find the object the hearer wants to know more about and track a new path to the end-state, after which a new utterance is generated.

During each language game, the hearer will ask 500 such questions, so that in the end, plenty of information about that particular state-of-affairs has been conveyed. This obviously does not constitute a realistic model of conversation. A more realistic number of questions per conversation could be used if we proportionally raise the average number of conversations during an agent's lifespan, but

this would negatively affect computational complexity.

Previously, we also mentioned a limited set of 200 randomly generated holistic utterances expressing general SVO situations (e.g. {eat [+animate],any}). A speaker agent can chose to use these holistic utterances to express (part of) the meaning. We do place an important restriction on the use of holistic tokens based on its probability: a holistic token can only be used if its probability is higher than the sum of the bigram probabilities between the names of its atomic concepts. The idea behind this is that it is only useful to use a holistic token, if it does not overly generalize the situation and miss out on salient information.

The probability of a holistic token will benefit from its general nature: a holistic token might therefore be more widely applicable. The atomic names on the other hand can be used less generally, but in a wider range of combinations. The experiments will show how these two conflicting forces work alongside each other: even though the experiments show compositionality is usually preferred, holistic artifacts do sometimes survive in the language of the society.

5. Experiments with GRAEL

We have performed ten different experiments, with roughly the same results. We single out a typical experiment for discussion. Each experiment was performed on a 10-agent society over the course of 5000 language game runs. During each run, each agent in the society is assigned as a hearer and as a speaker once. The GRAEL society is generation-based, but does not employ fitness functions, nor any type of genetic transmission of information. At some points in the society new agents are introduced, while other agents die at random intervals (lasting at least 200 runs, but with an upper-bound limit of 1000 runs on their life-span). This allows for a dynamic population size, which is only restricted by imposing a lower-bound society size of 5 agents.

5.1. *Quantitative Evaluation*

To check whether the agents are indeed developing some shared notion of word order, we halt the society after every 1000 runs and extract the agents. We then place (the same set of) 100 pre-defined meanings in their mind, for which they are asked to render sentences. The degree to which the sentences that are produced are the same among the agents, expresses the society's convergence on a word order.

To illustrate this convergence, we use the convergence diagrams in Figure 2. These diagrams display the 100 meanings, starting off with 4 attributes, followed by 36 3tuple relationships (4-21 have a [+animate] object, 22-39 a [-animate] object, all have [+animate] subjects), followed by 60 complex meanings. These meanings are ordered in increasing complexity of up to 15 relationships per utterance.

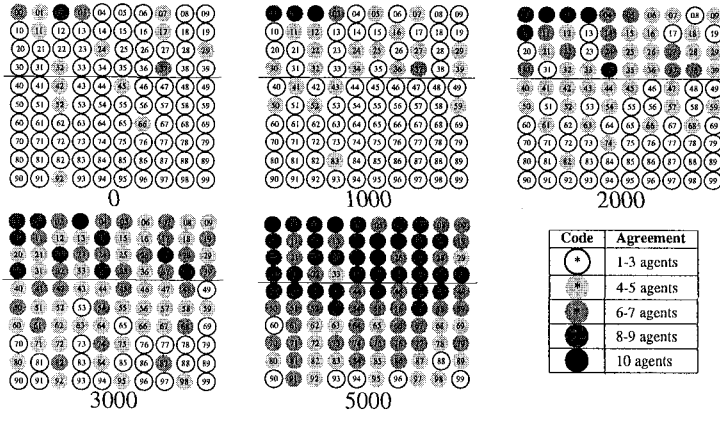


Figure 2. Convergence Diagram: shared word order in the society

The first diagram labeled '0' in Figure 2 displays the situation at the start of the GRAEL society. The 4 attributes (meanings expressed by at most two tokens) are trivially found by at least half of the population. The word order for some of the simple relations is also shared by several agents, either by a very lucky ordering of the words, or by several agents choosing the same holistic utterance to express a complex meaning, thereby reducing the randomness effect of word ordering.

After about 1000 language game runs, the word order for the attributes is shared by almost all agents. Looking at the diagram after 2000 runs, we notice that there seems to be a further tendency towards convergence for the word ordering of the more complex relationships as well. There is a noticeable difference between the meanings that express a relation between two **[+animate]** objects (meanings 4-21) and meanings that express a relationship with a **[-animate]** patient (meanings 22-39). On the whole, the word order seems less random for the former. This may be explained by agents having a much easier time distinguishing word order between two classes of objects, whereas it is not transparent from the word order alone which object is the agent and which is the patient.

The convergence diagram after 3000 runs shows that this positive trend of convergence continues. All simple meanings now display at least some limited degree of convergence, with many word orders shared by more than half of the population. It is mainly the young agents in the society that account for the randomness that is still present in the word ordering mechanism of the society. As convergence continues however, the newborn agents will be met with more consistent word order patterns, which helps them to pick up general tendencies more quickly. This is in line with the notion of the transmission bottleneck coined in Kirby and Hurford (2001).

At the end of the experiment, many complex meanings are being expressed

in the same word order by more than half of the population. There is less disagreement on how to order sentences in which relations themselves are used as agent or patient, so that these patterns can be picked up faster by the newborn agents. Convergence has improved on some meanings, while other (mostly complex) meanings seem to have become a bit more randomized (e.g. (60)) compared to the earlier situation at 3000 runs.

The experiments show that introducing an amplifying function of statistical processing in language, allows the society to evolve from chaotic word order to a generally shared, but not finitely converging concept of word order. It appeared indeed during many experiments that the society is drifting from one local maximum of convergence to the next. We do not consider this a problem, as a non-converging model of language evolution constitutes a more realistic one.

5.2. *Holistic Artifacts*

Holistic elements disappeared over time in about half of the GRAEL experiments, with only a few of them surviving in the other half. This can be explained by the large amount of different meanings that need to be expressed by the agents, allowing for high probability values of atomic names, which can easily overtake that of the holistic token.

In four experiments two holistic meanings had been maintained throughout the society over time. Analysis of the data did not reveal any reason as to why these particular phrases had been maintained and why this occurred in some experimental runs and not in others. We do not consider the survival of holistic items a problem, as Wray (1998) suggests holistic utterances are always present in natural language, even today, as living fossils of an earlier protolanguage.

6. Discussion

In this paper we introduced GRAEL as a computational model for the emergence of a fixed word order without using the problematic tool of shared meaning. Simply on the basis of general cognitive mechanisms like the ability to build mental constructs and the ability to register frequencies of utterances, we have shown how fixed word order can evolve in a society of babbling agents. This provides a previously unexplored data-driven point of view to the problem of modeling language evolution. Even though we have made some unrealistic assumptions in the implementation of GRAEL, most particularly the marginalization of the aspect of meaning in communication, we propose that simple distributional effects of language use can account for the emergence of inherently formal and syntactic constraints on language. The almost trivial effect of statistical conventionalization may indeed explain the convergence towards a fixed word order in our experiments and we tentatively propose that this "artifact" of language use may actually be a strong amplifying function in language evolution itself. Further experimentation in the same vein however is warranted to support this claim.

There are however some obvious problematic issues in our interpretation of the data we have purposely ignored up to now. When we observe agents using the same word order to express the same meaning, we might consider them to be able to understand each other. But we are faced with the empirical problem that this type of convergence may to a large extent be due to statistical conventionalization, rather than true understanding between agents. Whether or not GRAEL then constitutes a realistic model becomes an uncomfortably moot point. This raises an interesting question: the simple statistical point of view adopted in this paper is an aspect of computational models of language evolution that has up to now been largely ignored. If it is indeed possible to model the emergence of fixed word order solely on the basis of distributional properties of language use, we need to reflect on the exact nature of the governing role of semantics in this type of modeling and start situating advances in the field against the background of a quantifiable baseline. This paper has attempted to offer an example of such a baseline and hopes to instigate a more general discussion on the signal-meaning relationship in the research field of computational modeling of language evolution.

Acknowledgments

The research described in this paper was financed by the FWO (Fund for Scientific Research).

References

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (p. 111-172). Cambridge University Press.

De Pauw, G. (2002). *An agent-based evolutionary computing approach to memory-based syntactic parsing of natural language.* Doctoral dissertation, University of Antwerp, Antwerp, Belgium.

Dunbar, R. (1996). *Grooming, gossip and the evolution of language.* London: Faber.

Hashimoto, T., & Ikegami, T. (1996). The emergence of a net-grammar in communicating agents. *Biosystems, 38*, 1-14.

Juliano, C., & Tanenhaus, M. (1993). Contingent frequency effects in syntactic ambiguity resolution. In *Fifteenth annual conference of the cognitive science society* (p. 593-598). Hilsdale, NJ.

Kirby, S., & Hurford, J. (2001). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148). London: Springer Verlag.

Quine, W. V. O. (1960). *Word and object.* Cambridge: MIT Press.

Steels, L. (2004). Constructivist development of grounded construction grammars. In *Proceedings annual meeting association for computational linguistics conference* (p. 9-16). Barcelona.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication, 18*, 47-67.

SIMULATION MODEL FOR THE EVOLUTION OF LANGUAGE WITH SPATIAL TOPOLOGY

CECILIA DI CHIO
Department of Computer Science, University of Essex
Colchester, CO4 3SQ, UK
cdichi@essex.ac.uk

PAOLO DI CHIO
Dipartimento di Sistemi e Istituzioni per l'Economia, University of L'Aquila
L'Aquila, 67100, Italy
pdc@ec.univaq.it

In this paper, we present an agent-based simulation model for the evolution of language. This is based on a previous model proposed by the authors and inspired by Nowak's simplest mathematical model. We extend our previous work with the introduction of a significant characteristic: a world where the languages live and evolve, and which influences interactions among individuals. The main goal of this research is to present a model which shows how the presence of a topological structure influences the communication among individuals and contributes to the emergence of clusters of different languages.

1. Introduction

The genetic and linguistic systems follow two parallel evolutionary trajectories, i.e. they co-evolve (Cavalli-Sforza, 2000). Isolation, either social or geographic, causes independent evolution and genetic differentiation. The same happens with languages: isolation reduces cultural exchanges and the languages of isolated populations becomes more and more different. The study of the emergence of these isolated clusters of languages has been the motivation for our research. In order to achieve our goal, we have used the evolutionary theory of games, and in particular the theory of *evolutionary language games*, together with *agent-based simulation models*.

The theory of evolutionary language games arises from the union of *evolutionary game theory* (Maynard Smith, 1982) and the theory of *language games* (Wittgenstein, 1953).

The simulation model we present is an extension of our previous model (Di Chio & Di Chio, 2005), which was inspired by Martin Nowak's simplest mathematical model (Nowak & Krakauer, 1999), (Nowak, Plotkin, & Krakauer,

1999), (Nowak, 2000). In this new model, we introduce a characteristic that we think is crucial for a more realistic simulation of the evolution of language: a topological structure (the *world*) where the (individuals who speak different) languages live and evolve. This structure influences the interactions among the individuals and contributes to the emergence of clusters of different languages.

This model is an agent-based simulation. These kinds of models are characterised by a certain number of agents which can control their own behaviour according to their perception of the environment they live in. The goal of an agent-based simulation is to create agents which are able to interact with the environment in an intelligent way. For this reason, these simulations are widely used in modelling artificial intelligence and artificial life. Examples of agent-based simulation systems are cellular automata, ant systems (Bonabeau, Dorigo, & Theraulaz, 1999) and particle swarm systems (Kennedy & Eberhart, 2001).

The rest of this paper is organised as follow. In the next section we describe our simulation model. In section 3 we present some results, and we conclude in section 4.

2. The simulation model

Nowak's mathematical model describes quite accurately the emergence of a linguistic system but, at the same time, it is based on simple assumptions. In particular, there is no environment able to influence the communication among the individuals. Since isolation is one of the main reasons for the differentiation of languages and the emergence of linguistic groups, we developed a simulation model adding to Nowak's a *world*: an environment with a topological structure where the (individuals which speak the) languages live and evolve.

The world where the agents will live is a 2-D discrete grid whose x and y dimensions are exogenous parameters and which is topologically equivalent to a torus. Agents represent individuals as well as languages and do not move around, but are in a fixed *cell* location. They produce offspring, which will be generated and put into the environment according to a certain set of rules. As in the previous model (Di Chio & Di Chio, 2005), the size of the population is constant in time and each new generation completely replaces the old one.

Following Nowak's notation, each language $\mathcal{L}$ is defined by a relationship between a finite set of n *objects*, and a finite set of m *sounds* (the *vocabulary* of the language). The matrices $\mathcal{A}$, $\mathcal{P}$ and $\mathcal{Q}$ retain their meaning. The payoff of the language game between agents a_h and a_k will be written as

$$\pi(a_h, a_k) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} (p_{i,j} q'_{j,i} + p'_{i,j} q_{j,i}) \tag{1}$$

The similarities with the mathematical model end here. The computation of fitness, the generation of offspring and the positioning of the newborn agents, now

take into account the presence of the world. Let us examine each of these issues in detail.

> The fitness function is modified in order to be influenced by the distances between the individuals, in such a way that the contribution to the fitness of the agent a_h is higher for closer individuals it plays with. This mirrors the real world situation, where communication is more likely to happen between individuals which are closest to each other (using some suitable metric).
>
> The number of offspring that each agent generates will be proportional to the agent's fitness, but the factor of proportionality changes. It is no longer the global fitness (the fitness of the whole population) but a "locally global" fitness, i.e. the fitness of a suitable neighbourhood of the generating agent. To avoid too abrupt a separation among agents, we adopt a fuzziness in the definition of neighbourhood, weighting the fitness of pairs of agents with a smooth function. This is the key point which (eventually) leads to the emergence of many groups of individuals, and therefore to many cluster of languages.
>
> We position the newly generated agents following one of two strategies: (a) put all offspring in (a *list* in) the same cell as the parent or (b) put offspring in *neighbouring* cells. These have been chosen to mirror a more (the latter) or less (the former) strong isolation process.

Under these assumptions, we no longer have the emergence of a single, (possibly) optimum language spoken by the whole population. Instead, what we observe is the emergence of a certain number of *clusters* of different languages. This happens because nearby agents have fitnesses high enough to generate offspring, which in turn will be near to each other.

More formally, let $d(a_h, a_k)$ be the euclidean distance between the agents a_h and a_k and $\rho(a_h, a_k) = e^{-d(a_h, a_k)}$ the function of d we will use to weight the payoffs. The fitness ϕ for a_h is given by

$$\phi_{a_h} = \sum_{k \neq h} \pi(a_h, a_k)\rho(a_h, a_k) \tag{2}$$

We then need to weight each individual's global fitness in a similar way. Thus, not only will the communicative ability be influenced by the distance, but also the number of offspring.

To compute the number of offspring, we have to take into account the "locally global" fitness. If $\Phi(a_h)$ is the global fitness relevant to the individual a_h, and A_h is a suitable neighbourhood of a_h, we have

$$\Phi(a_h) = \sum_{a_k \in A_h} \phi_{a_k} \tag{3}$$

The number of offspring s_{a_h} for a_h is proportional to the ratio between the individual's fitness and the global fitness, that is

$$s_{a_h} = n_{A_h} \frac{\phi_{a_h}}{\Phi(a_h)} \propto \frac{\phi_{a_h}}{\Phi(a_h)} \tag{4}$$

where n_{A_h} is the number of agents in A_h. We do not know A_h, but we can "fuzzify" it and write (for the global fitness)

$$\Phi(a_h) = \sum_k \phi_{a_k} \rho(a_h, a_k) \tag{5}$$

and for n_{A_h}

$$\tilde{n}_{A_h} = \sum_{k \neq h} \rho(a_h, a_k) \tag{6}$$

In each generation, the population size N is constant. Therefore, we have

$$\sum_h s_{a_h} = \sum_h n_{A_h} \frac{\phi_{a_h}}{\Phi(a_h)} = N \tag{7}$$

whilst

$$\sum_h \tilde{n}_{A_h} \frac{\phi_{a_h}}{\Phi(a_h)} = M \tag{8}$$

Thus (to retain population size N per generation) the actual number of offspring for each individual is given by

$$s_{a_h} = \frac{N}{M} \sum_{k \neq h} \rho(a_h, a_k) \tag{9}$$

where N/M is a normalisation factor.

At each generation, the offspring of the same language will be close to each other, their fitnesses will be higher, and they will leave more offspring. This is a phenomenon which happens locally and therefore we expect to observe the process of language clustering.

Starting from a population of many different languages (i.e. from many different populations, each one made of just one language), the simulation shows how these languages spontaneously move (closer or further away) until the emergence of independent populations. This happens without any form of "artificial" constraint but thanks just to communication.

2.1. *Implementation of the model*

To implement the simulation model, we have used the *Swarm* platform (Swarm Development Group, 2000) and the *Objective-C* programming language, as we did for our previous model (Di Chio & Di Chio, 2005).

This simulation has a model swarm called `LangGameModelSwarm`. This creates the lists of the present, past and newborn languages, generates the offspring and manages the language game (the turns in the game for the languages). The agent `Language` deals with actions intrinsic to the language, such as creating the matrices $\mathcal{A}$, $\mathcal{P}$ and $\mathcal{Q}$ and sampling $\mathcal{P}$, playing the language games (i.e. computing the fitness), computing its own coordinates in the world and calculating the distance between itself and the other agents. The agent `LangSpace` represents the world which contains the languages. Finally, there are two different observer swarms, which perform similar actions, the only difference being that one makes the observation in a graphical model through the use of a GUI and therefore also deals with the function to manage the graphical interface. For more details on the implementation see (Di Chio, 2004).

3. Results

We have conducted several experiments to ensure the robustness of the model[a]. The settings for the parameters of the simulation are summarised in table 1.

Table 1. Parameters settings.

Parameter	Value
(objects, sounds)	$(5,5)$, $(10,10)$, $(25,25)$
Population size	100 individuals
Sampling parameter k	1, 4, 7, 10, 25
Generations	100
Iterations	20

We have run different experiments according to the positioning of the offspring in the world (neighbouring cells lookup or list) and whether the fitness is weighted or not with the distance.

The graphs in figure 1 show the results of the simulations (with the smallest vocabulary and sampling parameter values 1 and 25) when the distance influences both the individual's fitness ϕ and the locally global fitness Φ, and the population is replaced with neighbourhood lookup.
Those in figure 2 show the results of the simulations (with the same parameters as before) when the distance influences both the individual's fitness ϕ and the locally global fitness Φ, and the new population is positioned in a list.
The last two graphs (fig. 3) show the configuration of clusters in detail. In particular, we can observe that, if the replacement is with neighborhood lookup, it is possible to have clusters with more than one language, whilst if the population is positioned in a list, there is just one language in each cell.

[a] All the simulations have been run on a 2.4GHz Intel Pentium 4® CPU with 512MB RAM on the RedHat® Linux 9.0 operating system.

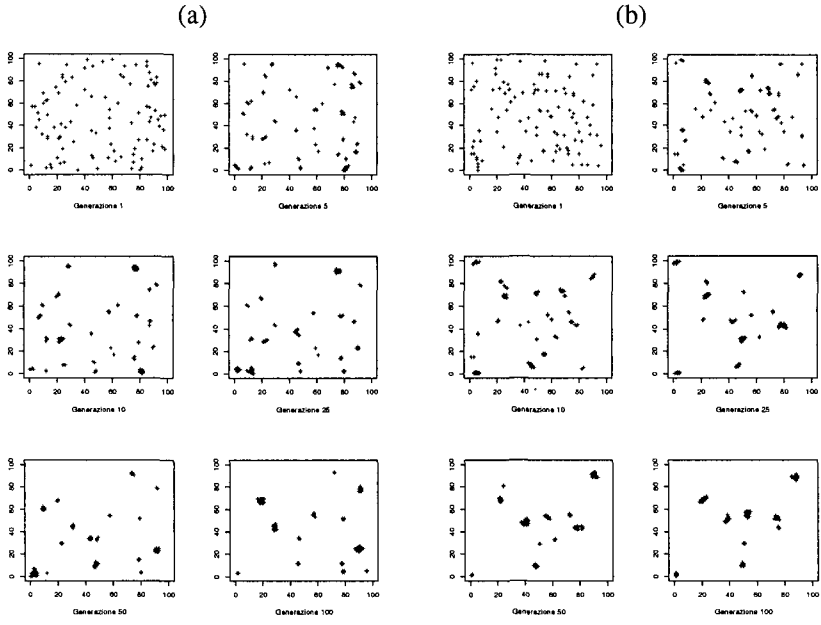


Figure 1. *Simulation model with (objects,sounds) = (5, 5) (a) $k = 1$ (b) $k = 25$. Distance influences both ϕ and Φ. Population replaced with neighbourhood lookup.*

For a more comprehensive set of graphs, as well as a complete list of clusters and theirs characteristics, refer to (Di Chio, 2004).

3.1. *Analysis of the results*

As the simulation results show, it is clear how important the presence of a topological structure is for the behaviour of the languages. We can in fact observe, by varying parameters, the emergence of different clusters of different languages.

The replacement with neighbourhood lookup causes the clusters to continually evolving. This happens because, by positioning the new individuals in the cells around their parents, the dimensions of the cluster are continuously varying, and therefore the distance among individuals in different clusters changes from one generation to the other. These variations help in the emergence of new languages in new positions (i.e. positions different from the starting ones).

On the other hand, positioning the new population in lists is a way to clearly highlight the process of cluster creation. Since all the offspring of an individual are placed in the same cell, the spatial dimensions of the clusters are constant (and equal to 1 cell). Therefore, in this situation we will not observe the emergence of new languages in new positions, but only the disappearance of isolated (weaker) languages.

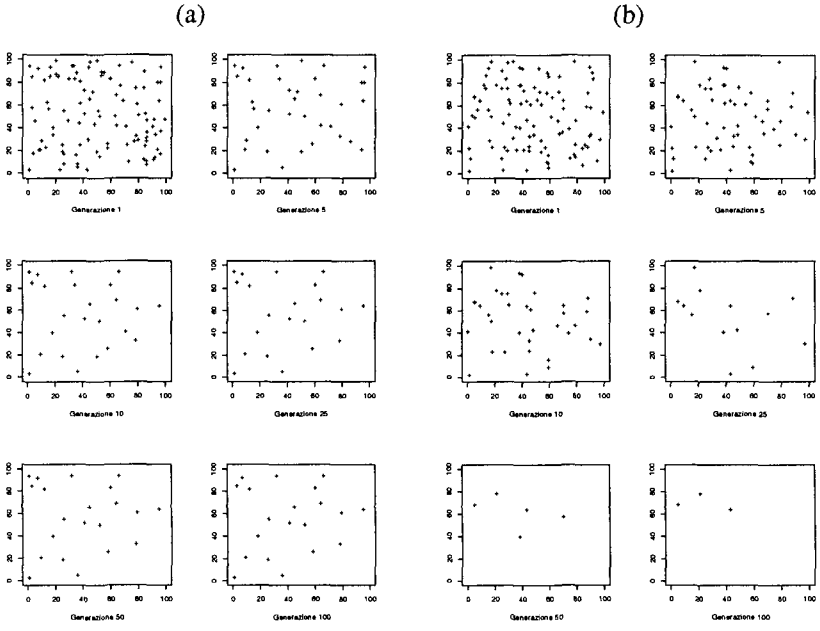


Figure 2. Simulation model with (objects,sounds) = $(5, 5)$ (a) $k = 1$ (b) $k = 25$. Distance influences both ϕ and Φ. Population positioned in a list.

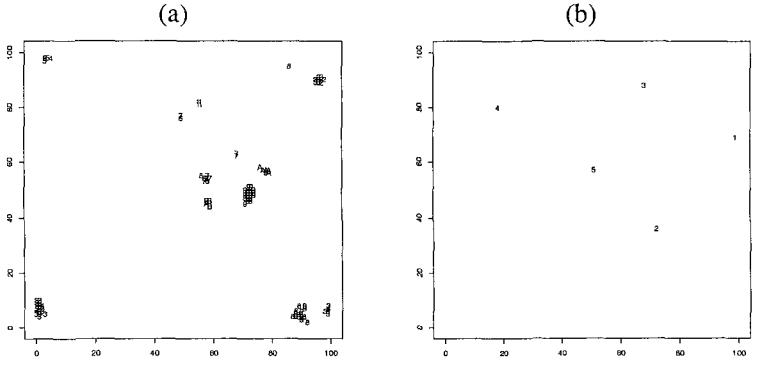


Figure 3. Simulation model with (objects,sounds) = $(5, 5)$ and $k = 7$ (a) Neighbourhood lookup, 11 languages and 14 clusters (b) List, 5 languages and 5 clusters.

4. Conclusion

We have presented a simulation model for the evolution of languages based on the theory of evolutionary language games. The model, inspired by a mathematical

model due to the biologist Martin Nowak, adds a topological structure in which the languages live. We have then studied how clusters of different languages emerge and evolve in the world, thanks to the influence of the environment on the communication among individuals.

Our results have shown the emergence of different configurations, according to the parameters acting on the system, e.g. the influence of the environment on the offspring generation and the way that the new languages are introduced to the world.

There are a number of interesting future directions we would like to explore:

- allow multiple parents and overlapping generations (population size no longer constant);
- separate the individual from the language, allowing an individual to speak more than just one language;
- study other linguistic phenomena such as dialects or pidgin/creole languages;
- expand our model to let agents move around, much like in a particle swarm system.

References

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: from natural to artificial systems.* Oxford University Press.

Cavalli-Sforza, L. (2000). *People, genes and languages.* University of California Press.

Di Chio, C. (2004). *Modelli di simulazione evolutiva per lo sviluppo del linguaggio.* (Tesi di Laurea, University of Roma "La Sapienza")

Di Chio, C., & Di Chio, P. (2005). *A simple simulation model for the evolution of language.* (Under revision)

Kennedy, J., & Eberhart, R. C. (2001). *Swarm intelligence.* Morgan Kaufmann.

Maynard Smith, J. (1982). *Evolution and the theory of games.* Cambridge University Press.

Nowak, M. (2000). Evolutionary biology of language. *Philosophical Transactions: Biological Sciences, 355,* 1615-1622.

Nowak, M., & Krakauer, D. (1999). The evolution of language. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 8028-8033.

Nowak, M., Plotkin, J., & Krakauer, D. (1999). The evolutionary language game. *Journal of Theoretical Biology, 200,* 147-162.

Swarm Development Group. (2000). *Brief overview of swarm.*

Wittgenstein, L. (1953). *Philosophical investigation.* Blackwell.

MOSTLY OUT OF AFRICA, BUT WHAT DID THE OTHERS HAVE TO SAY?

DAN DEDIU
Language Evolution and Computation Research Unit,
University of Edinburgh, 40 George Square,
Edinburgh, EH8 9LL, Scotland, UK
D.Dediu@sms.ed.ac.uk

The Recent Out-of-Africa human evolutionary model seems to be generally accepted. This impression is very prevalent outside palaeoanthropological circles (including studies of language evolution), but proves to be unwarranted. This paper offers a short review of the main challenges facing ROA and concludes that alternative models based on the concept of metapopulation must be also considered. The implications of such a model for language evolution and diversity are briefly reviewed.

1. Introduction

As is very well known, the modern human origins debate is now definitely closed and the general consensus is that the Recent Out of Africa model (Stringer & Andrews, 1988) explains perfectly well the genetic, palaeoanthropological and archaeological patterns observed. So, a fairly recent (around 200,000 years ago) and localized (a single population in (East) Africa) origin of modern humans followed by global expansion and replacement explains everything... But, is it really so?

2. The evidence

The issue of modern human origins is very important, profoundly influencing the range of explanations for the emergence, maintenance and evolution of language and the interactions between population genetic and linguistic structures. The impression outside the palaeoanthropological circles, is that the Recent Out-of-Africa model (henceforth ROA) is true, perception usually reinforced through the popularization press. In fact, there *is* a debate going on and the matters are *very far* from being settled.

I have selected the most recent papers (post 01.2000 but also a few earlier very important ones), dealing with cases where the ROA model does not fit or fits equally well as the alternative models. The search was not exhaustive and the further selection for inclusion in the review was rather strict, but still, the count

is quite large for a "closed" debate. This is the list of the main such points:

The transition to modern *Homo sapiens* was not sudden: the appearance of modern humans is sometimes clad as a heroic myth (McBrearty & Brooks, 2000), as a sudden transition, as a revolution. But there wasn't any such revolution (McBrearty & Brooks, 2000), neither morphologically, nor behaviorally, instead a mosaic of independent transitions to skeletal and behavioral modernity took place in Africa.

The modern humans originated from a structured population: the X chromosome disprove a single panmictic population, favoring models which "incorporate admixture between divergent African branches of the genus *Homo*" (Garrigan *et al.*, 2005a; Harris & Hey, 1999; Harding & McVean, 2004).

Some genes have very deep, non-African branches: the *RRM2P4* pseudogene has a MRC of ~2 MYA in East Asia (Garrigan *et al.*, 2005b), suggesting introgression from archaic local humans. The *dystrophin* gene presents a haplotype predating the ROA expansion and virtually absent from Africa. It might have left Africa earlier and introgressed later (Ziętkiewicz *et al.*, 2003). A noncoding region of the X chromosome (Xq21.1-21.33) shows a variant possibly arisen in Eurasia > 140 KYA (Yu, Fu & Li, 2002). Templeton (2002), applying nested clade analysis, finds a pattern of interbreeding between expanding and local populations.

Regional morphological continuity: one of the oldest claims against ROA-type models (Weidenreich, 1947). Wolpoff *et al.*, (2001) analyzed transitional cranial forms in two peripheral regions (Australia and Czech Republic) and concluded that they have dual ancestry. Wu (2004) concludes evolutionary continuity in China between *sapiens* and *erectus*. Demeter, Manni & Coppens (2003) supports regional continuity in the Far East with a morphometric analysis of 45 fossil crania. The most ancient European modern (Romania) presents a "mosaic of archaic, early modern human and possibly Neandertal morphological features" (Trinkaus *et al.*, 2003). The most well-known such case is the Abrigo do Lagar Velho infantile skeleton (Duarte *et al.*, 1999), showing a mixture of modern and Neanderthal morphological characters (Duarte *et al.*, 1999; Trinkaus & Zilhão, 2003), still accepted despite the critics. Given the burial context, the child was considered as a full community member.

There is also a series of arguments usually considered to support ROA, but which turn out not to be decisive:

Ancient Neanderthal *mtDNA* proves them a different species: the conclusion from extraction studies (Krings *et al.*, 1997; Lalueza-Fox *et al.*, 2005; Krings *et al.*, 2000; Ovchinnikov *et al.*, 2000) is that Neanderthal *mtDNA* is different from modern, seemingly supporting a replacement model. **But** Gutiérrez, Sánchez & Marín (2002) show ancient *mtDNA* is very sensitive to phylogenetic methods, diagenetic modifications have altered the sequences, and conclude that Neaderthal and modern *mtDNA* may overlap. Nordborg (1998) probabilistically proved that any single locus cannot resolve between replacement and admixture, being necessary to consider many loci in parallel (Wall (2000) suggests 50-100). *mtDNA* was extracted from a fossil modern gracile Australian *Homo sapiens* (Adcock *et al.*, 2001) and proved outside the modern pool. Later, the finds (LM3) were redated to 40±2 KYA (Bowler *et al.*, 2003) and the methodology contested (Cooper *et al.*, 2001), without denying that *mtDNA* lineages can be decoupled from other parts of the genome (Relethford, 2001a).

Based on living primates, the hominid clade was speciose: contested by Hunt (2003), who argues that if appropriate models are considered (the great apes), the hominin lineage may be seen "as a single, phenotypically diverse, reticulately evolving species" (Hunt, 2003).

Neanderthal morphology separates them from moderns: Harvati, Frost & McNulty (2004) used 3D primate craniofacial models and concluded Neanderthals and moderns to be separate species, but Ahern, Hawks & Lee (2005) considered this approach not capable of distinguishing between same or different species. Morphological differences could be due to non-genetic factors (Bogin & Rios, 2003): rapid dramatic morphological changes in modern Mayans accompanies migration to the USA, cautioning against morphological differences in fossil humans as diagnostic for species.

Genetic structure of living populations shows greater diversity in Africa and an African origin of human genes: generally, Africa harbors the greatest genetic diversity of living humans and most gene trees coalesce there (Jobling *et al.*, 2004) but this pattern is not true at least for the X chromosome. The greater genetic diversity of Africa can be explained by a greater long-term population size (Relethford, 2001b), also accommodating the majority coalescence (Takahata, Lee & Satta, 2001).

Modern humans are genetically very uniform: not precluding geographical differentiation (Bamshad *et al.*, 2003) and is usually considered the effect of a major population bottleneck, either a speciation or a migration/founder effect (Jobling *et al.*, 2004) or both. But this can be interpreted as a metapopulation

evolutionary history (Relethford, 2001b; Templeton, 2002; Harding & McVean, 2004; Eswaran, 2002), accommodating the small effective population size (Rousset, 2003) with a large enough adult population. Yu *et al.* (2003) shows the chimpanzees genetic diversity to have been overestimated.

There are some other arguments, like the **relative abundance of hybrids in primates** (Jolly, 2002), suggesting ubiquitous admixture in humans or the **unexpected diversity of our genus**, highlighted by the recent discovery of *Homo floresiensis* (Brown *et al.*, 2004), also pointing to advanced cognitive and technological capacities of *Homo erectus*, allowing him to cross Wallace's line.

3. The suggested class of alternative models

The data presented above (and more not included) suggests that an alternative class of models should be considered, but choosing it demands awareness to the influence of certain non-scientific factors, like political/moral (Wolpoff & Caspari, 1997), personality clashes/ambitions (Jobling *et al.*, 2004) and favored source (genetic, archaeological, fossil).

Generally, a polarity is described between the ROA model and *multiregionalsim* (Wolpoff & Caspari, 1997; Relethford, 2001b; Lewin, 1998; Jobling *et al.*, 2004), but, (Relethford, 2001b), there are *two* distinct dimensions: the *mode of transition* between archaic and modern humans and the *location and timing of this transition.* Our analysis suggests a recent African origin, a structured ancestral population (metapopulation), a mosaic/accretion of independent traits (morphological and behavioral/cultural) and is disfavoring a speciation event. It suggests a reticulate evolution, where constant gene flow between demes insures local adaptation and continuity while spreading globally the modern genetic-cultural complex. These seem to be satisfied by various models proposed (for example, Relethford, 2001b, Eswaran, 2002 and especially Templeton, 2002), but for our purposes, the following main points are relevant:

- no abrupt speciation event separating moderns from archaics;
- culturally, an accretionary evolution and not a sharp revolution;
- admixture between the migrating waves and locally adapted and differentiated archaics, insuring various degrees of regional continuity;
- metapopulational evolutionary model, whereby demes are constantly created, replaced and extinguished, maintaining genetic and cultural flows,

such that there is a global evolutionary accretion of genes and cultural traits without a "core" source population of the full package, Africa being demographically dominant.

4. Conclusions: implications for language evolution and diversity

Opposed to ROA, such a model can accommodate the language capacity as a mosaic of independent traits evolved in different demes. Language has a more or less specific genetic component, (Stromswold, 2001), confirmed by the *FOXP2* gene (Enard *et al.*, 2002) and seemingly supported by Williams syndrome (Bellugi, Korenberg & Klima, 2001). It is conceivable, for example, that the human-specific *FOXP2* mutations arose in different demes at different times and coalesced with the qualitatively different languages they allowed. The discovery (Mekel-Bobrov *et al.*, 2005; Evans *et al.*, 2005) of recent variants of two genes related to brain growth and development, with signatures of strong positive natural selection, not yet fixated and with marked population structures supports this mosaic evolutionary process.

There could exist minor inter-populational genetic differences in linguistic capacity (because of regional continuity, founder effect or not yet fixated advantageous alleles), offering new perspectives on language evolution, given that the basic requirement is heritable variation. Such a model highlights the early evolution of the language capacity and languages as two inter-related phenomena in metapopulations, leading to the modern linguistic capacity, able to support an immense linguistic (almost neutral) variation.

Another possibility is that besides the accidental correlations between genes and languages (Cavalli-Sforza *et al.*, 1994), there might also exist a slight non-accidental correlation, whereby specific genetic configurations favor/are favored by specific linguistic features. A fictional example could be a population with a high incidence of articulatory incapacity to produce a trilled /r/, which in turn will select for languages realizing the phoneme /r/ as an approximant. Conversely, speakers with such a deficiency will not incur any fitness penalty when immersed into a community speaking the /r/-approximant language. This hypothetical example can be extended to more plausible cases, like the better control of rapid orofacial movements (supposedly) brought by the human-specific mutation(s) in *FOXP2*.

Acknowledgements

My work has been supported by an ORS Awards Grant 2003014001 and a CHSS Studentship, The University of Edinburgh.

References

Adcock, G. J., Dennis, E. S., Easteal, S., Huttley, G. A., Jermiin, L. S., Peacock, W. J. & Thorne, A. (2001). Mitochondrial DNA sequences in ancient Australians. *PNAS*, 98, 537-542.

Ahern, J. C. M., Hawks, J. D. & Lee, S.-H. (2005). Neandertal taxonomy reconsidered... again. *Journal of Human Evolution*, 48, 647-652.

Bamshad, M. J., Wooding, S., Watkins, W. S., Ostler, C. T., Batzer, M. A.& Jorde, L. B. (2003). Human population genetic structure and inference of group membership. *American Journal of Human Genetics*, 72, 578-589.

Bellugi, U., Korenberg, J. R. & Klima, E. S. (2001). Williams syndrome. *Clinical Neuroscience Research*, 1, 217-229.

Bogin, B. & Rios, L. (2003). Rapid morphological change in living humans. *Comparative Biochemistry and Physiology Part A*, 136, 71-84.

Bowler, J. M., Johnston, H., Olley, J. M., Prescott, J. R., Roberts, R. G., Shawcross, W. & Spooner, N. A. (2003). New ages for human ocuppation and climatic change at Lake Mungo, Australia. *Nature*, 421, 837-840.

Brown, P., Sutikna, T., Morwood, M. J., Soejono, Jatmiko, Saptomo, E. W. & Due, R. A. (2004). A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature*, 431, 1055-1061

Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton: Princeton University Press.

Cooper, A., Rambaut, A., Macaulay, V., Willerslev, E., Hansen, A. J. & Stringer, C. (2001). Human origins and ancient human DNA. *Science*, 292, 1655-1656.

Demeter, F., Manni, F. & Coppens, Y. (2003). Late Upper Pleistocene human peopling of the Far East. *Comptes Rendus Palevol*, 2, 625-638.

Duarte, C., Maurício, J., Pettitt, P. B., Souto, P., Trinkaus, E., van der Plicht, H. & Zilhão, J. (1999). The early Upper Paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia. *PNAS*, 96, 7604-7609.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, *et al.* (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418, 869-872.

Eswaran, V. (2002). A diffusion wave out of Africa. *Current Anthropology*, 43, 749-774.

Evans, P. D., Gilbert, S. L., Mekel-Bobrov, N., Vallender, E. J., Anderson, Vaez-Azizi, L. M., *et al.* (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science*, 309, 1717-1720.

Garrigan, D., Mobasher, Z., Kingan, S. B., Wilder, J. A. & Hammer, M. F.

(2005a). Deep haplotype devrgence and long-range linkage disequilibrium at Xp21.1 provide evidence that humans descend from a structured ancestral population. *Genetics*, 170, 1849-1856.

Garrigan, D., Mobasher, Z., Severson, T., Wilder, J. A. & Hammer, M. F. (2005b). Evidence for archaic Asian ancestry on the human X chromosome. *Molecular Biology and Evolution*, 22, 189-192.

Gutiérrez, G., Sánchez, D. & Marín, A. (2002). A reanalysis of the ancient mitochondrial DNA sequences recovered from Neanderthal bones. *Molecular Biology and Evolution*, 19, 1359-1366.

Harding, R. M. & McVean, G. (2004). A structured ancestral population for the evolution of modern humans. *Current Opinion in Genetics & Development*, 14, 667-674.

Harris, E. E. & Hey, J. (1999). X chromosome evidence for ancient human histories. *PNAS*, 96, 3320-3324.

Harvati, K., Frost, S. R. & McNulty, K. P. (2004). Neanderthal taxonomy reconsidered: implications of 3D primate models of intra- and interspecific differences. *PNAS*, 101, 1147-1152.

Hunt, K. D. (2003). The Single Species Hypothesis. *Human Biology*, 75, 485-502.

Jobling, M. A., Hurles, M. E. & Tyler-Smith, C. (2004). *Human evolutionary genetics: origins, peoples and disease*. Garland Science.

Jolly, C. J. (2002). A proper study for mankind: analogies from the papionin monkeys and their implications for human evolution. *American Journal of Physical Anthropology*, 116, 177-204.

Krings, M., Capelli, C., Tschentcher, F., Geisert, H., Meyer, S., von Haeseler, *et al.* (2000). A view of Neandertal genetic diversity. *Nature Genetics*, 26, 144-146.

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking. M. & Pääbo, S. (1997). Neandertal DNA sequences and the origin of modern humans. *Cell*, 90, 19-30.

Lalueza-Fox, C., Lourdes Sampietro, M., Caramelli, D., Puder, Y., Lari. M., Calafell, *et al.* (2005). Neandertal evolutionary genetics. *Molecular Biology and Evolution*, 22, 1077-1081.

Lewin, R. (1998). *Principles of Human Evolution*. Blackwell Science.

Long, J. C. & Kittles, R. A. (2003). Human genetic diversity and the nonexistence of biological rances. *Human Biology*, 75, 449-471.

McBrearty, S., & Brooks, A. S. (2000). The revolution that wasn't. *Journal of Human Evolution*, 39, 453-563.

Mekel-Bobrov, N., Gilbert, S. L., Evans, P. D., Vallender, E. J., Anderson, J., R., Hudson, *et al.* (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. *Science*, 309, 1720-1722.

Nordborg, M.. (1998). On the probability of Neanderthal ancestry. *American Journal of Human Genetics*, 63, 1237-1240.

Ovchinnikov, I. V., Götherström, A., Romanova, G. P., Kharitonov, V. M., Lidén. K. & Goodwin, W. (2000). Molecular analysis of Neanderthal DNA

from the northern Caucasus. *Nature*, 404, 490-493.
Relethford, J. H. (2001a). Ancient DNA and the origins of modern humans. *PNAS*, 98, 390-391.
Relethford, J. H. (2001b). *Genetics and the search for modern human origins*. Wiley-Liss.
Rousset, F. (2003). Effective size in simple metapopulation models. *Heredity*, 91, 107-111.
Stringer, C. & Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science*, 239, 1263-1268.
Stromswold, K. (2001). The heritability of language. *Language*, 77, 647-723.
Takahata, N., Lee, S.-H. & Satta, Y. (2001). Testing multiregionality of modern human origins. *Molecular Biology and Evolution*, 18, 172-183.
Templeton, A. R. (2002). Out of Africa again and again. *Nature*, 416, 45-51.
Trinkaus, E. & Zilhão, J. (2003). Phylogenetic implications. In J. Zilhão and E. Trinkaus (Eds.), *Portrait of the Artist As a Child.* (pp. 497-518). Oxbow Books Ltd.
Trinkaus, E., Moldovan, O., Milota, Ş., Bîlgăr, A., Sarcina, L., Athreya, S., *et al.* (2003). An early modern human from the Peştera cu Oase, Romania. *PNAS*, 100, 11231-11236.
Wall, J. D. (2000). Detecting ancient admixture in humans using sequence polymorphism data. *Genetics*, 154, 1271-1279.
Weidenreich, F. (1947). Facts and speculations concerning the origin of Homo sapiens. *American Anthropologist*, 49, 187-203.
Wolpoff, M. H. & Caspari, R. (1997). *Race and human evolution*. Westview Press.
Wolpoff, M. H., Hawks, J., Frayer, D. W. & Hunley, K. (2001). Modern human ancestry at the peripheries. *Science*, 291, 293-297.
Wu, X. (2004). On the origin of modern humans in China. *Quaternary International*, 117, 131-140.
Yu, N., Fu, Y.-X.& Li, W.-H. (2002). DNA polymorphism in a worldwide sample of human X chromosomes. *Molecular Biology and Evolution*, 19, 2131-2141.
Yu, N., Jenesen-Seaman, M. I., Chemnick, L., Kidd, J. R., Deinard, A. S., Ryder, O., Kidd, *et al.* (2003). Low nucleotide diversity in chimpanzees and bonobos. *Genetics*, 164, 1511-1518.
Ziętkiewicz, E., Yotova, V., Gehl, D., Wambach, T., Arrieta, I., Batzer, M., *et al.* (2003). Haplotypes in the Dystrophin DNA segment point to a mosaic origin of modern human diversity. *American Journal of Human Genetics*, 73, 994-1015.

A COMPARISON OF THE ARTICULATORY PARAMETERS INVOLVED IN THE PRODUCTION OF SOUND OF BONOBOS AND MODERN HUMANS

DIDIER DEMOLIN

Department of Linguistics, Universidade de Sao Paulo, Brasil and Université libre de Bruxelles, Belgium

VÉRONIQUE DELVAUX

FNRS and Université de Mons-Hainaut, Belgium

Most studies of vocalizations with chimpanzees and Bonobos focus on the interpretation of the vocal behaviour of both captive and free-ranging groups to relate sounds produced to their semantic contexts. Spectrographic analyses reveal the acoustic structure of the vocalizations but rarely raise the question of the specific articulatory capacities of Bonobos in relation to the acoustics. This point is essential if one wants to understand the articulatory control that Bonobos have on their vocalizations. It is also important when the vocalizations of Bonobos and the sound produced by modern humans are compared.

1. Introduction

Fitch (2000) in a major paper reviewing the evolution of speech in a comparative perspective shows that two changes were necessary prerequisites for modern human speech abilities: the modification of vocal morphology and the development of vocal imitative ability. Focusing on the first point, one can wonder why do our closest animal relatives, like Bonobos, lack vocal output capabilities comparable with ours. Comparative data from living nonhuman primates will likely shed light on this question and will allow inferring the functions of vocalizations in early hominids. As stated by Fitch (2000: 260), the evolution of speech was closely tied to mechanisms of sound production and the most obvious speech-related difference between humans and other mammals concerns the structure of the human vocal tract. A central point is bout the fact that the human vocal tract anatomy differs from other primates. The human larynx rests much lower in the throat that in apes and human lack laryngeal air sacks. In addition, human speech requires sophisticated nervous control. Therefore as the relation between acoustic and articulatory aspects of human speech are better understood than that of living apes and particularly of Bonobos, it is worth evaluating the vocalizations of the latter from the known principles of the acoustics of human speech. For example taking the perspective

of the source-filter theory (Fant 1960) it is possible to discuss the nature of the source of vocalizations and to make inferences about how the latter is filtered in the vocal tract.

2. Material and method

2.1 Articulatory data

Articulatory data come from dissections, MRI data and 3D scans made on dead specimens from the Zoo of Planckendael in Belgium. Specimens were fixed for 6 months before being dissected. Our data come from 1 male, 1 female and 1 juvenile. 3D scan data come from the male subject and MRI data from the female and juvenile subject. These data allow an accurate description of the size and shape of the Bonobo's vocal tract. From an articulatory viewpoint, the point is to establish the position and the shape of the lips, the jaw, the velum, and the tongue. Two other parameters are considered: (i) the morphology of the larynx and of the vocal folds (ii)) the shape and the dimension of the vocal tract.

2.2 Acoustic and video data

2.2.1 Method

Three recording sessions (of approximately 1h30' each) were made with two males, two juveniles and several females, using a digital video camera with an incorporated microphone. These sessions were made at the Zoo of Plankendael in Belgium. One session was made outdoors and two indoors, out of which 45 sequences of vocalizations have been extracted to constitute the corpus for the present study. Each sequence may include several types of vocalizations (e.g. one staccato hooting, then 4 full screams, etc.) produced by different individuals. But we have excluded from the corpus those sequences in which several individuals vocalize together at the same time (mixed vocalizations). In about one third of the sequences approximately) a Bonobo is filmed by the camera while vocalizing (face or profile), i.e. by recording the sound and the corresponding image.

2.2.2 Repertoire

In order to compare our data with previous work on Bonobo's vocalizations, they were organized within the 16 categories (and subcategories) of De Waal (1988). Our data reveal several occurrences of 11 out of these 16 categories: low hooting, staccato hooting, legato hooting, wieew bark, whistle bark, (food) peep, peep yelp, peep scream, rasp scream, bark scream, full scream. What is missing in our data is: contest hooting (in De Waal 1988, contest hooting was uttered by one single individual); greeting grunt and panting laugh (this is aperiodic sounds of very low intensity hard to extract from ambient noise); pout moans (in De Waal 1988 pout moans were uttered by adolescent males only and we have two adults and two juvenile males but no adolescent in our group); alarm peeps

(there has been no opportunity to raise the alarm during our three recording sessions. Some vocalizations in our data do not match any of the categories from De Waal: (i) two kinds of soft sounds that we would define as: "soft bark" and "soft peep", (ii) the "piano", a sequence of non modulated peeps varying in intensity and pitch.
Data were processed with '*MovieSignalExplorer*' (Delvaux & Soquet), that is a customized application that allows (i) to analyze the acoustics of the vocalizations using spectrograms, spectra, energy analyzer and pitch analyzer; (ii) to display every (30 ms long) portion of the acoustic signal together with the corresponding digital picture recorded by the video camera.

3. Results

3.1 Articulatory data

The main features that can be observed from the articulatory measurements are the already known fact that the vocal tract shape of Bonobos is quite different from that of modern humans. The larynx is higher and the tongue is longer and narrower. However there is another important feature to be observed, that is the long flat and narrow palate and the smaller velum.
The larynx of Bonobos shows structures quite similar to that of modern humans with the important difference that the epiglottis seems to be less vertical than in humans as it can be seen in Figure 1. The V shape of the hyoid bone is also different when compared to the U shape of modern humans.

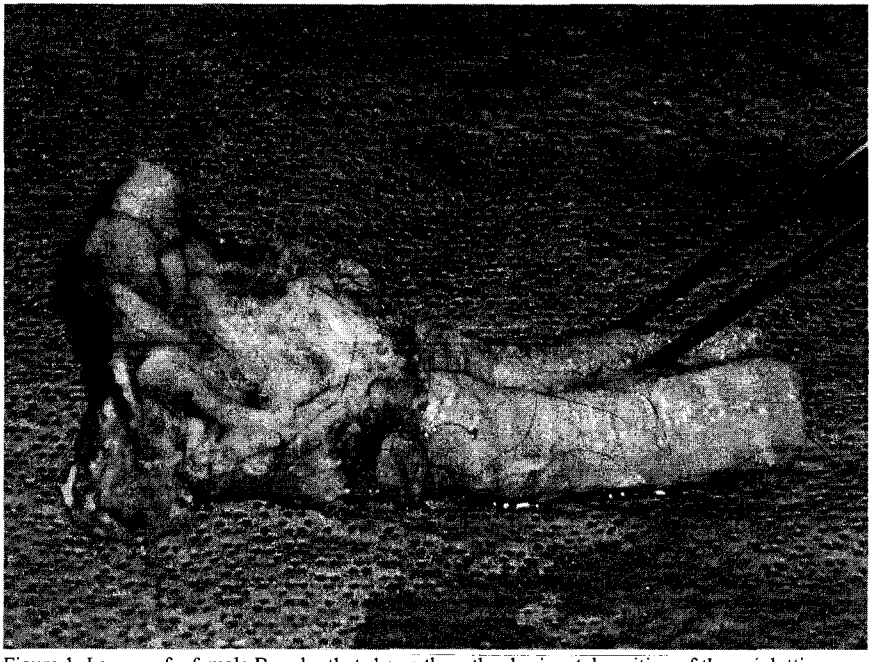


Figure 1. Larynx of a female Bonobo that shows the rather horizontal position of the epiglottis.

3.2 Acoustic data

Acoustic data show that Bonobos seem to modulate the vocal output on the temporal and spectral levels. Modulations of the fundamental frequency accompany variations of intensity. Acoustically it also appears that a second sound source (diplophonia) can frequently be observed in Bonobo's vocalizations. This second source is related to the resonances (formants) of the vocal tract that seem to have a uniform shape most of the time. Noticeable perturbations of the vocal folds vibrations are also frequently observed. Modulations of the harmonics and of the resonances are associated with modulations of the intensity of the sound source. Most of these features can be observed on Figure 2 and 3.

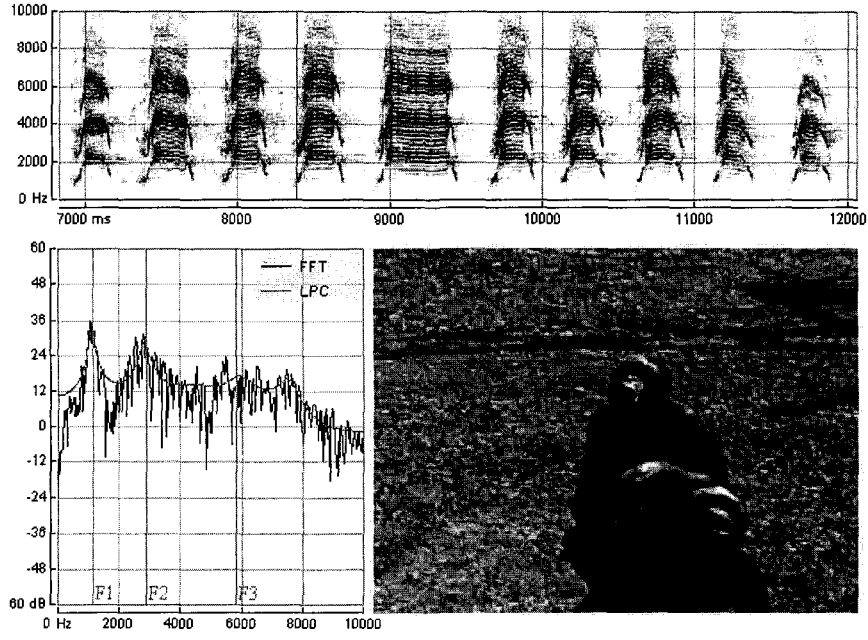


Figure 2. Spectrogram, FFT and LPC spectra of a bark scream produced by the Bonobo staying at the back of the picture. The spectrogram suggests that the formants and the set of harmonics are produced by two different sources. The barks are also modulated in duration.

4. Discussion

From the perspective of the source filter theory and comparing the relation between acoustic and articulatory data there are some striking observations that confirm and complement the observations made in a number of previous studies (e.g. De Waal 1988, Fitch and Hauser 1995, Owren and Rendall 1997 and Bermejo and Omedes 1998).

The source of vocalizations seems to be much more diverse than usually observed in modern humans as noted by Hauser (1996). The vocal fold vibrations that produce a quasi periodic sound and that consist in a fundamental frequency and a set of harmonics is present as a source. However when diplophonia is observed the origin of the second source must also be explained. Schön Ybarra (1995) and Mergell et al. (1999) have suggested that the thin upward extension of the vocal fold that is present in many primate species might serve as independent low-mass oscillators to support the quite high frequencies produced in vocalizations and that these vocal membranes underlie vibration at multiple frequencies, thus giving the observed diplophonia. For the moment, it is difficult to confirm whether this is true or not for the vocalizations of Bonobos. However we would like to draw the attention to the fact that various other

sources can be at play. The sphincteric phonation that implies the vibration of the arytenoids cartilages against the tubercle of the epiglottis can act a phonatory mechanism as well (Traill 1986). Esophageal phonation could also act as source in some vocalizations (Traill 1980). Another possible source is the epiglottis. Indeed as observed in the anatomical data, the epiglottis has a much less vertical position when it is compared to man. Even if it is separated from the root of the tongue in the process of a dissection, this position is striking. The length of the epiglottis that covers an area greater than the larynx itself is also to be noticed. This position could allow the epiglottis to act in a way similar to the opening of a flute to produce the high pitched whistling observed in Bonobo's vocalizations.

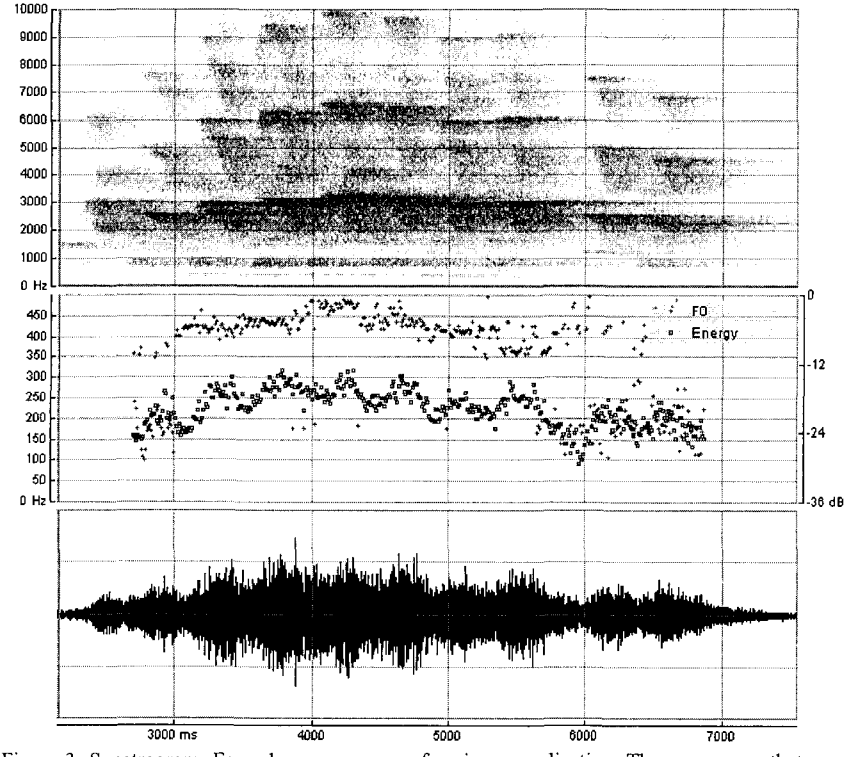


Figure 3. Spectrogram, Fo and energy curves of a piano vocalization. The resonances that can be observed on the spectrogram are accompanied by noise. Modulations of intensity (energy curve) correspond to modulations on the Fo curve.

Finally it has to be noted that if Fo variations are observed they are never independent of intensity. This suggests that there is a close association between Fo, intensity and subglottal pressure in Bonobo's vocalizations. This feature is not true for modern humans contrary to frequent claims (Ohala 1990, Demolin forthcoming). The fact that Fo can be controlled independently of subglottal

pressure is even a major evolutionary feature of human speech when compared to the vocalizations of primates.

When the filter properties are considered there are several important features to be noticed. The first is that no vowels and consonants are observed during Bonobos vocalizations. When formants are observed in primate vocalizations they almost always suggest that they are produced in a uniform acoustic tube, a feature that is very different of the vast majority of vowels produced by humans. Fitch and Hauser (1995) emphasize the importance of formants in the vocalization of primates and show that the low position of the adult human larynx gives the capacity to produce sounds that have different formant patterns easily. They conclude that the descent of the larynx was a key innovation in the evolution of speech. Fitch (2000: 260) also claims that the change in the larynx position greatly expands the phonetic repertoire of humans because the tongue can move both vertically and horizontally within the vocal tract after this larynx lowering. This is obviously true but it is only part of the explanation. In order for the tongue to move vertically and horizontally, the oral cavity must also be reshaped in these dimensions. This feature and the reshaping of the tongue itself are the consequences of the reduction of prognatism. This has been a major feature in the evolution of hominids and had important consequences on the production of speech because it allows the tongue to make closures and constrictions in the front part of the vocal tract where the vast majority of consonants are produced. Frayer and Nicolay (2000: 231) when comparing the upper palate dimension of modern humans with various hominids and chimpanzees conclude that most specimens older than roughly 250 000 years before the present do not differ from the chimpanzees.

Finally it is has to be noted that even if prosodic aspects involving intonation and Fo contours are not present, there are modulations of duration and intensity like those of Figures 2 and 3 that suggest that these parameters might controlled in vocalizations.

Acknowledgements

We wish to thank our colleagues Linda Van Elsaker and Kris d'Aout who made the work with the Bonobos of the Zoo in Plankendael possible and who provided the specimen for dissection. Sergio Hassid and Emile Godefroid of the Hopital Erasme in Brussels made the work dissection possible.
This research was supported by a grant 'Evolution des langues, du langage et de la parole' of the National Bank of Belgium.

References

Bermejo, M. and A. Omedes (1998). Preliminary vocal repertoire and vocal communication of wild bonobos (Pan Paniscus) at Lilungu (Democratic Republic of Congo). *Folia Primatologia*, 70: 329-357.

Demolin, D. (forthcoming). Phonological universals and the control and regulation of speech production. In Experimental approaches to Phonology. P. Beddor, M. Ohala and M.-J. Solé (Eds.). Oxford University Press.

De Waal, F. (1988). The communicative repertoire of captive Bonobos (Pan Paniscus), compared to that of chimpanzees. *Behaviour*, 106: 183-251.

Fant, G. (1960). *Acoustic Theory of Speech Production*. Mouton. The Hague.

Fitch, W.T. (2000). The evolution of speech a comparative perspective. *Trends in Cognitive Science*, 4 (7): 258-267.

Fitch, W.T. and M. Hauser (1995). Vocal production in nonhuman primates: acoustics, physiology and functional constraints on 'honest' advertisement. *American Journal of Primatology*, 37: 191-219.

Frayer, D. W. and C. Nicolay (2000). Fossil evidence for the origin of speech. In N. L. Wallin, B. Merker and S. Brown (eds.) *The origin of Music*. Cambridge: MIT Press.

Hauser, M. (1996) *The Evolution of Communication*. Cambridge, MA: MIT Press.

Mergell, P. , W. T. Fitch and H. Herzel (1999). Modelling the role of nonhuman vocal membranes in phonation. *Journal of the Acoustical Society of America*, 105 (3) : 2020-2027.

Ohala, J.J. (1990). Respiratory activity in speech. In W.J. Hardcastle and A. Marcahl (Eds.) Speech production and speech modeling. Kluwer Academic Publishers. 23-53.

Owren, M. J. and D. Rendall (1997). An affect-conditioning model of nonhuman primate vocal signaling. *Perpectives in Ethology*, 12: 299-346.

Schön Ybarra, M. (1995). A comparative approach to the nonhuman primate vocal tract: Implications for sound production. In E. Zimmermann, J.D. Newman & U. Jürgens (eds.), *Current topics in primate vocal communication*, New-York. Plenum Press: 185-198.

Traill, A. (1980). The Laryngeal sphincter as a phonatory mechanism in !xoo Bushman. In *Variation, Culture and Evolution in African Populations: papers in Honor of Dr Hertha de Villiers*. Johannesburg. Witwatersrand University Press: 123-131.

Traill, A. (1986). A note on pitch control in esophageal speech. *Die Suid-Afrikaanse Tydskrif vir Kommunikasieafwykings*, Vol 27: 95-98.

GENERALISED SIGNALLING: A POSSIBLE SOLUTION TO THE PARADOX OF LANGUAGE

JEAN-LOUIS DESSALLES
ENST (ParisTech)
46 rue Barrault
Paris, F-75013, France

The systematic and universal communicative behaviour that drives human beings to give honest information to conspecifics during long-lasting conversational episodes still represents a Darwinian paradox. Attempts to solve it by comparing conversation with a mere reciprocal cooperative information exchange is at odds with the reality of spontaneous language use. The Costly Signalling Theory has recently attracted attention as a tentative explanation of the evolutionary stability of language. Unfortunately, it makes the wrong prediction that only elite individuals would talk. I show that as far as social bonding is assortative in our species, generalised signalling through language becomes a viable strategy to attract allies.

1. Introduction: coping with the evolutionary paradox

Language is an idiosyncratic property of the human species. Though, since the reasons for its very existence have been for long considered beyond question, the only remaining issues concerning its origin are to find out when, where and how it emerged in our lineage. Even recently, some authors insisted on the "obvious" necessity of language.

> " It is possible to imagine a superintelligent species whose isolated members cleverly negotiated their environment without communicating with one another, but what a waste! There is a fantastic payoff in trading hard-won knowledge with kin and friends, and language is obviously a major means of doing so " (Pinker, 1994:367.)

> " The adaptive significance of human language is obvious. It pays to talk. Cooperation in hunting, making plans, coordinating activities, task sharing, social bonding, manipulation and deception all benefit from an increase in expressive power. " (Nowak & Komarova, 2001)

Unfortunately, there is no way to prove the existence of a general selection pressure towards language that only our lineage would have been responsive to. Moreover, no satisfactory explanation is currently available to account for the fact that individuals would find a selective advantage in giving away hard-won knowledge about their physical and social environment for free.

Traditional intuitions about evolution based on the species' welfare or on group selection have been dismissed for a long time (Williams, 1966), and even if attempts have been made to revive them as potentially sound (Sober & Wilson, 1998), it remains highly unlikely that such mechanisms produce any selection pressure[a] (Williams, 1966:112). The only traditional alternative left to explain altruistic talking to unrelated conspecifics is reciprocal cooperation (Trivers, 1971; Axelrod & Hamilton, 1981). The conditions for the stable existence of cooperation are well known (Dessalles, 1999): good benefit-to-cost ratio and good cheater detection. Spontaneous language is far from meeting those requirements. The informational content of most everyday chatter is unlikely to significantly change the hearers' life expectancy, and the cost of talking should not be underestimated, if only for the time devoted to it (Dunbar, 1998). But the second constraint is even more patently infringed. Cheating, in the case of language, means that the addressee takes the information given and does not give information in return, even with delay. At first sight, the dialogic nature of many conversations seems to support the cooperative view. A closer examination of spontaneous exchanges reveals a quite different picture. To avoid cheating, speakers should rely on good assessment and memory of who gave valuable information to whom. Such bookkeeping is compatible with conspiratorial whispering, not with open conversation. Observation of spontaneous exchanges reveals that speakers talk to more than two addressees on average (Dunbar, Duncan & Nettle 1995). If reciprocation were the underlying mechanism of language, talkative behaviour would have been selected against. There are more loquacious individuals around who are in search of an audience than people withholding valuable information and waiting for being prompted by avid hearers. Moreover, cheating detection seems to be performed by hearers rather than by speakers, as the former systematically point to any inconsistencies they perceive in what they hear (Dessalles, 1998). Conversation is more like scientific communication, in which authors strive to make a point that will be critically appraised by their colleagues, rather than like communication among spies or speculators, where information is exchanged on a strictly reciprocal basis.[b]

---

[a] Elementary calculus shows that behaviour in favour of the collective with no return for the individual may have a slight positive effect on the frequency of that behaviour, but this effect requires strong heterogeneity and low migration within the population; it is frequency-dependent and vanishes when the behaviour becomes predominant; it presupposes the absence of any alternative behaviour that would be selfish and profitable. Simulations show that the effect does not emerge from noise and that it disappears when individuals endure event small costs.

[b] Cooperation can be made more robust by introducing intense punishment, but punishment exerted by speakers is absent from spontaneous language. Correlations between players are also known to stabilise cooperation, but they amount to introducing some dose of kin selection into the system.

2. Costly signalling and human language

A new paradigm to explain non-kin altruism recently emerged from the pioneering work of Amotz Zahavi (Zahavi & Zahavi, 1997). Altruistic behaviour may be performed in the absence of any reciprocation as a way to gain prestige in the group. The main claim of the theory is that these conducts evolved to advertise some quality. Another aspect of the theory is that competition leads to costly performance, as the handicap of the cost makes the performance a reliable indicator of the corresponding quality.

> " [...] my suggestion that altruism is an investment in advertisement by the individual altruist (that is, a handicap) shows how altruism can provide a direct benefit to the altruist. The individuals that accord prestige to the altruist do so not to encourage altruism, [... they] benefit directly from the information advertised by the altruistic act in their own decision making. " (Zahavi, 2003)

This model has been shown to be theoretically consistent (Grafen, 1990; Gintis, Smith & Bowles, 2001). As stated in Zahavi's quotation, one crucial requirement is that the individuals that witness the performance benefit from making a good appraisal and reward good performers. The two sides of this requirement may be simultaneously fulfilled if the former make an asymmetrical alliance with the latter, *e.g.* becoming their disciples or conversely accepting them as coalition mates. In Gintis *et al.* model (2001), performers benefit from being chosen and discerning coalition partners benefit from choosing high quality performers.

We may say of those models that they are *political*, as they deal with fruitful alliances among individuals. The main reason for individuals to engage into coalition formation is that other individuals do so, what may create contexts in which isolated individuals are almost certain to lose. The crucial task each individual has to perform is to choose the best available allies, judging from available clues. As the individuals' true qualities are not directly observable, the selection task must rely on signals displayed by performers, as long as those signals are reliable. What Gintis *et al.* (2001) showed is that conditions do exist in which honest signalling is stable and provides a true indication of the individuals' quality. Though, realistic conditions that meet the model's requirements are still to be discovered.

Human language is spontaneously used without specific and obligatory consideration of genetic, spatial or aspect closeness.

Transposing Costly Signalling Theory (CST) to human language is not straightforward. Talking would be a way of advertising some crucial quality for forming efficient coalitions. Several problems must be solved: (1) Which quality is advertised through linguistic performance? (2) Is that quality positively correlated with the success of coalitions? (3) How is the reliability of signals enforced? (4) Why is language a generalised behaviour, while CST predicts that only high quality individuals should show off?

In previous work, we attempted to answer problems (1)-(3) (Dessalles 1998, 2000a, 2000b). (1) By striving to make relevant utterances, individuals would demonstrate their ability to extract information from their physical and social environment. Spontaneous conversations constitute an arena in which individuals compete to show that they were first to detect unexpected events that, without them, may have passed unnoticed. (2) This informational ability is supposed to be a useful quality for large coalitions. When five coalition mates or more are to take coordinated action, the presence among them of individuals who are fully in touch with the physical and social environment is a crucial asset. Conversely, the coalition is unlikely to perform well if none of its members knows about what is going on around. Through language, individuals continuously demonstrate their ability to extract relevant information from their surroundings or from their past experience. Even if the recounted events are most often futile, it is still a good way to show off as a valuable coalition mate. (3) In CST, reliability is guaranteed by the cost of the performance. However, as Zahavi observes (Zahavi & Zahavi 1997:223), it is easy to lie with words. If performance is appraised according to the informational unexpectedness or importance of the reported events, it is tempting to report highly unlikely events by merely inventing incredible situations and pretending to have witnessed them. We proposed that logic and argumentation evolved as a way to deter liars (Dessalles 1998), and that language may be honest in the absence of direct cost, as it is difficult to lie while remaining logically consistent.

Problem (4) remains a difficulty. In CST, signalling is honest as far as it allows telling apart high quality from low quality individuals. In an open and reliable competition, low quality individuals have no interest in showing off, as their performance would precisely reveal that their quality is low. How can we explain that language is a generalised behaviour in our species? CST would let us expect that only individuals capable of announcing big scoops or of making the soundest discussions would dare to talk. This is obviously not the case.

3. The problem of the signalling threshold

CST predicts a sharp cut-off in the level of signalling. Above a definite quality value, signalling becomes profitable, even if it is costly and altruistic, as the "political" return in terms of making efficient alliances compensates for the cost. Below that quality threshold, the probability of being chosen as coalition mate becomes too low and no longer compensates for the immediate cost. In previous unpublished attempts, we tried to circumvent the difficulty by adding noise and uncertainty both in the production of signals and in the formation of alliances. Though the consequence was an actual lowering of the signalling threshold, these efforts proved unsatisfactory. When signals become less correlated with the quality in demand, "customers" turn away from them even if communication is honest (Izquierdo *et al.*, 2005). Moreover, as CST shows, good quality signallers evolve to reveal their higher quality anyway, most often by enduring high costs that lower quality individuals cannot afford.

In the case of language, even if logical testing by listeners can ensure reliability in the absence of significant marginal cost, the signalling threshold problem remains intact: we would expect that the only individuals talking in conversation would be those who were able to gather genuine extraordinary or crucial experience, while it would be advantageous for those who have only banal events to report to remain silent. Human verbal interactions would look be quite different from what they are, something like alarm calls that occur infrequently.[c]

The thesis of this paper is that the existence of language, as it is practised hours each day by virtually all healthy individuals in traditional societies (Dunbar, 1998), can be explained within the CST framework under the assumption that social bonding is assortative.

4. Language and assortative social bonding

One crucial characteristic of human social behaviour is that it produces assortative matching. Even if you cannot become acquainted with prominent individuals many would dream of (for some it may be a famous actor, a politician, an important scientist or that person next door who ignores you), you still find some interest, and presumably some biological advantage, in

[c] In some forms of signalling, the threshold may be close to zero, as is possibly the case for costly sexual mating in some singing bird species. If this is the case, it is due to the fact that males have no alternative than song to have the slightest chance for mating. This situation does not obtain in human coalition formation, as a variety of qualities such as fidelity, courage, efficiency or actual social status offer alternatives to informational abilities for showing off.

establishing friendship bonds with people more like you. The point is that assortative matching gives a chance to everyone to be chosen, and so everyone benefits from displaying her/his quality.

The crucial parameter is the difference between the maximal number of links coming to a node and the maximal number of links going out of a node in the graph of social bonds. If the former is much higher, as when many people admire an opera singer, the competition generates an elite. Don't attempt to sing opera in public if you have a faint voice. Human friendship functions differently. The maximal number of friendship bonds an individual is able to establish corresponds roughly to the number of different friendship bonds individuals would accept. A consequence is that, at equilibrium, friendship bonds are assortative, *i.e.* they bind individuals of comparable "quality" (for whatever relevant quality considered). Fig. 1 shows how pairings may stabilize to be roughly assortative during a computer simulation, as individuals of various qualities (on the horizontal axis, below) tend to bind with individuals of comparable quality (on the horizontal axis, above). The left part of the figure was obtained by imposing one-to-one pairings. The right part of the figure shows a case of non-assortative pairing.

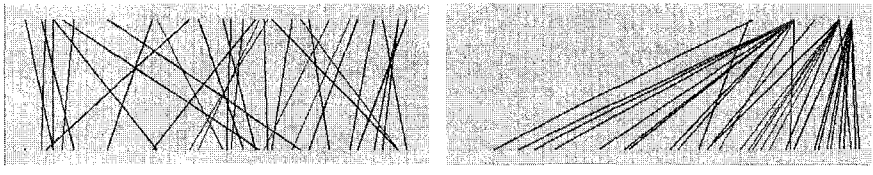


Figure 1. Examples of assortative (left) and non-assortative (right) pairings.

The political advantage for an agent of attracting allies may be written $f(q+\alpha Q)$, where q is the quality of the individual, Q the quality of the support provided by its allies, α is a constant and $f()$ an increasing function. In this simple model, individuals may reveal their quality by emitting a signal $s=gq$ that depends on their quality and on a genetically coded value g. Producing the signal involves the cost Cg, where C is a constant. Finally, the quality of the recruited allies, and thus their support, depends on s: $Q=h(s)$, where $h()$ is an increasing function. The benefit for the signalling individual is: $f(q+\alpha h(gq))-Cg$. At equilibrium, its first derivative in g is zero, and thus: $\alpha q h'(gq)\, f'(q+\alpha h(gq))=C$, which gives the optimal value of g to which evolution will lead. $h'(s)$ represents the marginal increase of the recruitment level. In the case of non-assortative paring, $h'()$ is zero or negligible below a threshold value of the signal, where it takes high values. As a consequence, if $f()$ does not change value abruptly, the

equation has no solution in g for small values of q and non-negligible cost C. This is the standard result of CST. In assortative pairing, $h()$ is expected to increase smoothly, as slightly better signals attract slightly better friends. $h'(s)$ is thus non-zero and the equation may have a solution for any value of q, which means that signalling is now generalised.

Computer simulations confirm this result. Plots on Fig. 2 show how much individuals invest in communication, depending on their quality: low (black), medium low (white), medium high (red) and high (blue). The two first plots (left and middle) were obtained for two different benefit functions $f()$. The plot in the middle shows that there are conditions in which, though communication remains honest, it is more profitable for low quality individuals to invest more in communication than higher quality individuals. The plot on the right corresponds to non-assortative links. It shows a typical CST result where only high quality individual are signalling.

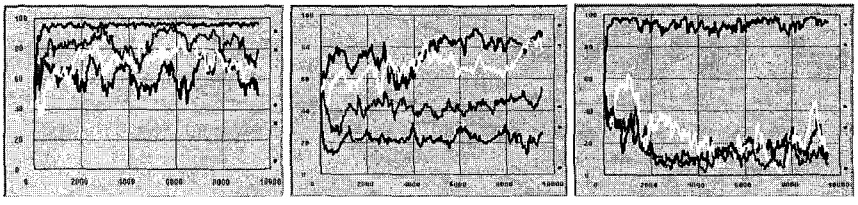


Figure 2. Three examples of evolution of investment in communication depending on quality.

5. Conclusion

The main point of this paper was to show that generalised signalling is possible within the framework of CST thanks to assortative social bonding, and that the existence of language could be a consequence of this phenomenon. The crucial difference with various forms of kin-based cooperation, in which individuals are bound to interact with resembling or proximal partners, is that *coalition mates choose each other*. The fundamental hypothesis is that *language controls social bonds*, while actual bonds are not a prerequisite for the former to exist.

Several points must be more precisely investigated. For instance, several aspects of the social game, such as the selection of best partners by displaying individuals, are merely enforced in the simulation, instead of being left to emerge. We think nevertheless that this simple experimental setting offers new support to the idea that human language emerged as a signalling device to advertise information abilities.

References

Axelrod, R. & Hamilton W. D. (1981). "The evolution of cooperation". *Science* 211, 1390-1396.

Dessalles, J-L. (1998). "Altruism, status, and the origin of relevance". In: J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds), *Approaches to the evolution of language: social and cognitive bases*. Cambridge: Cambridge University Press, 130-147.

Dessalles, J-L. (1999). "Coalition factor in the evolution of non-kin altruism". *Advances in Complex Systems* 2(2), 143-172.

Dessalles, J-L. (2000a). *Aux origines du langage : Une histoire naturelle de la parole*. Paris: Hermès-sciences.

Dessalles, J-L. (2000b). "Language and hominid politics". In: C. Knight, M. Studdert-Kennedy & J. R. Hurford (Eds), *The evolutionary emergence of language: social function and the origins of linguistic form*. Cambridge: Cambridge University Press, 62-79.

Dunbar, R. I. M., Duncan N. & Nettle D. (1995). "Size and structure of freely forming conversational groups". *Human nature* 6(1), 67-78.

Dunbar, R. I. M. (1998). "Theory of mind and the evolution of language". In: J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds), *Approaches to the evolution of language: social and cognitive bases*. Cambridge: Cambridge University Press, 92-110.

Gintis, H., Smith E. A. & Bowles S. (2001). "Costly Signaling and Cooperation". *Journal of Theoretical Biology* 213, 103-119.

Grafen, A. (1990). "Biological signals as handicaps". *Journal of Theoretical Biology* 144, 517-546.

Izquierdo, S. S., Izquierdo L. R., Galán J. M. & Hernández C. (2005). "Market failure caused by quality uncertainty". In: P. Mathieu, B. Beaufils & O. Brandouy (Eds), *Artificial Economics*. Springer LNEMS 564, 203-213.

Nowak, M. A. & Komarova N. L. (2001). "Towards an evolutionary theory of language". *Trends in cognitive sciences* 5(7), 288-295.

Pinker, S. (1994). *The language instinct*. New York: Harper Perennial, ed. 1995.

Sober, E. & Wilson D. S. (1998). *Unto Others. The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Trivers, R. L. (1971). "The evolution of reciprocal altruism". *The Quaterly Review of Biology* 46, 35-57.

Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton University Press, ed. 1996.

Zahavi, A. & Zahavi A. (1997). *The handicap principle*. New York: Oxford University Press.

Zahavi, A. (2003). "Indirect selection and individual selection in sociobiology". *Animal behaviour*. 65, 859-863.

INNATENESS AND CULTURE IN THE EVOLUTION OF LANGUAGE

MIKE DOWMAN, SIMON KIRBY
Language Evolution and Computation Research Unit, Linguistics and English Language
The University of Edinburgh, Edinburgh, EH8 9LL, UK

THOMAS L. GRIFFITHS
Cognitive and Linguistic Sciences, Brown University
Providence, RI 02912, USA

Is the range of languages we observe today explainable in terms of which languages can be learned easily and which cannot? If so, the key to understanding language is to understand innate learning biases, and the process of biological evolution through which they have evolved. Using mathematical and computer modelling, we show how a very small bias towards regularity can be accentuated by the process of cultural transmission in which language is passed from generation to generation, resulting in languages that are overwhelmingly regular. Cultural evolution therefore plays as big a role as prior bias in determining the form of emergent languages, showing that language can only be explained in terms of the interaction of biological evolution, individual development, and cultural transmission.

1. Introduction

Why is language the way it is and not some other way? Answering this *why* question is one of the key goals of modern linguistics. We can reframe this question as one about *language universals* in the broadest sense. Universals are constraints on variation, and include fundamental structural properties of language, such as compositionality, recursion, and semi-regularity. Language has arisen from the interactions of three complex adaptive systems: individual development, biological evolution, and cultural transmission, so a satisfactory approach to language should take into account each of these three systems, and how they interact.

Within generative linguistics, language structure is explained in terms of innately determined properties of the acquisition mechanism, and the constraints it places on developmental pathways (Chomsky, 1965). Whilst the generative approach is often contrasted with *linguistic functionalism*, which focuses on language use rather than acquisition, functionalism aims to explain aspects of language structure in terms of processing capacity, which is also an innately determined property of individual language users, though not necessarily one that is specific to language (see Kirby, 1999 for discussion). These approaches shift the burden of explaining linguistic structure to one of explaining how our

innate learning biases arose. Therefore explanations of linguistic structure are shifted to the domain of biological evolution, in which our innate prior biases are shaped. These in turn affect individual development, and therefore ultimately the universal properties of human language.

2. Cultural Transmission and the Bottleneck Effect

The arguments made so far have neglected the problem of explaining through what mechanism the properties of individuals give rise to properties of languages (Kirby, 1999). Recent work has addressed this issue using *iterated learning models*, computer programs comprising simple models of language-learning individuals that are then placed in a simulated population so that the dynamics arising from their interactions can be studied (e.g. Batali 1998; Kirby, 2001). The simulated individuals can learn from one-another, and so language can be passed from one generation of speakers to the next, a process which gives rise to the cultural evolution of language. Fig. 1 shows how iterated learning forms a bridge between individual bias and language structure.

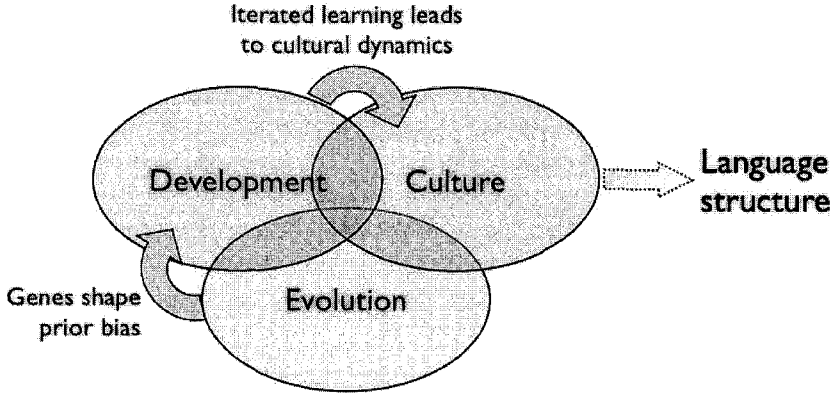


Figure 1: The universal structure of human language arises out of the process of iterated learning, an adaptive system operating on a cultural time-scale, driven by individual biases that are ultimately shaped by biological evolution.

One of the most basic and pervasive properties of human languages is *regularity*. That is, if one meaning is expressed using a particular rule or construction, then it is likely that another meaning will also make use of the same pattern. We might therefore expect regularity to arise from some fundamental aspect of our language faculty – in other words our prior bias might be expected to reflect this central universal strongly. However, in a wide variety of models using a diverse set of learning algorithms and assumptions about signal and meaning spaces (e.g. Batali, 1998; Kirby, 2001), the overwhelming conclusion is that strong biases are not necessary to explain the emergence of pervasive

regularity. It seems that regularity emerges whenever the number of training samples that the learners are exposed to is small. If there is too little data stable languages do not emerge, while if there is too much training data, the emergent languages are not regular, but instead express meanings in an ad hoc way. Kirby, Smith & Brighton (2004) explain this behaviour in terms of adaptation to the *bottleneck* – the limited amount of linguistic examples from which each speaker must learn the language. A regular rule can persist into the next generation so long as the learner sees only one example of it, but an irregular expression can only persist into the next generation if the learner is exposed to exactly that expression.

3. Iterated Bayesian Learning

While a wide range of learning algorithms, with a correspondingly wide range of biases, have resulted in the emergence of regular linguistic structure, it is very hard to determine exactly what the biases inherent in the various models of learning actually are. Because of this, it is difficult to be sure of the generality of the result. To address this issue, we have begun to explore a Bayesian version of iterated learning, which has the advantage that we can make the biases of learners completely explicit, and manipulate them freely.

In the Bayesian framework, constraints on both acquisition and processing can be characterised as a probability distribution over possible languages. In this view, the language learner is faced with the task of forming a hypothesis about the language of her speech community on the basis of data that she is exposed to. She aims to assess the *posterior* probability, $P(h|d)$, of a hypothesis (i.e. a language) h, based upon the observed data d. Bayes' rule indicates that this should be done by combining two quantities: the *likelihood*, $P(d|h)$, being the probability of the observed data d given hypothesis h, and the *prior probability*, $P(h)$, indicating the strength of the learner's *a priori* bias towards the hypothesis h. According to Bayes rule the posterior probability of a hypothesis is proportional to the product of the associated likelihood and prior probability, as shown in Eq. (1). In other words, the learner must take into account how well each possible hypothesis predicts the data seen *and* how likely each hypothesis is *a priori*. Conceived in this way, the influence of the learner's language acquisition device (Chomsky, 1965) and language-processing machinery is characterised as the prior probability distribution over hypotheses.

$$P(h \mid d) \propto P(d \mid h)P(h) \tag{1}$$

Griffiths & Kalish (2005) investigated iterated learning under the assumption that the learners first calculate the posterior distribution over languages, and then *sample* from this distribution (i.e., they choose a language with a probability equal to its posterior probability). By viewing the process of iterated learning as a Markov chain (c.f. Nowak et al., 2001), they were able to

prove that such a process will result in a distribution of languages that exactly mirrors the prior bias.[a] This is a startling result, and one that renders the results from previous simulations mysterious, as the process of cultural evolution makes no independent contribution to the emergent languages. In particular, the number of training samples – the bottleneck size – has *no effect whatsoever* on the probability of each type of language emerging. Why then is the bottleneck size the crucial factor in all the simulation models of iterated learning?

The answer to this puzzle turns out to hinge on our assumptions about what a rational learning agent should do when faced with a choice between languages. It might seem more rational for learners to pick the language with the maximum *a posteriori* probability (which in Bayesian learning theory is called the MAP hypothesis), rather than sampling from this distribution, as in Griffiths and Kalish's approach. This would maximise the chance of picking the same language as that spoken by the previous agent. This small difference between the sampling and MAP learner turns out to have huge implications for the dynamics of iterated learning.

To demonstrate why this is the case, we can work through a simple idealised model which nevertheless reflects the general case of iterated Bayesian learning. The first step in constructing such a model is to decide on the form of the space of logically possible languages. For this example, we aim to explain the origins of *regularity* in language. Regularity here can be seen as an umbrella concept that covers a variety of aspects of language. We will treat a language as a deterministic function from discrete *meanings* to discrete *classes*. Depending on how the model is interpreted these could be thought of as classes across a morphological paradigm, or an indication of the form of a compositional encoding of a particular meaning. The idea is that a language in this model is completely *regular* if all its meanings belong to the same class, and is completely *irregular* if all the meanings belong to different classes.

Learners are exposed to a set of m utterance-meaning pairs, each of which consists of a meaning and the class used to express it. They then use Bayes rule to find the most likely language to have generated this set of meaning-class pairs. Finally, this language is used to generate a new set of meaning-class pairs. This is done by sampling m meanings at random and using the hypothesized language to generate classes for each. In addition, we model noise in the system by randomly picking a different class to the correct one for each meaning with probability $\cdot$. The dynamics of this system can be completely characterized for a particular value of m (the bottleneck) and $\cdot$ (the noise factor), by looking at the

[a] Strictly speaking, this is only the case if the prior bias results in a dynamical system that is *ergodic*. Simplifying, this essentially means that every language in the space must be at least potentially "reachable" by cultural evolution.

probability that a language l_1 spoken by one agent will give rise to a learner choosing a language l_2, for all pairs of languages (l_1,l_2), as shown in Eq. (2). $\boldsymbol{x} = \{x_1,\ldots,x_m\}$ is the set of meanings chosen at random, and $\boldsymbol{y} = \{y_1,\ldots,y_m\}$ is the corresponding set of classes output by the speaker, and $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the possible sets of meanings and output classes respectively.

$$P(l_2 \mid l_1) = \sum_{\boldsymbol{x}\in\boldsymbol{X}} \sum_{\boldsymbol{y}\in\boldsymbol{Y}} P(l_2 \text{ is MAP} \mid \boldsymbol{x},\boldsymbol{y}) P(\boldsymbol{y} \mid \boldsymbol{x}, l_1) P(\boldsymbol{x}) \tag{2}$$

The probability that a language is the MAP language will normally be either 1 or 0, but where several languages are tied with equally high posterior probability, this value will be equal to 1 divided by the number of such languages. Eq. (3) shows how the classes are produced for a given language. We can think of $P(l_2|l_1)$ as a matrix of transitions from language to language (what Nowak et al. 2001 call the Q-matrix), defining a Markov chain over languages. The first eigenvector of this transition matrix gives the *stationary distribution* for the Markov chain, indicating the distribution over languages that will emerge out of iterated learning (provided the underlying Markov chain is ergodic). We now have everything in place to determine what universal properties will emerge for a given bottleneck, noise-term and prior bias.

$$P(\boldsymbol{y} \mid \boldsymbol{x}, l) = \begin{cases} 1-\varepsilon & \text{if } \boldsymbol{y} \text{ is the class corresponding to } \boldsymbol{x} \text{ in } l \\ \dfrac{\varepsilon}{|\boldsymbol{y}|-1} & \text{otherwise} \end{cases} \tag{3}$$

What is a reasonable prior bias for the language model we have described? One possibility is to have a completely uninformative prior, with every language being equally likely. Unsurprisingly, this results in no clear preference for one language over another in the final result. We can infer from this directly that the prior does indeed matter. Since language exhibits regularity, human language learners cannot be "blank slates". Instead, we make the minimal assumption that learners will expect future events to be similar to previous events. If we have n meanings and k classes, this assumption is embodied in the prior specified by Eq. (4), where n_j is the number of meanings assigned to class j, α is a parameter of the prior, and $\Gamma(x)$ is the generalised factorial function, with $\Gamma(x) = (x\text{-}1)!$ when x is an integer. The specific form of the prior can be justified both from the perspective of minimum description length (Rissanen, 1978), and from the perspective of Bayesian statistics, where $P(l)$ appears as a special case of the Dirichlet-multinomial distribution (Johnson & Kotz, 1972). α controls the strength of the prior, or alternatively how much of a bias towards regularity is built into the model. Low values of α create a strong regularity bias, and high values a much weaker one.

$$p(l) = \frac{\Gamma(k\alpha)}{\Gamma(\alpha)^k \Gamma(n+k\alpha)} \prod_{j=1}^{k} \Gamma(n_k + \alpha) \tag{4}$$

The key question which we wish to address is: how strong does this bias towards regularity need to be? As we mentioned above, Griffiths and Kalish (2005) have shown that, for a sampling learner, the expected distribution of languages exactly reflects the prior. So, since languages are overwhelming regular, this suggests the prior bias must be very strong. With the MAP learner, however, this turns out not to be the case as long as there is a bottleneck on linguistic transmission.

For example, we looked at a simple model of languages with four possible meanings ($n = 4$), and four possible classes for each meaning ($k = 4$). From the perspective of regularity, there are five different types of language in this space: all meanings in the same class; three meanings in one class and one in another; two meanings in one class and the other two in a second class; two meanings in one class and the other two in two different classes; and all four meanings in different classes. For shorthand, we label these types: *aaaa, aaab, aabb, aabc,* and *abcd* respectively. The first row of Table 1 shows the prior probability of these five types of languages under the prior $P(l)$ described above, with $\alpha = 10$, so that there is only a weak preference for regularity.

Table 1. The predicted distribution of language types for different bottleneck sizes, in terms of the probability of a particular language of each type. The prior distribution is shown for comparison. As the bottleneck on linguistic transmission tightens, the preference for regularity is increasingly over-represented in the distribution of languages.

Languag e	*aaaa*	*aaab*	*aabb*	*aabc*	*abcd*
Prior	0.00579	0.00446	0.00409	0.00371	0.00338
m = 10	0.145	0.00743	0.000449	0.000324	0.0000335
m = 6	0.175	0.00566	0.000150	0.000158	0.0000112
m = 3	0.209	0.00329	0	0.0000370	0

Given this prior, the expected distributions of languages by type for different bottleneck sizes and an error-term ($\cdot$) of 0.05 are shown in the remainder of Table 1. What is immediately obvious from these results is that the prior is *not* a good predictor of the emergent properties of the languages in the model. The *a priori* preference for regularity is being hugely over-represented in the languages that evolve culturally. In fact the strength of the bias often has no effect on the resulting stationary distribution – it is the ordering it imposes on languages that is most important. Furthermore, as with the simulation models, the tightness of the bottleneck on linguistic transmission is a determining factor in how regular the languages are.

Of course, languages typically are not completely regular. Irregularity is a common feature of morphological paradigms. This irregularity is not randomly

distributed throughout a language, but rather appears to correlate with frequency – for example, the ten most common verbs in English are all irregular in the past tense. Previous iterated learning simulations have suggested that this can be explained in terms of adaptation to cultural transmission. Put simply, frequent verbs can afford to be irregular, since they will have ample opportunity to be transmitted faithfully through the bottleneck (Kirby, 2001).

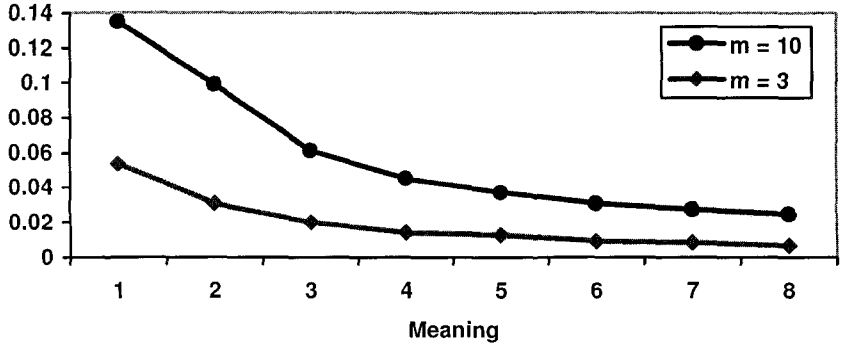


Figure 2. Irregularity correlates with frequency. These are results from simulations using 8 meanings and 4 classes, with • = 0.05 and $\alpha \approx 1$. They show the proportion of languages in which each meaning was expressed using an irregular construction. The frequency of each meaning was inversely proportional to its rank (counting left to right). A meaning was counted as regular if its class was in the majority.

To test this intuition, we made a simple modification to our model. Rather than picking meanings at random from a uniform distribution, they were skewed so that some were more common than others. The results in Fig. 2 show how often each meaning was irregular (i.e. in a minority class) for a language model with 8 meanings and 4 classes. The frequency of each meaning decreases from left to right in this graph, demonstrating that the model results in a realistic frequency/regularity interaction.

4. Conclusion

Language involves three adaptive systems: biological evolution, individual development, and cultural transmission. An adequate account of the origins of linguistic structure must crucially focus on the interactions between these systems. Our contribution has been to demonstrate that the innovation of cultural transmission radically alters the relationship between our innate learning biases and our linguistic behaviour.

The implications of this for the study of language evolution are, firstly, that our innately given bias cannot be directly inferred from our phenotype (i.e. language). More specifically, weak innate biases can nevertheless lead to strong

universals wherever there is a bottleneck on the cultural transmission of language. This leads naturally to an explanation of frequency-related patterns of regularity and irregularity in language assuming only a weak expectation of predictability on behalf of learners. Finally, this result demonstrates that we must be cautious in assuming that adaptive structure necessitates an explanation in terms of the selective evolution of innate traits that are specifically linguistic. Regularity is an adaptive feature of language but we have shown that the mechanism for adaptation need not be biological.

In this paper we have not looked at the final interaction in Fig. 1 – between culture and evolution. Much work remains to be done. However, we note that our result shows iterated learning can shield the strength of innate biases from the view of natural selection. The implications of this are beginning to be worked out, but it is clear that an account of the biological evolution of the human language faculty cannot be complete if it fails to take into account the interactions between innateness, culture and linguistic structure.

References

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy and C. Knight (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge: Cambridge University Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Griffiths, T. and Kalish, M. (2005). A Bayesian view of language evolution by iterated learning. In B. G. Bara, L. Barsalou and M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson, N. L. and Kotz S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. New York, NY: John Wiley & Sons.

Kirby, S. (1999). *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford: Oxford University Press.

Kirby, S. (2001) Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2): 102--110.

Kirby, S., Smith, K. and Brighton, H. (2004) From UG to Universals: Linguistic Adaptation through Iterated Learning. *Studies in Language*, 28(5): 587-607.

Nowak, M. A., Komarova, N. L., and Niyogi, P. (2001). Evolution of universal grammar. *Science*, 291: 114–118.

Rissanen, J. (1978). Modelling by shortest data description. *Automatica*, 14: 465–471.

EARLY HUMAN LANGUAGE WAS ISOLATING-MONOCATEGORIAL-ASSOCIATIONAL

DAVID GIL

Max Planck Institute for Evolutionary Anthropology
Leipzig, 04103, Germany

Isolating-Monocategorial-Associational (IMA) Language is language with the following three properties: (a) *morphologically isolating*, without word-internal morphological structure; (b) *syntactically monocategorial*, without distinct syntactic categories; and (c) *semantically associational*, without distinct construction-specific semantic rules, compositional semantics relying instead on the association operator, which says that the meaning of a composite expression is associated with the meanings of its constituents in an underspecified fashion. IMA Language is present in the following five domains: (a) *phylogeny:* at some stage in evolution, early language was IMA Language; (b) *ontogeny:* at some stage in acquisition, early child language is IMA Language; (c) *semiotics:* some artificial languages are IMA Language; (d) *typology:* some languages are closer than others to IMA Language; and (e) *cognition:* IMA Language is a feature of general human cognition. This paper presents arguments pertaining to the first of these domains, namely phylogeny, citing evidence from the linguistic behaviour of captive apes which points towards the conclusion that early human language was IMA Language.

Imagine a hypothetical language, either natural or artificial, with the following three properties:

(1) (a) *Morphologically Isolating*
No word-internal morphological structure;
(b) *Syntactically Monocategorial*
No distinct syntactic categories;
(c) *Semantically Associational*
No distinct construction-specific rules of semantic interpretation (instead, compositional semantics relies exclusively on the Association Operator, defined in (2) below).

Such a language may be referred to as *Isolating-Monocategorial-Associational*, or for short, *IMA*.

Does IMA Language exist? Obviously, English is not an IMA Language, nor, to the best of my knowledge, has any other natural language been proposed to be in complete possession of the three defining properties in (1) above. However, in Gil (2005) it is argued that the notion of IMA Language is of relevance to a variety of domains: phylogeny, ontogeny, semiotics, typology and cognition. The present paper focusses on the first of these domains, suggesting

that IMA Language represents an earlier stage in the evolution of modern human language.

The three defining properties of IMA Language pertain to three different linguistic domains, morphology, syntax and semantics; logically, they are thus independent of each other. Accordingly, one may imagine various other kinds of hypothetical languages with different subsets of the three properties, for example a language that is isolating but not monocategorial or associational.

The defining properties of IMA Language represent the limiting points of maximal simplicity within each of the three domains, morphology, syntax and semantics. Hence, for each domain, one may imagine languages approaching these end points along a scale of decreasing complexity. Accordingly, a language is increasingly isolating as it has less and less morphological structure, increasingly monocategorial as its syntactic categories decrease in number and importance, and increasingly associational as its construction-specific rules of semantic interpretation become fewer and less distinct. Alongside *Pure IMA Language*, as in (1) above, one may thus entertain the possibility of a range of *Relative IMA Languages*, approaching Pure IMA Language to various degrees within each of the three domains. Thus, for example, in Gil (2005) it is argued that Riau Indonesian is one such Relative IMA Language.

The first defining property, *morphologically isolating*, is the one that is most familiar, since it pertains to a typology that has been the focus of considerable attention in the linguistic literature. As is well known, isolating languages such as Vietnamese have considerably less word-internal morphological structure than synthetic languages such as Russian, which in turn have considerably less morphology than polysynthetic languages such as Mohawk. However, no natural language is purely isolating, as per (1a); all known isolating languages still have *some* morphology — affixation, compounding, or other kinds of processes such as reduplication, stem alternation, and so forth.

The second defining property, *syntactically monocategorial*, pertains to a domain within which the presence of cross-linguistic variation has only recently, and still only partially, been recognized. In the past, syntactic categories have generally been presumed to be universal, often in accordance with the eight parts of speech of traditional Latin grammar. Indeed, the assumption that syntactic categories must be the same in all languages has lingered on into much current linguistic work, in schools as diverse as linguistic typology and generative grammar; this assumption is in evidence whenever a linguist analyzing a language says that one word must be a noun because it means 'chicken' while another word must be a verb because it means 'eat'.

However, in recent years an increasing body of literature has begun to examine the ways in which the inventories of syntactic categories may vary across languages. One important issue that has attracted considerable attention

has been the viability and nature of the category of adjective, the extent to which words denoting properties such as 'big', 'red', 'good' and so forth exhibit distinct adjectival behaviour, or, alternatively, are subsumed within larger categories of noun or verb. Another major focus has been on the universality of what is generally considered to be the most fundamental categorial distinction, namely that between noun and verb; such work has typically dealt with languages which seem, prima facie, to lack a noun/verb distinction, from families such as Munda, Austronesian, Salish and Wakashan. It is of course languages lacking a noun/verb distinction which come closest to being syntactically monocategorial. However, to the best of my knowledge, no language has ever actually been proposed to be purely monocategorial. In particular, most or all descriptions of languages without a noun/verb distinction still involve, at the very least, a distinction between a single open syntactic category (encompassing the equivalents of both nouns and verbs) and one or more closed syntactic categories containing various "grammatical" or "functional" items.

The third defining property, *semantically associational*, although rooted in various common-place observations concerning the ways in which expressions derive their meanings, is nevertheless of a more novel nature. Consider the best translation of a basic transitive sentence such as 'Mary hit John' into the language of your choice. How do you know who hit whom? If you chose Mandarin, then, like in English, the agent is differentiated from the patient by linear order: the agent precedes the verb while the patient follows it. However, if you chose Russian, then linear order provides no semantic information; instead, the agent is differentiated from the patient by its case marking, nominative as opposed to accusative, and by the fact that it triggers gender agreement on the past-tense form of the verb. The various rules according to which agents and patients are differentiated in English, Mandarin, Russian and other languages constitute examples of *construction-specific rules of semantic interpretation*, as specified in (1c) above, in that they apply specifically to active transitive clauses. Most languages contain many such construction-specific rules, which, together, govern the compositional semantics of clauses, phrases, and other, more specific constructions, accounting for semantic features such as thematic roles, tense, aspect, number, definiteness, and numerous others.

Now imagine you are confronted with a three-word sentence in an unfamiliar language, armed only with a rudimentary dictionary. Somehow, you identify three word stems, meaning 'Mary', 'hit' and 'John'; however, these three word stems bear rich additional morphological structure, and you know nothing about the grammar of the language. Can you figure out the meaning of the sentence? At first blush, the answer would seem to be no. With no information on thematic roles, tense, aspect, number, definiteness, and other such features, the sentence could mean anything from 'Mary hit John' through 'John will repeatedly try to hit Mary' to 'John and Mary aren't hitting anybody' and so on

and so forth. Still, the meaning of the sentence is hardly unconstrained: it is not very likely to mean 'The rain in Spain falls mainly in the plains'. Thus, although you have no knowledge of the grammar of the language, it is a safe bet, in fact a near certainty, that the meaning of the sentence, whatever it is, *has to do in some way with* 'Mary", 'hit' and 'John'.

The semantic relationship of "having to do with" may be formally represented by means of the *Association Operator*, defined as follows:

(2) *The Association Operator A:*
Given a set of n meanings $M^1 ... M^n$, the Association Operator A derives a meaning A ($M^1 ... M^n$) read as 'entity associated with M^1 and ... and M^n'.

Two subtypes of the Association Operator may be distinguished, the *Monadic Association Operator*, in which n equals 1, and the *Polyadic Association Operator*, for n greater than 1.

In its monadic variant, the Association Operator is familiar from a wide variety of constructions in probably all languages. Without overt morphosyntactic expression, it is manifest in cases of metonymy such as the often cited *The chicken left without paying*, where the unfortunate waiter uses the expression *the chicken* to denote the person who ordered the chicken. More commonly, the Monadic Association Operator is overtly expressed via a specific form, which is commonly referred to as a *genitive*, *possessive* or *associative* marker. Some idea of how unconstrained the association is can be obtained by comparing the obvious meanings of the English enclitic possessive marker *'s* in phrases such as *John's father, John's nose, John's shirt, John's birthday, John's suggestion* and so forth, or by considering the range of meanings of a single phrase such as *John's book*, which could denote the book that John owns, the book that John wrote, the book that's about John, or, in more specific contexts, the book that John was assigned to write a review of, and so forth.

In its polyadic variant, the Association Operator provides for a basic mechanism of compositional semantics in which the meaning of a complex expression is derived from the meanings of its constituent parts. In accordance with the Polyadic Association Operator, whenever two or more expressions group together to form a larger expression, the meaning of the combined expression is associated with, or has to do with, the meanings of each of the individual expressions. Obviously, polyadic association applies in a default manner throughout language; it is hard to imagine how things could be otherwise. Thus, in the little thought experiment described above, it is what made it possible to be sure that in an unfamiliar language, in the absence of any specific grammatical information, in a sentence with three words whose meanings were based on 'Mary', 'hit' and 'John', the meaning of the sentence

would still be associated in some way with 'Mary', 'hit' and 'John', or 'entity associated with Mary, hitting and John'.

One grammatical domain in which the Polyadic Association Operator is overtly visible is in genitive constructions. In many languages, genitive constructions are formed by the bare juxtaposition of the two expressions, in which case the derived meaning may be represented by means of the Polyadic Association Operator applying without any overt morphosyntactic expression. More generally, the Polyadic Association Operator may be considered as a universal default mechanism for semantic interpretation, but one that is in most cases overridden and narrowed down substantially by the application of additional construction-specific rules. A purely associational language would be one in which there were no such further construction-specific rules of semantic interpretation, and in which, therefore, the compositional semantics were effected exclusively by the Polyadic Association Operator. It is almost certainly the case that no natural language is purely associational; however, as argued in Gil (2005), some languages, such as Riau Indonesian, may come closer to being purely associational than is generally assumed.

In general, then, Pure IMA Language represents a limiting case of maximal simplicity within the domains of morphology, syntax and semantics. One may indeed wonder whether IMA Language is capable of fulfilling the multifarious functions associated with human language in the diverse contexts in which it is used. Nevertheless, IMA Language is in fact more widespread than might be expected, and can indeed fulfil a wider range of functions than might seem, prima facie, to be the case.

IMA Language, or a system that comes close to IMA Language, is manifest in the following five distinct ontological realms:

(3) (a) *Phylogeny*
At some stage in evolution, early language was IMA Language;
(b) *Ontogeny*
At some stage in acquisition, early child language is IMA Language;
(c) *Semiotics*
Some artificial languages are IMA Language;
(d) *Typology*
Some languages come closer than others to IMA Language;
(e) *Cognition*
IMA Language is a feature of general human cognition.

This paper focusses on the first of these domains, namely phylogeny; the remaining four domains are discussed in more detail in Gil (2005).

Although we have preciously little direct evidence of any kind concerning the evolution of natural language, it is reasonable to suppose that early human language was IMA Language. More precisely, the following two logically distinct hypotheses may be formulated:

(4) (a) *Evolution of Linguistic Abilities*
At some stage in evolution, the cognitive abilities of humans or pre-humans were limited to the representation of IMA Language;
(b) *Evolution of Actual Languages*
At some stage in evolution, all natural languages were IMA Language.

While hypothesis (4a) is about the evolution of cognition, or, more specifically, mental grammar, sometimes referred to as I-language, hypothesis (4b) is about the evolution of actual languages, also known as E-languages.

A commonly held position, most often associated with Chomsky and his followers, is that contemporary human linguistic abilities emerged ex nihilo in a single gigantic leap, presumably associated with a unique genetic mutation. Such a view is clearly inconsistent with hypothesis (4a); however it is agnostic with respect to hypothesis (4b), since even if human linguistic abilities went straight from nothing to what they are now, actual languages might have taken a variety of incremental paths over the course of time in order to make use of such abilities (indeed this process may still be far from complete); and one of those possible paths could easily have involved IMA Language as an evolutionary way station.

A more refined position is put forward by Bickerton (1990), who argues that man's linguistic abilities evolved into their contemporary shape through an intermediate stage which he refers to as *protolanguage*. Structurally, Bickerton's protolanguage is a form of IMA Language; however, it embodies at least one significant further restriction that is not part of IMA Language, namely that it does not permit syntactic recursion. Ontologically, too, Bickerton's protolanguage is akin to IMA Language, in that he considers it to be manifest in a variety of realms, including three of the five listed in (3) above: phylogeny, ontogeny, and cognition. Notably, however, Bickerton has nothing to say about the other two domains, semiotics and typology. Moreover, he expressly denies the existence of any "interlanguage" between protolanguage and contemporary linguistic abilities; thus, like Chomsky, his position is inconsistent with hypothesis (4a), though in the case at hand, what is at issue is a single, albeit very important structural feature, namely, syntactic recursion. Conversely, hypothesis (4a) is consistent with, but does not necessarily entail, the existence of a stage, prior to IMA Language and the evolution of recursion, corresponding to Bickerton's protolanguage.

So how might we seek support for the two evolutionary hypotheses in (4)? Although we cannot go back in time, we can jump across the branches of our evolutionary tree to see what our nearest relatives, the various primates, have accomplished in the realm of language. Many species have a lexicon of predator cries; however, since these usually involve individual cries in isolation, there is no compositionality, and hence nothing near the possible richnesses of IMA Language. A somewhat more interesting case, reported recently by Zuberbuhler (2002), is that of the male Campbell's monkeys, who appear to be able to juxtapose two different calls, a predator cry preceded by a "boom" sound, to produce a complex cry whose meaning seems to involve some kind of attenuation or even negation of the predator-cry meaning. However, to this point at least, no clear examples of productive compositionality of meaning-bearing signs have been attested in the naturally-occurring repertoire of non-human primates, or any other animals.

However, amongst primates in captivity, there is an increasing body of evidence suggesting that they can be taught to master compositionality, and concomitantly also IMA Language. Two of the more celebrated cases are those of the bonobo Kanzi (Greenfield and Savage-Rumbaugh 1990) using lexigrams, and the orangutan Chantek (Miles 1990) using American Sign Language. Some examples of Kanzi's spontaneous lexigram production are given below:

(5)	(a)	LIZ HIDE	*agent - HIDE*
	(b)	WATER HIDE	*patient - HIDE*
	(c)	HIDE AUSTIN	*HIDE - agent*
	(d)	HIDE PEANUT	*HIDE - patient*

Kanzi's usage of lexigrams provides no evidence for morphological structure or for distinct syntactic categories; it is thus isolating and monocategorial. Moreover, as suggested by examples such as the above, it is also associational. The above examples form a miniature paradigm (schematized to the right) in which the same sign hide is either preceded or followed by a participant, which, as indicated by the context of the utterance given by the authors, may, in either position, be understood as either the agent or the patient. Thus, there would seem to be no evidence for any grammatical assignment of thematic roles in the lexigram usage of Kanzi. Rather, the semantic relationship between the two signs is vague. Like in the language of pictograms and example (4) above, the juxtaposition of lexigrams has a single general meaning that may be represented in terms of the Polyadic Association Operator as, for (5a), A (liz, hide), 'entity associated with Liz and with hiding'. Thus, the bonobo Kanzi's use of lexigrams satisfies the three properties of IMA Language. Similar observations hold also for the orangutan Chantek's usage of ASL.

It would seem, then, to be the case that both bonobos and orangutans are endowed with the cognitive abilities to represent IMA Language, even though they apparently have not made any use of these abilities to create any actual IMA Languages in the wild. Given that the common evolutionary ancestor of bonobos and orangutans is shared also by humans, it is thus likely that this common ancestor also had the cognitive abilities to represent IMA Language without having any actual IMA Languages. (The alternative, less parsimonious scenario would involve positing the independent development of IMA Language abilities in at least two separate evolutionary lineages.) Quite obviously, however, no primates, even in captivity and with the dedicated efforts of their caregivers, are capable of acquiring the full-blown complexities of natural human language. Thus, the linguistic capabilities of captive apes support the reconstruction of a stage in human evolution, perhaps eight or ten million years ago, in which the abilities to represent IMA Language were already present, in accordance with hypothesis (4a). Alongside the above, the linguistic capabilities of captive apes also increase the plausibility of hypothesis (4b), though the alternative logical possibility remains that pre-human cognitive abilities may have developed past IMA Language before actual languages ever reached the IMA stage.

It should be noted, though, that since, to the best of my knowledge, the linguistic behaviour of captive apes does not provide any evidence for the mastery of syntactic recursion, the abilities of Kanzi, Chantek and other such captive apes may equally well be characterized in terms of the more restrictive protolanguage of Bickerton. In order to provide specific support for the existence of an evolutionary stage of IMA Language, either in addition to or instead of protolanguage, evidence of a different kind is called for: at present I am not familiar with any such evidence.

References

Bickerton, D. (1990) *Language and Species*, University of Chicago Press.

Gil, D. (2005) "Isolating-Monocategorial-Associational Language", in H. Cohen and C. Lefebvre eds., *Categorization in Cognitive Science*, Elsevier, Oxford.

Greenfield, P.M., & Savage-Rumbaugh S. (1990) "Grammatical Combination in Pan Paniscus: Processes of Learning and Invention in the Evolution and Development of Language", in S. T. Parker and K. R. Gibson eds., *"Language" and Intelligence in Monkeys and Apes, Comparative Developmental Perspectives*, Cambridge University Press, Cambridge, 540-578.

Miles, H.L.W. (1990) "The Cognitive Foundations for Reference in a Signing Orangutan", in S. T. Parker and K. R. Gibson eds., *"Language" and Intelligence in Monkeys and Apes, Comparative Developmental Perspectives*, Cambridge University Press, Cambridge, 511-539.

Zuberbuhler, K. (2002) "A Syntactic Rule in Forest Monkey Communication", *Animal Behaviour* 63:293-299.

COMPUTATIONAL SIMULATION ON THE COEVOLUTION OF COMPOSITIONALITY AND REGULARITY

TAO GONG, JAMES W. MINETT, WILLIAM S-Y. WANG

Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong, China

Compositionality and regularity are universals in human languages; in most languages, complex expressions are determined by their structures and their components' meanings. Based on a multi-agent computational model, the coevolution of compositionality and one type of regularity, word order, is traced during the emergence of compositional language out of holistic signals. The model modifies some questionable aspects in the Iterated Learning Model and Fluid Construction Grammar by considering the conventionalization in horizontal transmission and the gradual formation of syntactic categories which mirror the semantic categories. The model also implements a bottom-up syntactic developmental process, i.e., the global orders for regulating multiple arguments are gradually formed from simple local orders between two categories.

1. Introduction

There are two mainstream views on language emergence (Minett & Wang 2005). *Innatism* (e.g., Anderson & Lightfoot 2002) ascribes linguistic universals to certain innate mechanisms unique to humans (e.g., *Language Acquisition Device*). However, recent evidence from anatomical, biological and comparative studies of the human communication system and those of animals (Oller & Griebel 2004) suggests that language might have evolved from domain-general abilities, and these studies support *Emergentism* (e.g., Knight et al. 2000). Emergentism also gains support from many behavior-based computational models which have discussed the effects of some communicational constrains and certain domain-general competences of language user(s) on the formation of linguistic universals. For instance, Kirby's *Iterated Learning Model* (ILM) (2003) implied that the *bottleneck effect* (the restricted exposure of the previous generation's language to the next generation) in vertical transmission is sufficient to trigger a compositional language; Luc Steels' *Fluid Construction Grammar* (FCG) (2004) developed an artificial language with simple argument constructions through iterated description games between two agents. Despite the significant insights they provided, these models neglected some important processes that might affect language emergence, and built in some questionable assumptions.

In ILM, it is insufficient to claim that compositionality has emerged based on the limited facts that agents have acquired compositional rules and that these rules have high expressivity. Using compositional materials to exchange *integrated meaning* (which describes a complete event containing an action (*Predicate* (Pr)), its instigator (*Agent* (Ag)) and sometimes the entity undergoing such action (*Patient* (Pt))) also requires regulating these materials under similar mechanisms (e.g., syntactic or morphological operations); without regularity, the sharing of common compositional materials is useless. For example, without syntactic information, it is unclear "who are chasing whom" when hearing "dogs chase cats" or "cats chase dogs". The emergence of regularity occurs in the production and comprehension of integrated meanings based on linguistic and nonlinguistic information available during the communications. Due to the assumption of *direct meaning transference*, in which the speaker's intended meaning is always transparent to the listener, ILM has a simplified comprehension process, which makes it unsuitable for studying how agents develop regularity during the comprehension based on linguistic information. The inseparability of the lexicon and grammar has already been discussed in some empirical studies on language acquisition (e.g., Bates & Goodman 2001). Besides, cultural evolution covers both vertical and horizontal transmission. *Conventionalization* in horizontal transmission (during which an individual's language conforms acceptably to the language of the community) is important for shaping the language in one generation. If certain mechanisms during this conventionalization already incline toward compositional materials, the bottleneck effect shown in ILM might be weakened.

FCG, similar to the Construction Grammar theory (Fried & Östman 2004), assumed that syntactic structures mirror semantic categories, but it did not demonstrate how such matching was established. FCG also built in a preference for linguistic information — reference to visual events occurs after the linguistic comprehension fails. How this preference is built needs further discussion.

Regarding these limitations, based on our previous work in Evolang5 (Gong & Wang 2005, Gong et al. 2005), we present a multi-agent model to study the simultaneous emergence of compositionality and one type of regularity (word order) via horizontal transmission. This model adopts an indirect meaning transference, in which the comprehension is based on both linguistic and nonlinguistic information. Based on simple imitation ability, compositional materials (in the form of rules) emerge when agents identify recurrent patterns in the *Meaning-Utterance mappings* (M-U mappings) acquired during previous communications. Considering that some empirical study has shown that some

primates can notice and manipulate simple orders (Hauser 1996), this model studies whether such simple sequential ability can lead to the complex syntax in human languages by adopting a categorization mechanism. Through categorization, rules which have the same semantic roles and which are used similarly in forming utterances are grouped into a category. The relative word orders of elements belonging to different categories are acquired as syntactic rules. Similar to the Verb Island hypothesis (Tomasello 2003), this categorization process traces how the semantics-syntax match becomes established. Meanwhile, complex orders that regulate elements from multiple categories can gradually form based on the simpler word orders between pairs of categories. This bottom-up syntactic development is consistent with the "tinkerer" view of evolution (Jacob 1977), which states that complex features can gradually develop out of simple available abilities. It also matches the "carpentry" theory of language evolution (O'Grady 2005), which points out that sentences can be resolved by simple processing forces driven by efficiency. In addition, similar developmental processes have already been traced in children's language acquisition, e.g., children progress from a "two-word" stage to a "multi-word" stage in the acquisition of lexical items and simple sentences (Clark 2003).

2. Model Description

2.1. *Linguistic Rules and Categories*

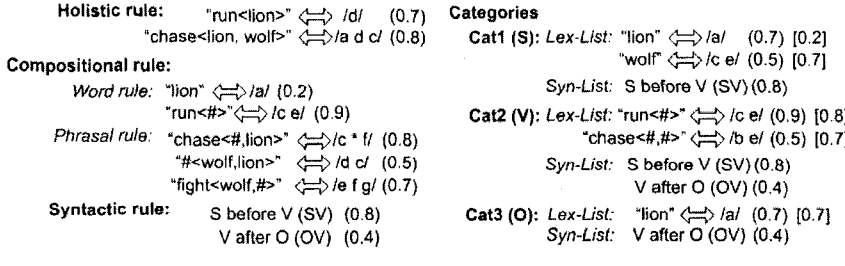


Figure 1. Linguistic rules and categories. "#" can be replaced by other semantic items, and "*" by other syllable(s). Numbers enclosed by ( ) denote the rule strengths, and those by [ ] denote the association weights of rules to their associated categories.

M-U mappings are used to represent language. The semantic space consists of a set of integrated meanings, each having an occurrence frequency. These integrated meanings are of two types: Type-I: "Pr_1<Ag>", e.g., "run<lion>"; and Type-II: "Pr_2<Ag, Pt>" e.g., "chase<wolf, lion>". *Utterances* (combinable syllables chosen from a signal space) map to either integrated meanings or some

semantic item(s) (e.g., "lion" or "run<#>"). Different semantic roles ("Ag", "Pt" and "$Pr_{1/2}$") correspond to different syntactic roles ("Subject" (S), "Object" (O) and "Verb" (V)) in utterances.

Agents use linguistic rules to produce utterances encoding integrated meanings and comprehend these meanings from utterances. Linguistic rules (see Fig. 1) include *lexical rules* (M-U mappings plus strengths) and *syntactic rules* (local word orders plus strengths). Lexical rules are either *holistic*, encoding an entire integrated meaning, or *compositional*, encoding part of an integrated meaning; the latter can be further divided into *word* and *phrasal rules* (see Fig. 1). Every syntactic rule uses one of four simple orders—*before*, *after*, *middle* or *surround*—to regulate how lexical members from two categories with different syntactic roles may be combined. The global order that regulates members from multiple categories is the combination of these local orders. Sometimes the global order is *precise*, e.g., SV+VO results in the unique global order SVO; sometimes, it is *imprecise*, e.g., SV+SO results in either SVO or SOV.

A category consists of a list of lexical rules (*Lex-List*) and a list of syntactic rules (*Syn-List*). *Categorization* is a process of acquiring linguistic knowledge so that the local orders of lexical rules can be regulated. Agents can build up a category to associate some lexical rules with some syntactic rules, if these lexicons have the same semantic role and are found under the regulation of those syntactic rules' local orders. In addition, agents can merge categories the having same syntactic roles if those categories share some lexical members that are regulated by a common syntactic rule. These categorization mechanisms are similar to those in the Verb Island hypothesis, but are not restricted to verbs.

Agents acquire linguistic rules and syntactic categories during communication. During production, similar to ILM, agents occasionally create holistic rule to express the whole integrated meaning when they have insufficient compositional rules to encode every semantic item in the chosen integrated meaning. If compositional rules are available but the associated syntactic rules are insufficient to regulate the utterance, agents occasionally create new syntactic rules in order to do so. In comprehension, based on the available M-U mappings acquired in previous communications (*previous experience*) stored in a *buffer*, agents can detect recurrent semantic item(s) and utterance syllable(s) that occur in two or more M-U mappings, and then map them as compositional rules that are stored in their *rule lists*. This detection of recurrent pattern is discussed in detail in Gong et al 2005. After that, based on the previous experience, agents may categorize some lexical rules into categories, or reorganize the available categories by merging some of them.

2.2. *Communication*

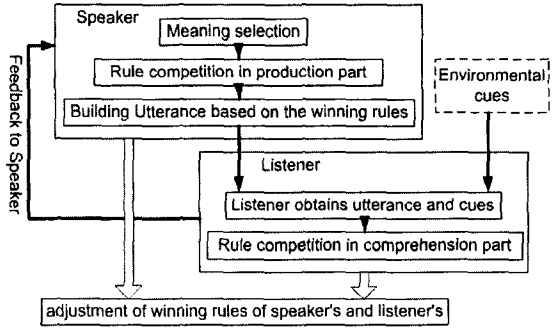


Figure 2. Information exchange in communications.

Each communication contains many rounds of information exchange (see Fig. 2). After choosing an integrated meaning based on its occurrence frequency, the speaker selects his winning linguistic rules based on their combined strength, builds up the utterance accordingly and sends the utterance to the listener. The listener receives the utterance and sometimes, some environmental cues. *Cues* are comprehension hints for the listener; the probability that one cue corresponds to the speaker's intended meaning is manipulated by the *Reliability of cues* (*RC*). Then, the listener selects the set of rules that allow her to comprehend an integrated meaning with the highest combined strength. The listener's calculation of the combined strength considers both linguistic rules and cues. Then, if the combined strength exceeds a certain *confidence threshold* (*CT*), the listener sends a positive feedback to the speaker and both of them reward their winning rules by increasing their strengths; otherwise, a negative feedback is sent and these rules are penalized.

There is no direct meaning check in the whole process; the strengths of linguistic rules that are rewarded sufficiently eventually exceed those of cues. Then, in order to comprehend utterances, agents will refer to these linguistic rules in comprehension, rather than the cues. Therefore, a preference for linguistic information is gradually established. At this stage, another feature of human language, *displacement* (Hockett 1960), is also established — agents are able to describe events not happening in their immediate space or time, and may still be understood. The model therefore simulates how linguistic communication turns into a reliable medium to exchange information, not just assistance to visual information to describe or discriminate simple concepts as shown in some models (e.g., Vogt 2005), though the assisting role may emerge earlier and might be a prerequisite for what we simulate here.

Syntactic categories play important roles for the emergence of global orders from local ones in communications, although categories themselves do not participate into the strength-based competition. In production, when choosing his candidate rules, the speaker first activates related compositional rules which can encode the chosen integrated meaning's semantic items. Then, considering these rule's syntactic categories, she activates some syntactic rules, the combination of which can resolve some global orders to regulate these compositional rules' syllables. Then, she judges whether these lexical and syntactic rules can win the strength-based competition. Similarly in comprehension, when choosing her candidate rules, the listener first activates some of her lexical rules whose syllables partially match the heard utterance. Some local orders can be detected based on these syllables' locations in the heard utterance. If some of these orders match the syntactic rules of categories to which these lexical rules belong, both the categories and syntactic rules are activated. Then, according to the categories, the semantic roles of the lexical items that are comprehended can be specified, particularly is the distinction of "Ag" from "Pt". Meanwhile, if some cue contains a meaning that corresponds to the comprehended meaning, its strength can participate into the calculation of combined strength of these rules. Then, the listener judges whether this set of rules can win the strength-based competition based on its combined strength.

3. Results and discussions

A semantic space containing 16 Type-I and 48 Type-II integrated meanings (consist of 4 "Ag"="Pt", 4 "Pr_1", 4 "Pr_2", the type ratio is 1:3) is used, in which the probability of choosing a Type-I or Type-II meaning are equal (by setting the token ratio of the two types to 3:1). Rule strengths are constrained to the interval [0.0, 1.0], the initial rule strength is set to 0.5 and the update step of strength is 0.1. 10 agents randomly communicate in pairs, and there are 20 information exchanges in one communication. The total number of communication is 3,000. Each agent initially shares 8 holistic rules but has neither categories nor syntactic rules. The buffer size is 40, and the rule list size is 60. After a certain number of communications, the *Rule Expressivity* (*RE*), *No. of common rules*, *Global Orders' Understandability* (average percentage of meanings comprehended using different global orders) and *Understanding Rate* (*UR*, average percentage of meanings accurately understood using linguistic rules only) are calculated.

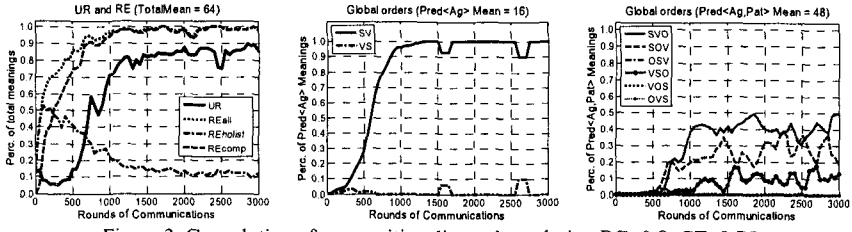


Figure 3. Coevolution of compositionality and regularity. RC=0.8, CT=0.75.

3.1. *Coevolution of compositionality and regularity*

The result of the simulation under the above parameters is shown in Fig. 3. In Fig. 3(a), the increase of the RE of the compositional rules is consistent with, but is not sufficient to prove, the transition from holistic language to compositional language. The understandability of the emergent language is traced by UR, which follows a U-shaped curve during the competition among holistic, compositional and syntactic rules and a sharp S-curve after some dominant orders are established and some lexical rules are shared. During the increase of UR, regularity emerges and some global orders become dominant (shown in Fig. 3(b)(c), agents use SV, SVO and SOV orders to comprehend most of the integrated meanings). The high value of UR (not RE) indicates the emergence of a compositional language which has both common compositional rules and consistent global orders. This emergent process occurs much later than the increase of RE, and indicates coevolution of compositionality and regularity, i.e., the acquisition of compositional rules and the formation of dominant orders boost each other, and are achieved simultaneously during language emergence.

4. Conclusions and future directions

This paper demonstrates the coevolutionary emergence of compositionality and regularity in a multi-agent system. A conventionalized language with common compositional materials and consistent syntax emerges out of holistic signals. Iterated communications in horizontal transmission provide opportunities for agents to get exposed to the utterances produced by others and to affect others with their own knowledge. Categorization shapes their categories to achieve a semantics-syntax match. This match can be imperfect, as still exists in modern languages, e.g., the Sex-Gender mismatch in German. The paper also implements a bottom-up syntactic developmental process in which complex syntax is formed based on local orders. How different sets of local orders affect the formation of global orders is discussed by Minett et al. in this volume and the linguistic or social factors that trigger the mismatch is under study.

Acknowledgements

This work is supported by grants from RGC Hong Kong: CUHK-1224/02H & CUHK-1127/04H. The authors thank Professors J. H. Holland, T. Lee, R. Chen and lab mates F. Wong & L. Shuai for useful discussions and suggestions.

References

Anderson, S. R., & Lightfoot, D. W. (2002). *The language organ*. Cambridge, MA: Cambridge University Press.

Bates, E., & Goodman, J. C. (2001). On the inseparability of grammar and the lexicon: evidence from acquisition. In M. Tomasello & Bates, E. (Eds.), *Language development: the essential readings* (pp. 134–162). Oxford: Blackwell.

Clark, E. V. (2003). *First language acquisition*. Cambridge, MA: Cambridge University Press.

Fried, M., & Östman, J. (Eds.) (2004). *Constructional approaches to language*. Amsterdam, Philadelphia: John Benjamins Pub.

Gong, T., & Wang, W. S-Y. (2005). Computational modeling on language emergence: a coevolution model of lexicon, syntax and social structure. *Language and Linguistics, 6(1)*, 1–41.

Gong, T., Minett, J. W., Ke, J-Y., Holland, J. H., & Wang, W. S-Y. (2005). A computational model of the coevolution of lexicon and syntax. *Complexity, 10(6)*, 50–62.

Hauser, M. D. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*, 88–96.

Jacob, F. (1977). Evolution and tinkering. *Science, 196(4295)*, 1161–1166.

Knight, C., Studdert-Kennedy, M., & Hurford, J. R. (Eds.) (2000). *The evolutionary emergence of language: social function and the origins of linguistic form*. Oxford, UK: Oxford University Press.

Minett, J. W., & Wang, W. S-Y. (Eds.) (2005). *Language acquisition, change and emergence*. Hong Kong: City University of Hong Kong Press.

O'Grady, W. (2005). *Syntactic carpentry: an emergentist approach to syntax*. Mahwah, NJ.: Erlbaum.

Oller, D. K., & Griebel, U. (Eds.) (2004). *Evolution of communication systems: a comparative approach*. Cambridge, MA: MIT Press.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterative learning: a framework for the emergence of language. *Artificial Life, 9(4)*, 371–386.

Steels, L. (2004). Constructivist development of grounded construction grammar. In D. Scott, W. Daelemans & Walker M. (Eds.), *Proceedings annual meeting association for computational linguistic conference* (pp.9–16).

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Vogt, P. (2005). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence, 167(1–2)*, 206–242.

AN EPISTEMOLOGICAL INQUIRY INTO THE 'WHAT IS LANGUAGE' QUESTION AND THE 'WHAT DID LANGUAGE EVOLVE FOR' QUESTION

NATHALIE GONTIER

Research Assistant of the Fund for Scientific Research – Flanders (F.W.O.-Vlaanderen), Centre for Logic and Philosophy of Science, Vrije Universiteit Brussel, Pleinlaan 2, Brussels, 1050, Belgium

Although Hauser, Chomsky and Fitch (HCF/FHC) and Pinker and Jackendoff (PJ/JP) differ in the epistemic questions they ask concerning, respectively, the nature of language (what language is), and the evolution of language (what language evolved for), it will be argued that both questions are part of the same methodological framework. This framework resembles the classical manner in which scientific knowledge is to be obtained while newer epistemological methods are suggested that can complement the study of the characteristics of language and evolutionary transitions that led to language.

1. Introduction

The current debate between, on the one hand, Steven Pinker and Ray Jackendoff (PJ, 2005; JP, 2005) and on the other, Noam Chomsky, Marc Hauser and Tecumseh Fitch (HCF, 2005, FHC, 2005), concerning (amongst other things) the original distinction made by HCF (2002) between the Faculty of Language in the Broad sense (FLB), and the Faculty of Language in the Narrow sense (FLN), is taken as point of departure in order to evaluate the different epistemic frameworks used by these authors.

Methodologically, two different paradigms, each with its own specific epistemic questions, can be distinguished. Proponents of the "language is an adaptation"-paradigm mainly focus on the question of *what* language evolved *for*, the most popular answer being, that it evolved for communication (Pinker & Bloom, 1990; PJ, 2005; JP, 2005). Adherents of the "(aspects of) language is (are) an exaptation"-paradigm basically study *what* language is, which aspects are uniquely human, and which cognitive domains are unique to/for human language. Here, it will be argued that a large part of the discussion revolves around a failure to acknowledge that each paradigm entails different goals when the *what* and the *what-for*-question are asked.

2. Aristotle and the nature of the scientific enterprise

In 350 b.C., Aristotle (1970: II 3) stated that to know something is to examine the 4 causes (namely, the material, the formal, the efficient, and the final cause) that underlie that something. These causes can be extracted by asking specific

questions: to know the true nature or essence of a thing, one needs to ask the what-question (that relates to the material and formal cause). To know how something came into being, one needs to ask the how-question, and to obtain the final goal of something, one needs to ask the what-for-question.

Important for the purpose of this article is the following: Aristotle carefully demonstrated that the what-question and the what-for-question are inextricably related, that is, the essence of a thing, and the final goal of a thing, converge. A human infant, for example, has the potential to grow into an adult member of society, and this is also the final goal of that infant, that its "potential" will become "actualized". Thus, even before the child is an adult, it has the goal of becoming one. Put differently, Aristotle's world view is a teleological one, namely, essence precedes existence (the goal is already known before it exists).

3. Which questions are central to functional and evolutionary biology?

It is general knowledge that with Newton's work, the teleological framework was replaced by a mechanistic one. With the rise of causality-thinking, only the how-question (that relates to Aristotle's efficient cause), that asks about the mechanism, is regarded as scientific. All other questions are rejected.

This is true for all scientific disciplines, except for biology (e.g. Gontier, 2004: 180-195). To investigate which questions are acknowledged as scientific within biology, we need to make a distinction between Neodarwinian theory (that relates to the Modern Synthesis, founded by the population geneticists) and Postneodarwinian theory (the selectionist models that developed after the founding of the Modern Synthesis). In the latter theory, the Modern Synthesis is expanded to include, for example, the "gene's eye view" (Dawkins, 1982).

In 1961, Ernst Mayr, a Neodarwinian, still clearly distinguished between functional biology and evolutionary biology. Functional biologists emphasize the important role that the how-question plays in biology. This in turn, means that the whole is divided into its component parts, and then the question is asked concerning how these parts function individually, and how they function together. When this is known, the whole is known. Mayr (1961: 360):

> "The functional biologist is vitally concerned with the operation and interaction of structural elements [...]. His ever-repeated question is 'How?' How does something operate, how does it function? [...H]is approach is essentially the same as that of the physicist and the chemist."

Things get more complicated, when we turn to evolutionary biologists and their inquiry. Here the why-question is still posed. Mayr (1961: 360):

> "The evolutionary biologist differs in his method [...] His basic question is Why?'. [...] It may mean 'how come?' but it may also mean the finalistic 'what for?'. It is obvious that the evolutionist has in mind the historical 'how come?' when he asks 'why?'".

The why-question, can be interpreted in two different ways (Gontier, 2005); it can be a question that is directed towards the past, or it can be directed towards the future. When the why-question is directed towards the future, this means that *why* becomes a synonym for the *what-for-question*, and thus here it relates to the Aristotelian final cause. Thus, when the what-for-question is raised, one does not merely investigate the goal/function/or utility of something, rather one also asks what the essence or true nature of that something is. This is the case, because ever since Aristotle, the what-for and the what-question are intertwined, and both questions together lead to a teleological approach.

Mayr, however, stresses that within biology, there is no place for the what-for-question. The why-question, can also be directed towards the past, and here it is equated with the how-question, namely how certain characteristics - especially adaptive characteristics- emerge historically in evolution of life. Hence, with Mayr, the how-question is elaborated upon to include not only the functioning of the component parts, but also the historical origin of certain traits.

*Post*neodarwinians, on the contrary, defend a wholely different viewpoint. We take Richard Dawkins' work as an example. The difference with Mayr is that Dawkins defends a position in which the methodology used by functional and evolutionary biologists is the same. Dawkins (2000: 17):

> "... what kind of explanation for complicated things would satisfy us. We have just considered the question from the point of view of mechanism: how does it work? [...] But another kind of question is how the complicated things came into existence in the first place. [...] the same general principle applies as for understanding mechanism."

Even more so, contrary to Mayr, Dawkins re-introduces the what-for-question, directed towards the future:

> "The theory of natural selection provides a mechanistic, causal account of how living things came to look *as if* they had been designed *for* a purpose. So overwhelming is the appearance of purposeful design that, [...] we still find it difficult indeed boringly pedantic, to refrain from teleological language when discussing adaptation. Bird's wings are obviously 'for' flying, [...]." (Dawkins, 1982: 161, emphasis added)

Just like Mayr, the importance of the how-question is emphasized by Dawkins, in order to, on the one hand, investigate the mechanical functioning of a certain trait, and on the other hand, to investigate the historical, evolutionary origin of certain characteristics. The difference between the two biologists arises when Dawkins stresses that questions concerning the historical and evolutionary origin also require a functional approach.

To be sure, Dawkins (e.g. 2000: 6-7) repeatedly emphasizes that natural selection does not work in a goal-oriented or teleological manner, and thus that the use of the what-for-question is not justifiable. However, he does not succeed in over-bridging a certain ambiguity that surfaces in his writings, concerning the

use of the what-for-question, an ambiguity that, as we shall discuss under 4.2., especially re-appears in the implementation of his theories by PJ. For now, it suffices to say that in Dawkins' work, the what-for-question remains posed together with the how-question.

4. How are these paradigms put to use in language (origin) studies ?

4.1. *The Chomskyan tradition and the essentialist "what is language"-question*

Chomsky's theory of generative grammar explains language as an innate human capacity, that forms an organ in the brain. The young Chomsky's main goal was to criticize behaviourism, an approach that preferred to investigate the external behaviour an organism displays. Chomsky, on the other hand, emphasized that language needs to be studied as an internal component of the brain, also called I-language.

> "The shift in locus was from the study of E-language to the study of I-language, from the study of language regarded as an externalized object to the study of the system of knowledge of language attained and internally represented in the mind/brain. [...] UG is a characterization of these innate, biologically determined principles, which constitute [...] the language faculty." Chomsky (1986: 24)

Although language needs to be studied from within biology –as the "bio-linguistics"-discipline dictates, this does not immediately or even necessarily imply that the human language faculty also needs to be studied from within *evolutionary* biology. On the contrary, in his earlier writings, Chomsky emphasizes that the LAD is uniquely human, and that a survey into the evolutionary origin of the LAD from within a *comparative* evolutionary framework would thus be meaningless. Chomsky argued that evolution had been saltational, leading to a qualitative difference, which excluded an explanation of the origin of the LAD by means of natural selection that follows a gradual pace.

The (Aristotelian) quintessential of language, that *what* makes language what it is, is the universal grammar, and this aspect becomes the principal focus, the central topic of investigation, in contrast to being concerned with *how* this faculty evolved. *"Hence, the logically prior task of elucidating precisely what evolved has taken research priority over elucidating how it evolved."* (Newmeyer, 2003: 60).

In other words, within this tradition, the what-is-language-question is posed as the most important question. In 2002, HCF (2002: 1569-71) specified even more what is uniquely human and what is thus essential to the human language faculty, by making a distinction between the FLB and the FLN. FLB encompasses the sensory-motor system, the conceptual-intentional system and the computational mechanisms required for recursion, while the FLN only

encompasses recursion. It is the FLN that is defined as uniquely human and thus as that *what* human language is. It is assumed that most (if not all) other elements of the FLB are shared with other animals and that, although perhaps necessary for language, they are not sufficient to cause human language on the one hand, or to define human language, on the other. Moreover, what is shared with other animals, is not a priori understood to have evolved in the course of evolution *for* human language (rather, an exaptationist explanation is preferred).

Hence, here we find that, contrary to earlier works of Chomsky, a comparative approach is pursued, albeit in a *via negativa* way:

> "[...] a basic and logically inelimenable role for comparative research on language evolution is this simple and essentially negative one: A trait present in nonhuman animals did not evolve specifically for human language, although it may be part of the language faculty and play an intimate role in language processing." (HCF, 2002: 1572)

Basically, within this paradigm, the comparative method seems to be useful only in so far that it distinguishes the unique properties of human language from those properties that we share with other animals, and hence to answer the "*what* is unique to human language-question", the answer being "recursion". Thus within this framework, it is assumed that it is possible to distinguish essential properties from accidental ones, the latter being necessary but not sufficient to produce human language or to explain the evolution of human language (not that recursion is either). This last point is rather important, and is, I think, overlooked in the recent criticism given by JP (2005) and PJ (2005). Namely that once we know what language is, what makes human language unique (i.e. recursion), we really do not know anything about the evolution of language, because according to HCF (2002) recursion did not cause the evolution of language, and language did not evolve *for* recursion, or in order to have recursion. Language just has recursion. Within HCF's theory, "recursion" merely is the answer one needs to give when it is asked what language is, not what language evolved for.

4.2. *The Postneodarwinist tradition and the "what did language evolve for"–question*

In this section, we turn to the *Post*neodarwinians, and here, too, as we shall see, the how-question becomes of secondary importance, this time because of the emphasis that is given to the "*what* did language evolve *for*"-question.

Although bio-linguistics, instructionist models and Neodarwinian models already flourished well before Pinker and Bloom's 1990 article, it was these latter authors who for the first time clearly introduced a selectionist account on the origin of language from within a *Post*neodarwinian tradition.

Their primary goal was to accomplish a synthesis between Chomsky's theories on the human language faculty and selectionist accounts, by stressing that natural selection can lead, gradually, to the evolution of such an adaptively

complex device. Inspired by Dawkins (1982, 2000), Pinker and Bloom (1990) state that language shows design, and that, therefore, language is a result of natural selection, because only language can explain such "complex design".

> "[...] natural selection remains the only evolutionary force capable of generating complex design, in which a feature of an organism (such as the eye or heart) has a non-random organization that enables it to attain an improbable goal that fosters survival and reproduction [...] (Pinker, 2003: 24).

Natural selection has positively selected for a module in the brain that carries the rules for universal grammar, and this adaptive trait is genetically underpinned (the *FOXP2* gene being just one example), hence natural selection can gradually go about its business (Pinker 2003).

Again following Dawkins, in evolutionary language research, too, the what-for-question is introduced. A partly physical character such as language (physical because it is related to brain structures or certain sets of genes, etc.) can only show "design" if it was selected *for* a certain function, and that function needs to be beneficial for the carrier in the struggle for existence. Thus, the primary question within selectionist frameworks becomes: *what* did language evolve *for*, to what *end*? Assuming that language is an adaptation, *what* is the evolutionary benefit that language gives to its carriers, so that it can be selected?

Within selectionist approaches, numerous answers have already been given, the most prominent are that language evolved *for* communicative and/or social reasons, an explanation also provided by PJ (2005: 223): *"... the language faculty evolved gradually in response to the adaptive value of more precise and efficient communication in a knowledge-using, socially interdependent lifestyle."* Language was selected for the adaptive advantages that the lexicon and the grammar (distinguishable in syntax, morphology and phonology) provide, because these elements form the design of language (Pinker, 2003).

4.3. *Adaptation or exaptation*

Dividing the FLB into its 3 different subsystems, HCF (2002: 1573) also raises the question of whether these different subsystems, each on its own, underwent a different evolution. If so, the possibility arises that each subsystem once fulfilled functions other than the ones it now performs in today's human language apparatus. If we explore this direction further, acknowledging that the subsystems did fulfil other functions, then we also need to take into account that these functions *evolved* for other reasons, and also were *selected* for reasons other than language, in the course of evolution. And this in turn means that language (FLB and/or FLN) is neither necessarily an adaptation, nor that language got selected for the enhancement of communication as PJ propose. According to HCF (2002), especially recursion can turn out to be the result of an exaptation, rather than an adaptation.

The difference between an adaptation and an exaptation was first introduced by Gould and Vrba in 1982. *"A feature is an adaptation only if it was built by natural selection for the function it now performs."* (Gould and Vrba, 1998: 53). Postneodarwinians assume that an adaptation involves the selection of a *function*, more specifically the function that it currently performs. According to PJ, the function of language is communication and this means that language can only be an adaptation if it actually was selected for its communicative possibilities (possibilities that need to be genetically based in order for natural selection to be able to do its work).

An exaptation can take on two different forms (Gould and Vrba 1998: 54-55): on the one hand, an adaptation can lose its current function and can obtain a different function, and this new function in turn can become the target of positive selection. On the other hand, exaptation can mean a trait, that never really had a function or that never really was an adaptation, can gain a function and become the target of positive selection. According to HCF, a large part of the FLB and also recursion evolved in an exaptationist manner.

5. Can we overcome traditional thinking concerning the nature and evolution of language?

Basically, there are thus two classic questions asked in current language origin studies. Within the Chomskyan tradition, there is a primacy of the what-question, which is mainly posed in order to obtain a proper definition of human language, and this in turn is obtained by looking for the difference between animal communication systems and human language. Within the PJ-tradition, there is a primacy of the "what did language evolve for" question, one possibility being communication in a knowledge-using socially based community. However, in neither of these traditions does the question of *how* language actually did emerge, receive the prior status that is should receive. The recent HCF-PJ discussion shows that this is the case because both traditions assume that *how* language evolved will be derived, either from the *what* language is, or from the *what* language evolved *for* question. That is why misunderstandings between the two traditions occur.

I have shown show that, beginning with Aristotle, both the what- and the what-for-question have been intertwined. Indeed, this is one of the major reasons HCF and PJ end up clashing with each other. But the most important question in research regarding the origin of language should be the how-question, which is the one we have discussed with the use of Ernst Mayr's work. This how-question, should be (re-)directed towards the past, not towards the future (because it is here that the what-for-question emerges). As such, this question should be integrated directly into language origin studies.

Acknowledgements

Sincere thanks to the FWO-Vlaanderen, the R&D-Department of the VUB, and the CLWF. A warm thanks also goes out to Roslyn Frank.

References

Aristotle. (1970). *Aristotle's Physics. Book I and II.* Oxford: Clarendon.

Chomsky, N. (1985). *Knowledge of Language.* Westport: Praeger.

Chomsky, N., Hauser, M. D., & Fitch, W. T. (2005). Appendix. The minimalist program. Unpublished, available at www.wjh.harvard.edu/~mnkylab.

Dawkins, R. (1982). Replicators and vehicles. In R. Brandon and R. Burian, Eds., *Genes, organisms and populations* (pp. 161-79). Cambridge MA: MIT.

Dawkins, R. (2000). *The Blind Watchmaker.* London: Penguin Books.

Fitch, W. T., Hauser, M.D., Chomsky, N. (2005). The evolution of the language faculty: clarifications and implications. *Cognition, 79 (2),* 179-210.

Gontier, N. (2004). *De oorsprong en evolutie van leven: 15 van het standaardparadigma afwijkende thesen.* Brussels: Vubpress.

Gontier, N. (2005). Introduction to evolutionary epistemology, language and culture. In N. Gontier, J.P. Van Bendegem and D. Aerts (Eds.), *Evolutionary Epistemology, Language and Culture* (pp. 1-29). Dordrecht: Springer.

Gould, S. J., & Vrba, E. S. (1998). Exaptation: a missing term in the science of form. In D. Hull and M. Ruse (Eds.), *The philosophy of biology* (pp. 52-71). Oxford: Oxford University Press.

Hauser, M., Chomsky, N., & Fitch, W. (2002). The faculty of language: what is it, who has it, and how did it evolve ? *Science, 298,* 1569-1580.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition, 97 (2),* 211-225.

Mayr, E. (1961). Cause and effect in biology. In E. Mayr (ed.), (1997), *Evolution and the diversity of life* (pp. 359-382). Harvard: Harvard University Press.

Newmayer, F. (2003). What can the field of linguistics tell us about the origin of language? In M. H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 58-76). Oxford: Oxford University Press.

Pinker, S., (2003). Language as an adaptation to the cognitive niche. In M. H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 16-37). Oxford: Oxford University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707-784.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition, 95,* 201-236.

MINIMALIST FOUNDATIONS OF LANGUAGE EVOLUTION: ON THE QUESTION OF WHY LANGUAGE IS THE WAY IT IS

WOLFRAM HINZEN

Department of Philosophy, Universiteit van Amsterdam, Nieuwe Doelenstraat 15, Amsterdam, 1012 CP, The Netherlands

I describe and assess the Minimalist Program (MP) as an approach to the evolution of language. The MP is less about evolution than explanation, but if its attempt to vindicate a certain idea of 'design perfection' was successful, a deeper level of explanation would be achieved than historical narrative and functional explanation affords, and the evolution problem would be solved along the way. Arguably, a minimalist methodology is also a necessary component in any explanatory approach to language evolution, no matter its theoretical orientation. While these are clear virtues, I question the MP's central explanatory claim, that language can be understood as an optimal solution to the problem of satisfying interface conditions imposed by pre-linguistic cognitive systems.

1. Minimalism and language evolution

1.1. *Beyond explanatory adequacy*

Describing variance in the world's languages is one thing. Explaining it is quite another. In the generative tradition, this task has taken the form of *deducing* it: the generative linguist has variance follow from (i) the interaction of a universal set of structural principles (laws) with a universal set of parameters and linguistic data, plus (ii) a lexicon specifying idiosyncratic (non-predictable) features of the language in question. If the principles and parameters were known (still a distant goal) and the deduction achieved, a form of 'explanatory adequacy' would be reached.

Yet, this would leave the question open why language is the way it is. A familiar answer to this latter question is that language evolved under particular adaptive constraints. There is or was some job that it did optimally, or more optimally at least than anything else did, and its structure reflects that task specification. Thus Jackendoff (2002, ch.8) proposes a gradualist evolutionary story with a number of incremental steps each of which has a rationale with regard to the 'goal' of increasing expressiveness in the communication of complex propositions.

Here the answer to our question is a historical, evolutionary, and functionalist one, and it may seem quite natural if we view, as the generative

tradition has done, the language faculty as merely another '"organ of the body", one of many sub-components of an organism that interact in its normal life' (Chomsky, 2005:1). On the other hand, this answer is more an answer to the question of why language exists rather than why it is the way it is. On a Darwinian view, structures prove adaptive without arising on adaptive grounds: adaptive function does not rationalize origins or design primitives. Necessarily, natural objects will exhibit signs of design and adaptation, but this will be a consequence of evolution, not a cause.[a] It is a prevailing logical error that function accounts for genesis, or internal causal mechanisms, hence can be a cause of evolutionary novelty (Hinzen, 2006, Müller, 1990).

1.2. *Answering 'why'-questions, MP-style*

Consider that, in general, the evolution of some organismic structure, X, need not actually be essential to understanding its design. For any evolved organic system, the language faculty included, we can distinguish three factors that enter into its growth and development (Chomsky, 2005): (i) external data, (ii) genetic endowment, (iii) general principles of structural architecture not specific to the system and/or the species in question.

Which of these factors is explanatorily dominant with respect to some particular system is an entirely empirical question, with consequences for the instrumental role that natural selection actually played. Furthermore, it is arguably a conceptual necessity that any language organ newly inserted in a primate or pre-linguistic cognitive architecture has to 'satisfy interface conditions' imposed by pre-existing cognitive systems with which language will have to interact to be used. One natural such condition is that expressions generated by the language faculty have to be *legible* at all interfaces it forms with existing systems. In addition to (i-iii), then, we can consider a further causal factor, (iv) principles relating to the way the system has to interface with other systems already in place to be usable at all by them.

Suppose now that the language faculty *precisely has* those features it *has* to have to be usable at all – plus features of a *domain-general* sort that are not specific to humans or even organic nature as such – and has *no features besides*. Then we *understand why* it is the way it is. We will not have to look *further* for such an explanation in external factors like adaptational benefits, contingent evolutionary pathways, or brain physiology (Hinzen, 2006). In this sense, any

[a] As Walsh et al. (2002) argue, it is a purely statistical notion relevant to the re-grouping of populations with variants already there.

feature in category (iii) that the language faculty plausibly *has* to satisfy in order to be usable at all, or that is not specific to it (category iv), will, in Chomsky's terms, have a 'principled explanation'.

Note that neither an explanation by appeal to (iii) nor to (iv) invokes evolution or history in any intrinsic way. Indeed it could reflect the vision of a super-human engineer handed the task of designing the language faculty from scratch: surely, he would use general principles coming 'for free', and then try building the language system into the brain so as to interface with systems already in place in an optimal way, without any further, redundant structure. The more we can show that the language faculty is the work of an engineer reasoning in some such fashion, the less we will regard path-dependent evolution, the tinkering of Jacobian 'bricolage', as essential to it.

All that would be without being 'anti-evolutionary', indeed because it would simply be a pursuit of a kind of explanation deeper than historical narrative or functional explanation. If some natural object has only properties that are arguably *necessary* (or else not specific to it), this means we are not *puzzled* by its properties any more, and we will feel much less incited to inspect history for the pathways and contingent adaptive pressures that led to this object.[b]

The task of principled explanation in the sense that the MP arises with respect to *any* theoretical account of the language faculty. A 'minimalist' theorist, whatever his account of the language faculty, will in each case proceed by asking for the *barest essentials* that are needed for a descriptively and explanatorily adequate account of language and its various components. The more minimal our descriptive apparatus becomes, the less has to evolve for language, hence the explanatory burden is reduced, and questions of principled explanation for the remaining features can be asked.

As has been stressed at nauseam for this very reason, the MP and its explanatory vision is something one can choose to be interested in or not, but it is not a theory one can accept or reject. As a non-theory specific approach to language evolution, it indeed has given room to quite different visions on the minimization task (see e.g. Culicover and Jackendoff's 2005 minimalist vision).

In sum, it is a mistake, both to say that the MP is a controversial theoretical account of language evolution, and to say that 'why'-questions *need* to have a Neo-Darwinian evolutionary (and functionalist) answer (Pinker and Jackendoff

[b] If, in discussions of language evolution a 'lack of data' is explored, we can see the MP as making the point that in the specific case of language data may actually not be *needed*, and might not actually tell us much even if found: the system is as such already understood, on what are arguably principled grounds.

2005). There is no *a priori* constraint to that effect, and the functionalist argument, according to which language was not merely selected, but positively shaped by natural selection for some function and explainable in these terms, must be argued on entirely empirical grounds.

1.3. *Functional explanation in practice*

How good the prospects for such an argument in the case of language are – even leaving general doubts about the conceptual coherence of functional explanations aside – remains questionable. The more we find that the structural principles of language are domain-general and not specifically human, they won't be selected for (let alone explained by) the functions that they specifically enable in humans and in the domain of language. And the more we can motivate the structures of language that we empirically find in terms of external constraints imposed on language by an existing pre-linguistic cognitive architecture, the less we will have to explain features special to it that it would otherwise exhibit. As a *system*, language may possibly still be an 'adaptation' in some sense, but that rationale won't transfer to the parts it recruited in evolution for its novel purpose.

Consider again Jackendoff's (2002:ch.8) detailed attempt to provide a functional rationale for language, the communication of thought. That communication should be of a unique or even special significance in the study of language would be surprising, since any species whatsoever communicates in one way or another. What makes human language special is clearly not that it is a communication system but that it is a linguistic one. Moreover, it is clear that even though virtually anything can be used for a communicative purpose, as Pinker and Jackendoff (2005) concede, the thoughts that can be conveyed by language are of a very special kind: they are propositional thoughts, and there is simply no necessity that thoughts (nor communicated ones or specifically adaptive ones) would ipso facto be propositional, as a look at animal communication systems, or logically possible ones, immediately shows. This raises again the doubt that these special kinds of thought contents may actually be at least partially *due* to language, as opposed to being the reason why it exists.

Yet, it is the latter claim that the functionalist makes. But that language should have evolved 'for the communication of complex propositions' is surprising in the light of the foregoing, since it assumes that *pre*-linguistic conceptual-intentional systems are actually propositionally structured. But what we see is that the derivation of each and every expression that can be used in a self-standing utterance and can express a propositional thought involves

transformational syntax (technically, head-, A- and A'-movement). Thus, *the killing of that man*, or *was killed that man*, which do not involve transformations, also do not express a propositional, truth-evaluable thought. If that is of any significance, and it seems to be, it questions the plausibility that thought is fully propositional in creatures lacking transformational (context-sensitive) syntax (presumably, all non-human creatures).

Moreover, as Carstairs-McCarthy (1999) as argued, there are many different syntactic formats that language might have had, without any negative implications for its expressive power or informativeness. E.g., languages might universally not have had two major lexical categories (Nouns and Verbs), but only Nouns, which have argument structures and involve predicational relations as well. In fact they have had no categories at all. In that case, no sentence/Noun Phrase distinction would have evolved, and propositionality in the sense of the ability to think truth-bearing thoughts expressed by sentences, wouldn't have either.

It is thus simply unclear how we can hope explaining language evolution as a gradual increase in expressiveness with respect to the goal of expressing complex propositions. In the absence of language, there is plausibly no such goal set. The explanation is simply circular in the absence of comparative evidence that thoughts of non- or pre-linguistic animals are propositional in nature (has contents of a sort we specifically express by means of sentences). Comparative evidence that there is seems to point in the exactly opposite direction (Terrace, 2005).

Syntactic constraints in general seem largely unmotivated, if not unintelligible, in the light of the putative 'goal' of language, the expression of complex thought.[c] Nor need perfectly grammatical expressions be readily usable, which they often are not. Language design, for all we can tell, is not optimized for *use*. The MP vision of design optimality is of a radically different nature: linking pre-linguistic sensory-motor and conceptual-intentional systems (the prime task of syntax according to Chomsky's version of the MP) *can* well be optimal, in the sense of exploiting a minimal set of grammatical mechanisms and invoking no levels of representations apart from interface levels, without the result being readily usable, as performance factors intervene.

---

[c] Most major syntactic constraints (locality, subjacency, the Case-filter, the Minimal Link Condition, etc.) forbid the use of numerous expressions that would, were it not for their ungrammaticality, quite interpretable in a communicative context. Syntax is as much a hindrance to communication as an aid to it.

2. Some cracks in the foundations of the MP

2.1. *Interface conditions don't motivate all of syntax*

All that said, the very idea of 'motivating' properties of syntax from the need to 'meet interface conditions' can be objected to virtually on the same grounds on which Jackendoff's functionalism was criticized above. It is to start with a *non sequitur* to conclude, from the fact that a system not satisfying certain interface conditions will therefore not be usable, that the properties that will make it usable will therefore come into existence.

Moreover, consider a pre-linguistic primate or hominid mind that has a rich conceptual-intentional (C-I) system. Then suppose the language system arises as the MP proposes, as a method of *linking* this system to a phonological one. The computational system (syntax), CS, of language, will then be centrally be motivated by the need to meet (in particular) the demands of the C-I system externally imposed on it. But what syntax do we predict this method will induce?

Possibly, none. If human syntax is motivated by interface conditions imposed by outside systems (systems internal to the mind, but outside the linguistic system), then the syntax resulting from and explained by this can only be as rich as these very outside systems. That means that *either* the C-I system involves relevantly structured mental representations, or they are non-syntactic. In the former case, appealing to the C-I system won't explain these structures.[d] In the latter, what kind of information will be 'legible' at this interface? Presumably, only information that does not demand a grasp of compositionally interpreted phrasal structures generated by the syntactic system. In the limit, the only CS 'satisfying interface conditions' would be a system having no syntax at all, consisting of lexical items in certain relations alone. Put differently and more positively, if the 'thoughts' a non-linguistic animal can think do not reach up to those complex meanings that humans express by means of syntactically complex phrases and sentences, syntax itself will play an explanatory role for what kind of complex meanings we can think, and the entire explanatory vision of the MP will be in jeopardy.

---

[d] Even positing merely verbal argument structure there ('who does what to whom' structure), if relevantly like argument structure in language, might be circular, as argument structure may itself be syntactically conditioned (as per Hale and Keyser 2002).

2.2. *The fate of hierarchy in MP*

We shouldn't be surprised then to actually find, if we look at current Minimalist practice, a tendency to reduce syntax as much as possible to no syntax, i.e. to eliminate structural complexity where it is not plausibly 'motivated from interface conditions'. Collins (2002) in particular has reduced the result of combing a verb like *kill* and a noun like *Bill* by the operation Merge to the set {kill, Bill}. The Merge of *Jill* yields the nested set {Jill, {kill, Bill}}, which can now merge with T(ense), and so on. The nested recursive structure that arises is formally similar to the von Neumann construction of ordinal numbers out of a single initial object, the empty set ∅, which we may call '1'. Merge yields the set {∅}, call it '2', to which we can apply the same operation again to get the set {∅, {∅}}, which may be called '3', etc.

Chomsky (2005) has used this very analogy to suggest that one-place Merge applied to a single lexical item, ∅, yields the 'minimal language'. His own proposal differs from Collins',[e] but even in his version of phrase structure, it is not clear how hierarchy gets off the ground. Just as we would expect, given the general methodology of 'motivating syntax from interface conditions', only nested sets labeled by lexical items and trivially interpreted by some analogue of the successor operation, not categorial projections and ordered hierarchies of them, ever arrive at the interface.

It is unclear here whether we have explained language, or explained it away. Forming phrase structural hierarchies is not only required to license the verb's arguments, but also correlates with emergent semantic effects and novel semantic entailments, none of which seem plausibly purely semantic or non-linguistic. Thus, when a causative verb X combines with its internal argument Y, a novel integrated event forms with an integral participant and a telic structure which nothing in the lexical items X and Y as such or their set predicts: e.g., *clearing the screen* intrinsically implicates that *the screen clears*, which in turn intrinsically ends with the *screen clear*.

Phrase structure as minimalized above won't explain why this is so. Traditionally, what syntax construed as above on analogy with the successor

---

[e] Namely, that the results of Merge are not merely sets but labeled ones. However, as a consequence of the 'Inclusiveness' requirement, which makes all structures assembled by CS and delivered at the C-I interface a mere re-organization of lexical features with no other features added in the course of the derivation, the result of the phrasal projection of *kill*, or the label of the resulting set, would be actually identified with *kill*, i.e. be a lexical item again: Merge thus yields the set {kill, {kill, Bill}}.

function in arithmetic yields is the 'syntagmatic' dimension of language, whereas the hierarchies we have just mentioned – e.g. causative verbs contain subevents, achievements contain accomplishments contain activities contain states – belong to the 'paradigmatic' core of language. While systematic, it has been a moral of the downfall of generative semantics that they are not systematic in the same way that syntagmatic processes are. Another, paradigmatic, syntax is needed.

3. Conclusion

In sum, even if we accepted recent minimalist accounts of Merge that eliminate phrasal projections or identify them with lexical items, then another syntax has to be added to them (for a suggestion, see Hinzen and Uriagereka, to appear). It is no clearer here than in the case of propositionality how this syntax would follow from anything in non-syntactic C-I-systems, or in the structure of the outside world. 'Principled explanation', if wedded to the idea of 'satisfying interface conditions', may be at a far more distant horizon than the current MP suggests, as crucial parts of the core syntax of language may follow from nothing other than that syntax itself.

References

Carstairs-McCarthy, A. (1999). *The Origins of Complex Language*, Oxford: OUP.

Culicover, P. & R. Jackendoff (2005). *Simpler syntax*, Oxford: OUP.

Chomsky (2005), 'On Phases', Ms.

Collins, C. (2002). Eliminating Labels. In S. Epstein and D. Seely (Eds.), *Derivation and Explanation in the Minimalist Program*, 42-64. Oxford: Blackwell.

Hauser, M., Chomsky, N., & W. T. Fitch (2002). The Faculty of Language: What is it, who has it, and how did it evolve? *Science* 298, 1569-1579.

Hinzen (2006). *Mind design and minimal syntax*. Oxford: OUP.

Hinzen, W., J. Uriagereka (to appear). On the metaphysics of linguistics, Ms.

Jackendoff, R. (2002). *Foundations of Language*. Oxford: OUP.

Müller, G. B. (1990). 'Developmental mechanisms at the origin of morphological novelty: A side-effect hypothesis'. In M.H. Nitecki (Ed.), *Evolutionary Innovations* 99-130. Chicago.

Pinker, S., and R. Jackendoff (in press). 'What's special about the human language faculty?', *Cognition*.

Terrace, H. (2005). Metacognition and the evolution of language. In Terrace&Metcalfe (Eds.), *The missing link in cognition*, 84-115. Oxford University Press.

Walsh, D.M., T. Lewens, and A. Ariew (2002). The Trials of Life: Natural Selection and Random Drift. *Philosophy of Science* 69, 452-473.

WHY HAS AMBIGUOUS SYNTAX EMERGED?

STEFAN HOEFLER

Language Evolution and Computation Research Unit, University of Edinburgh
40 George Square, Edinburgh EH8 9LL, Scotland, UK
stefan@ling.ed.ac.uk

Ambiguity is a defining property of natural language distinguishing it from artificial languages. It would seem to be dysfunctional, and therefore its ubiquity in language poses an evolutionary puzzle. This paper discusses the implications of a typical iterated learning model on the conditions under which syntactic ambiguity emerges and stabilises in language. It contrasts the purely nativist stance that language imperfections such as syntactic ambiguity are artifacts arising from internal constraints of the genetically determined language faculty with the view that they are frozen accidents persisting because they are easily learnt.

1. Introduction

Ambiguity is a striking property of natural language distinguishing it from artificial languages. The mathematically well defined case of *syntactic* ambiguity is present in sentences that can be structurally analysed in more than one way:

(1) The officer watched the spy with the telescope.

(2) The word of the Lord came to Zechariah, son of Berekiah, the prophet.

(3) The passengers who left the boat first were old men and women.

In an optimal communication system, a feature like syntactic ambiguity would seem to be dysfunctional. Effective coding implies that one signal corresponds to one meaning and can therefore be interpreted deterministically. Syntactic ambiguity, one would assume, should not be found in a successful communication system like human language. Its ubiquity thus poses an evolutionary puzzle: why has such an apparent imperfection emerged in language?

One view of language imperfections is that they are artifacts arising from *internal* constraints of the innate language faculty (Chomsky, 2002). Nativists conclude from the poverty of stimulus argument that language acquisition is not primarily a matter of learning but rather of setting predefined, genetically determined parameters of the innate language faculty.

This view is contrasted by an approach taken in studies of language evolution which explain hallmarks of language by the fact that language undergoes cultural

transmission (Hurford, 2002; Brighton, Kirby, & Smith, 2005). This idea is based on the observation that language acquisition represents a special class of learning problem as the output of the language learning of one generation is the input to the learning of the next generation. Properties of language are exhibited because language itself, as opposed to its users, adapts to be learnable. Brighton (2003) points out that in such an iterated learning framework, language imperfections reflect residues of linguistic evolution through cultural transmission.

This paper introduces syntactic ambiguity as an example of an imperfection of language and discusses the implications of a typical iterated learning model on two questions raised by the evolutionary puzzle that comes with its ubiquity. Under what conditions does syntactic ambiguity emerge? Despite its ubiquity, only a sparse number of grammatical rules in human language are actually involved in syntactic ambiguity. The second question must therefore ask why syntactic ambiguity has been prevented from becoming more pervasive.

The remainder of this paper falls into three sections. In the first, I present a computational model to study the emergence and transmission of syntactic ambiguity. The subsequent section describes observations made in our simulations and illustrates them with three example experiments. These results are summarised and discussed in the last section of the paper.

2. An Iterated Learning Model of Emergent Syntactic Ambiguity

Our intuitions about the behaviour of complex dynamic systems, and any verbal theorising built on them, tend to be faulty. On the other hand, the formalisation of analytical mathematical approaches for such systems (Nowak & Komarova, 2001) proves to be difficult. Computer simulations therefore offer a useful alternative to study the evolution of language (Cangelosi & Parisi, 2001). To this end, Kirby and Hurford (2002) have developed the Iterated Learning Model (ILM). Iterated learning models have been applied to simulate how pivotal properties of language such as recursive syntax or compositionality can originate from cultural transmission (Kirby, 2002). Roberts, Onnis, and Chater (2005) have presented a simplicity-based model to explain the emergence of quasi-regular constructions. They point out that in such a model, the transmission bottleneck is a necessary prerequisite for the emergence of linguistic idiosyncrasies.

In the model presented here, two agents are 'alive' at any one time: a speaker and a learner. Each agent is endowed with an induction algorithm to infer context-free grammars from linguistic data it has observed, and with the ability to produce sentences from that grammar. There is no biological evolution in the agents. The speaker produces a certain number of sentences from its internalised grammar. The learner uses its induction algorithm to infer a grammar on the basis of the speaker's output. The number of sentences the learner is exposed to constitutes the learning bottleneck through which language is transmitted. After one iteration, the learner becomes the new speaker, a new learner is created and the process is

started again.

The algorithm for grammar induction applied in this model is described in detail in Kirby (2002) and ultimately based on Wolff (1982). However, apart from minor simplifications, two noteworthy modifications have been made to the original algorithm. First, our algorithm is not enriched with any *explicit* semantics. The notion of syntactic ambiguity is intrinsically rooted in the principle of compositionality, which says that the meaning of a complex expression is a function of the meaning of its constituents and the way they are combined. Different syntactic structures correspond to different complex meanings. The meaning distinctions caused by *syntactic* ambiguity are therefore implicitly expressed in the compositional structures assigned to a sentence, if we presuppose the principle of compositionality for a model.

The second major modification of Kirby's original algorithm is that grammar induction in our model is not necessarily deterministic. Multiple hypothetical grammars can arise in a learner where rule subsumption can be carried out in more than one way. This will be illustrated in section 3. The induction algorithm also produces permutations of the original order in which the linguistic data was presented to the learner and induces alternative grammar hypotheses from these. In multi-generation simulations, hypothetical grammars are either selected for simplicity or for maximal expressivity.

The notion of expressivity adopted in the model corresponds to the number of structurally distinct sentences a grammar can produce. Counting distinct syntactic structures, rather than distinct strings, entails that each interpretation of an ambiguous sentence contributes to the expressivity of its grammar separately. A different syntactic structure also yields a different compound meaning according to the principle of compositionality.

Simplicity as included in the model is based on a non-probabilistic notion of the Minimum Description Length (MDL) principle (Rissanen & Ristad, 1994). The MDL of context-free grammars is calculated according to the methods set up for regular grammars in Teal and Taylor (2000). In its MDL condition, the model is thus very similar to the one presented in Zuidema (2003).

Our model of the learner does not acquire vocabulary or induce lexical categories. We take them to be already learnt. Terminal symbols in the examples of the following section are therefore to be thought of as lexical (or basic syntactic) categories rather than individual words.

3. Example Experiments

This section describes the behaviour observed in the simulations. I will first discuss *single learning* simulations, which were carried out to study the conditions under which structural ambiguity *emerges* in a typical model like ours. In a second step, we will analyse the *stabilisation* of ambiguous grammars in *iterated learning* simulations, where language is transmitted through a bottleneck over generations.

3.1. *Emergence of Syntactic Ambiguity in a Single Learning Model*

Single learning simulations have shown that a range of factors influence the emergence of structural ambiguity during grammar induction. I will use a simple example language to visualise the behaviour of these simulations. The initial example input consists of the data *ba, bca, bab, bac*. After the encounter of the first two strings *ba, bca*, the learner has incorporated the rules $S \rightarrow ba$ and $S \rightarrow bca$ to its grammar. These rules can be compressed by subsumption in two different ways, and hence induction yields two different hypothetical grammars at this stage of the learning process:

$S \rightarrow Xa$	$S \rightarrow bX$
$X \rightarrow b$	$X \rightarrow a$
$X \rightarrow bc$	$X \rightarrow ca$

We track the further development of the left-hand side hypothesis grammar. The next string *bab* is generalised by replacing the substring *b* with the already established non-terminal symbol *X*. The new rule $S \rightarrow XaX$ is incorporated and compressed by subsumption with the existing rule for *S*:

$S \rightarrow XY$
$Y \rightarrow a$
$Y \rightarrow aX$
$X \rightarrow b$
$X \rightarrow bc$

When encountered, the last string *bac* is generalised to $S \rightarrow XYc$. It is this generalisation and the subsequent rule subsumption which introduce ambiguity to the grammar:

$S \rightarrow XZ$
$Z \rightarrow Y$
$Z \rightarrow Yc$
$Y \rightarrow a$
$Y \rightarrow aX$
$X \rightarrow b$
$X \rightarrow bc$

This final grammar can generate 12 structurally different sentences. The ambiguous strings *babc* and *bcabc* are produced in two different ways where the substring *abc* is either structured as *(a(bc))* or as *((ab)c)*. Such a grammar can account for ambiguous constructions like the ones in the English example sentences (1)–(3), if we replace its symbols with appropriate phrasal and lexical categories. If we track the right-hand side hypothesis above, we obtain an unambiguous grammar with a 6 sentences expressivity.

Running such experiments, we have been able to identify conditions under which syntactic ambiguity occurs in a model like the one presented. There is one

necessary prerequisite for the emergence of syntactic ambiguity: generalisation during grammar induction. No ambiguous grammar emerges without generalisation during the process of its acquisition. But in a typical iterated learning model, learners have to apply generalisation because of the transmission bottleneck. In such a model, syntactic ambiguity is thus one possible implication of the cultural transmission of language.

The likelihood that an ambiguous grammar will be induced is dependent on an interaction of (1) the *properties of the input strings* and (2) the *properties of the induction algorithm.* Structural ambiguity is the result of several coinciding rules induced from the data, rather than something a single learning event would elicit.

We observe that in cases where the hypotheses for a set of linguistic data comprise ambiguous as well as unambiguous grammars, the ambiguous grammars are more expressive than the unambiguous grammars. Similarly, more expressive grammars are more likely to be ambiguous than their less expressive counterparts. Cursorily viewed, this could lead to the assumption that ambiguity would increase over generations in an iterated learning model in which hypothesis grammars are selected for maximal expressivity. We will see below that this is not the case.

3.2. *Stabilisation of Syntactic Ambiguity in an Iterated Learning Model*

The described learner is placed in an iterated learning simulation to analyse the stabilisation of syntactic ambiguity over generations. The number of sentences heard by each generation has proven to be critical in all simulations. For the simulations described here, the bottleneck size was chosen such that the example language is stable if the learners select their hypothesis grammar for MDL. Since the induced ambiguous hypothesis grammars are usually of higher MDL than their unambiguous counterparts, the grammars evolved under such conditions are mostly unambiguous.

The same conditions were then applied to simulations in which the learners select their hypothesis grammars for maximal expressivity. Fig. 1 charts the emergence and stabilisation of syntactic ambiguity in three such simulations, based on the initial example input data *ba, bca, bab, bac.* The grammars evolved under these conditions stabilise in either minimally ambiguous or unambiguous form.

If expressivity was the only pressure to have an impact on language evolution, the given conditions would lead to highly productive and ambiguous languages. Undoubtedly, such languages would be positively dysfunctional and impose insurmountable problems on communication. But, as Brighton et al. (2005) observe, a language is stable if it is expressive *and learnable.* Remember that in our model the pressure for expressivity is realised by the selection of the most expressive grammars. Learnability on the other hand is implemented by the bottleneck through which languages are transmitted.

The example simulations in Fig. 1 illustrate the observation that languages stabilise at a relatively low level of expressivity. Highly expressive languages

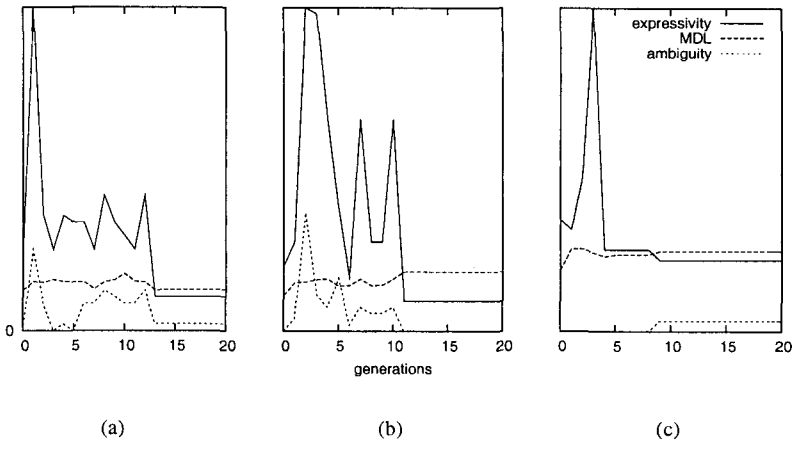


Figure 1. Emergence and stabilisation of syntactic ambiguity in three example simulations, relative to the expressivity and simplicity (MDL) of the evolved grammars. The dotted line denotes ambiguity, the dashed line MDL and the solid line expressivity. The y-dimensions of the lines are relative to each other and therefore not indicated in absolute values where they are > 0. Hypothetical grammars are selected for maximal expressivity in each generation. Due to the learning bottleneck, the evolved grammars stabilise on a low level of expressivity within the first 20 generations of 1000 in total. They are either minimally ambiguous or unambiguous. (a) The evolved stable grammar is minimally ambiguous. (b) The evolved stable grammar is unambiguous. (c) A minimally ambiguous stable grammar evolves from unambiguous unstable grammars.

cannot pass the learning bottleneck successfully and do not reach a stable state. We witness the impact of two competing pressures on the evolution of language: expressivity and learnability.

We can distinguish three types of behaviour in the simulations. The example case in Fig. 1(a) shows how a stable *minimally ambiguous* grammar emerges within the first 20 of 1000 generations of the simulation. The stable grammar emerges after a significant decrease in expressivity and ambiguity. This behaviour can also be observed in the example experiment in Fig. 1(b), where a stable *unambiguous* grammar evolves under the same conditions. We conclude from the equal distribution of these two types of behaviour that minimal syntactic ambiguity does not constitute an impediment to the successful transmission of a language. In cases represented by Fig. 1(c), a stable minimally ambiguous grammar evolves from unstable unambiguous grammars. Expressivity is slightly lowered at the moment of the introduction of ambiguity in generation 9. We have seen before that

ambiguity tends to occur in more expressive hypothetical grammars induced from the *same* data. However, in this example it is introduced during the transmission of language from one generation to the next. Syntactic ambiguity occurs in a grammar that is less expressive than the one of the previous generation. This increases the learnability of the example language and stabilises it.

4. Discussion and Conclusion

We have found evidence that *in a typical iterated learning model*, the same phenomenon, the learning bottleneck, is responsible for the two evolutionary puzzles set out in the introduction of this paper. Given its dysfunctionality, why has syntactic ambiguity emerged? And given its persistence in human language, what has prevented syntactic ambiguity from becoming more pervasive? Our simulations suggest that the answer to both questions is the bottleneck through which language is transmitted in such a model.

In our *single learning simulations* illustrated by example experiments, syntactic ambiguity emerges during grammar induction due to coinciding properties of the data and the learning algorithm. The necessary precondition for its emergence is the process of generalisation. Learners need to generalise during language acquisition because they are only exposed to a limited set of linguistic data. The presented type of grammar induction thus implies that syntactic ambiguity reflects a residue of an accidental but not improbable coincidence in the evolution of language through iterated learning.

Does our model therefore oppose the nativist view of language imperfections? It seems that, like any iterated learning model, it takes an intermediate stance. It is potentially nativist in the explanatory emphasis it puts on the specifics of the learning algorithm. But at the same time, it adds a complementary explanatory process, cultural evolution through iterated learning.

The observations made in *iterated learning simulations* suggest that against a pressure for expressivity, the transmission bottleneck ensures that syntactic ambiguity does not become too pervasive in a language once it has emerged. The stable ambiguous grammars emerging from the simulations only exhibit sparse ambiguity. The evolutionary pressure on language to be learnable prevents it from becoming too productive and therefore also constrains ambiguity. Strikingly, the learning bottleneck seems to enable us to explain why syntactic ambiguity is not more pervasive despite its persistence in language. Under the presented assumptions, we can thus disregard such notions as ease of disambiguation or communicative dysfunctionality. The example experiments in this paper illustrate how the stabilisation of ambiguity can be subject to the fluctuation caused by two competing pressures on the evolution of language, learnability and expressivity. In a typical iterated learning model, both pressures are exhibited due to the learning bottleneck through which language is transmitted.

Acknowledgements

Stefan Hoefler was supported by Cogito Foundation Research Scholarship 101/04.

References

Brighton, H. (2003). *Simplicity as a driving force in linguistic evolution.* Unpublished doctoral dissertation, The University of Edinburgh, Edinburgh.

Brighton, H., Kirby, S., & Smith, K. (2005). Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In M. Tallerman (Ed.), *Language origins: Perspectives on evolution.* Oxford: Oxford University Press.

Cangelosi, A., & Parisi, D. (2001). Computer simulation: A new scientific approach to the study of language evolution. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 3-28). London: Springer.

Chomsky, N. (2002). *On nature and language.* Cambridge University Press.

Hurford, J. R. (2002). Expression/induction models of language evolution: dimensions and issues. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models.* Cambridge University Press.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models.* Cambridge University Press.

Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: An overview of the Iterated Learning Model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148). London: Springer.

Nowak, M. A., & Komarova, N. L. (2001). Towards an evolutionary theory of language. *Trends in Cognitive Sciences, 5*(7), 288-295.

Rissanen, J., & Ristad, E. S. (1994). Language acquisition in the MDL framework. In E. S. Ristad (Ed.), *Language computation.* Philadelphia: American Mathematical Society.

Roberts, M., Onnis, L., & Chater, N. (2005). Acquisition and evolution of quasi-regular languages: Two puzzles for the price of one. In M. Tallerman (Ed.), *Language origins: Perspectives on evolution.* Oxford: Oxford University Press.

Teal, T. K., & Taylor, C. E. (2000). Effects of compression on language evolution. *Artificial Life, 6*(2), 129-143.

Wolff, J. G. (1982). Language acquisition, data compression and generalization. *Language and Communication, 2*(1), 57–89.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems 15.* Cambridge, MA: MIT Press.

PROTO-PROPOSITIONS

JAMES R HURFORD

Language Evolution and Computation Research Unit,
University of Edinburgh
Adam Ferguson Building, George Square
Edinburgh, EH8 9LL, Scotland, UK.

Before the evolution of languages as public conventional communication systems, pre-humans had somewhat complex private mental schemes for representing the external world. What is known about human and some animal vision suggests that proposition-like cognitive structures existed for the mental representation of perceived scenes before the advent of complex language. The structures traditionally adopted by formal Logic can be modified to conform to known constraints on the visual representation of scenes. While this modification slightly reduces the expressive power of representations (in that the meanings of some complex sentences cannot naturally be represented), it provides a unified, ontologically parsimonious, primitive notation for cognitive representations, suitable for later recruitment by complex syntactic language. The most basic semantic elements later mapped onto sentences are all present in the prelinguistic mental representation, which reflects the workings of the visual attention system

1. The Magical Number 4 – How big is a single thought?

First order predicate logic (FOPL), not being an explicitly psychological theory, places no upper bound on the number of entities that can be involved in a simple proposition. Thus, beside natural 1-place, 2-place, and 3-place predicates, in principle 14-place, even million-place predicates are countenanced by the metatheory of FOPL. In practice, however, the examples discussed by logicians hardly ever involve predicates of degree greater than 4. This is in fact a psychologically natural limit, conforming to the limits on subitizing, the fast process of global visual attention which can take in a very limited number of objects at one glance. The limit in humans is around 4.

Cowan (2001) has argued that the real limit on human short-term memory is also "the magical number 4", and not seven plus or minus two, as earlier argued by Miller (1956). The human subitizing limit of 4 is shared by animal species relatively closely related to humans (Dehaene 1997).

If simple clauses in human languages are taken to correspond to propositions describing simple scenes, the typical size and scope of human simple clauses can be explained by this limitation on visual attention and short-term memory, shared with non-linguistic creatures. (The complex sentences of modern languages, with embedding and coordination of simple clauses, are not the issue here. The evolutionary basis for simple clauses is an issue independent of the evolution of processes which build more complex structures with them.)

2. Predicate-Argument Structure in Animal Brains

Propositions, as theorized by logicians, are structured asymmetrically, consisting of elements of two distinct kinds, a predicate and one or more arguments. The semantic functions of predicate and arguments differ. Predicates are associated with properties and relations, whereas arguments typically refer to individual objects. The combination of a predicate and its arguments results in a formula with a truth value, which depends on whether the property or relation denoted by the predicate holds of the object(s) referred to by the arguments. As argued in Hurford (2003), neural correlates can be found in human and animal visual (and auditory) systems of these two logical types. One visual mechanism, the so-called 'dorsal stream' (or the 'where stream') pre-attentionally picks out objects (up to about four at a time) and assigns mental indices to them, for the purposes of tracking them in the scene. The dorsal stream, apart from thus indexing objects in the scene, does not categorize them in any way. Once visually indexed, objects can be the targets of focal attention, and at this stage properties are assigned to them by processes in the ventral stream (or 'what stream'). Thus one brain mechanism picks out objects – this corresponds to the referring function of logical arguments. Another, separate, mechanism, assigns properties to objects. This division of labour by the visual system is shared by many mammals. The visual representation of a scene is the overall complex of neural activity delivered by the interacting dorsal and ventral streams. The represented scene is composed by a process delivering objects as arguments and a process delivering predicate-like judgements about the properties of those objects. This visual system forms a natural evolutionary platform for representations of meanings, subsequently expressed by human sentences. This whole argument is presented in detail in Hurford (2003).

3. Local and Global Attention to Objects and Scenes

Humans and some animals are capable of parallel attention to whole scenes (global attention) and to the objects within scenes (local or focal attention). And humans and some animals (e.g. pigeons) can shift back and forth rapidly between local and global attention. This activity exhibits a kind of two-level representation. For instance, a pattern of dots forming a rectangle can be simultaneously represented as both a rectangle and a collection of dots. In a pathological condition known as simultanagnosia, patients 'cannot see more than one object at a time'. Such a patient cannot simultaneously see both a rectangle

formed with dots and the individual dots (for example not being able to count the dots), and cannot see the relations between multiple objects in a scene. This is a failure in the normal coordination of global and local attention. Normal humans can simultaneously categorize a scene by its global properties, and assign properties to the constituent objects. It seems likely that other primates, and many other vertebrates, have visual attention organized in this way, with a division between global and local attention and an ability to shift between them and combine their outputs into a single representation. This, as will be seen in a later section, suggests a parallel with current event-based semantics, and can lead the way to a re-thinking of traditional logical representations.

4. Animal Truth and Reference

Truth and reference are paradigmatic qualities of human propositions. Can the prelinguistic mental representations of animals refer, or be true? Yes. Some non-human animals have mental categories. This is most obviously shown by their systematic behaviour in relation to various coherent classes of input stimuli from the environment. Most neuroscientists would talk of patterns of neural activity, rather than mental categories; but if these patterns of neural activity are categorially distinct (which they usually are, with a little fuzziness at the edges), this is only a terminological difference. While some philosophers have been reluctant to attribute concepts to non-humans, ethologists have usually had no such qualms where it can be shown that the relevant behaviour cannot be accounted for in terms of simple triggering mechanisms. Herrnstein (1991) defined a scale with five points, and labelled the fourth point on the scale 'concepts'. He concluded that many animals showed evidence of having concepts, but that only humans reach his fifth point, the level of abstract relations. More recent work indicates that some non-humans can sustain representations long after the triggering stimulus has faded (e.g. Zuberbühler et al. 1999), and can represent abstract relations. In some cases, animals have been shown to control complexities beyond simple recognition of the concrete properties of objects. Thus, there is evidence in some animals for primitive versions of metacognition (sensitivity to ones own mental states), and control over more abstract relational concepts, such as 'same-colour-as'.

An animal can hold an object in (focal) attention. The relationship between the external object and the brain mechanism directing and controlling attention to it is the basis for the later evolution of deictic reference in communicative acts, like that involved with the English demonstrative pronouns *this* and *that*. When such pronouns are successfully used in conversation about concrete things, as when one says "This is hot!", both speaker and hearer attend to the same external object. It seems reasonable to call the relation between the brain mechanism of a non-linguistic creature attending to an object a relation of **proto-reference** to the object in question. What is lacking in proto-reference, as

opposed to deictic reference in language, is an accompanying communicative signal. Ballard et al. (1997) use the term *deictic* for the relation between mental variables tracking objects and the objects themselves. More complexly, humans can perform parallel "attentive tracking" of a limited number of objects in a scene. Theoretical models of this capacity describe it as involving the mental assignment of spatial tags to targets; see Culham et al. (1998). Here again it is appropriate to call the relation between the mental tags and the external objects 'proto-reference'. Neuroscientists sometimes apply the term *refer* to this relation between a mental process and an external object. Pylyshyn (2000:205), for example, writes "... the visual system ... needs a special kind of direct reference mechanism to refer to objects without having to encode their properties."

The properties assigned to an attended-to object, combined with the mental index of the object in the scene, constitute a proposition-like representation, in which the evolutionary beginnings of **truth** can be discerned. The truth of human linguistic utterances depend on conventionally coordinated assignments of denotations to predicates. If I point to an object and say "That is a mouse", the truth of my utterance depends on what class of objects is normally associated in this society with the word *mouse*. Uncommunicative creatures have no such shared conventions, but it is possible to appeal to the idea of a single adult creature's usual patterns of behaviour in relation to classes of objects which we human researchers can identify as nondisjunctive natural classes. If such a move can be justified, we can maintain that non-human animals can have true or false beliefs. The cat that chases a wind-blown leaf may well be mentally categorizing the leaf as a mouse. That is, the neural activity normally reserved for mouse-like things gets 'wrongly' triggered by the wind-blown leaf. In this case, it seems reasonable to say that the representation in the cat's brain is false.

In claiming that an animal acts as if X were true, or that it believes X, two points must be acknowledged: firstly, no linguistic representation of X is present in the animal's mind, and secondly, it follows that whatever form of words we choose to describe X is only an approximation to the conditions influencing the animal's behaviour. Only an impossibly omniscient external agent can infallibly judge the truth or falsity of propositions in creatures' minds. We scientists can approach the fact of the matter whether animals have true or false beliefs on particular occasions, based on our systematic observations of their normal responses to categories of objects which we have reason to believe they perceive more or less as we do. See Dennett (1996)'s section on 'The Misguided Goal of Propositional Precision' for concurring arguments.

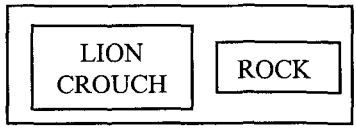
5. A parsimonious *Begriffschrift* for proto-propositions

Frege introduced a notation for the representation of concepts. His graphically baroque notation has since been translated (without loss or increment) into the

more familiar representations of FOPL. For Frege, concepts were not psychological, due to a phobia of 'psychologizing', understandable at a time when so little was known about the brain. However, given modern knowledge of the workings of the visual system, we are now in a position to devise a notation for the mental representation of visual scenes, comparable to Fregean FOPL. The notation proposed is comparable, preserving what is apt about the Fregean conception, but rejecting what is psychologically implausible.

It is useful to adopt a minimal element of iconicity in the proposed notation. The central idea is to replace any individual variable by a box, and to write predicates applying to that variable inside the box. Thus, the following two representations are equivalent, both translatable as *There's a crouching lion and a rock.*

$\exists x \exists y$ LION(x) & CROUCH(x) & ROCK(y)



The outer box represents the whole scene, the object of global attention. As will be seen shortly, predicates can be applied to the whole scene as well as to the individual objects in it.

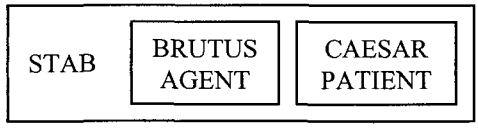
6. Getting Rid of Individual Constants

As argued elsewhere (Hurford 1999, 2001, 2003), individual constants, logical terms corresponding (roughly) to proper names in languages, such as *Queen Elizabeth II* and *Saul Kripke*, are psychologically implausible. In Logic, such individual constants are assumed to have an absolutely rigid designation, recovered by users with absolute reliability. This demands an impossible omniscience on the part of users. Even animals who can 'recognize' other individuals with impressive reliability can be easily fooled by experimenters. The alternative is to treat proper names as predicates. Thus MARY and JOHN, for example, have the same status in prelinguistic mental representations as WOMAN and MAN, but are more specific as a matter of degree, not of kind. In extensional terms, MARY picks out members of a subset of the set picked out by WOMAN. MARY does not necessarily identify a singleton set; it would include her identical twin as well.

7. Getting Rid of Ordered Arguments and Role Markers

Logic sequentially orders the arguments of a predicate, as a device for making it clear what roles the denoted individuals play in asymmetric relations. But the individuals in a visual scene are simultaneously present, not ordered in any sequential way. Linguists indicate 'who did what to whom' by Thematic role labels such as AGENT and PATIENT. This obviates the need for notational ordering of the arguments of a predicate. It is, further, common practice in event-based semantics (e.g. Parsons (1990)) to treat these Thematic role markers as predicates, and to label a whole event with a separate predicate, often corresponding to a verb in language. Translating a typical event-semantic representation into our box notation, we get the following equivalents.

$\exists e$ STAB(e) & AGENT(brutus, e) & PATIENT(caesar, e)



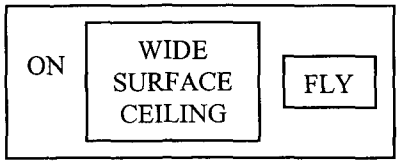
In a common way of paraphrasing, these would represent a scene describable by *There was a stabbing event in which Brutus was the Agent and Caesar the Patient*, or just plain *Brutus stabbed Caesar.* Note the ontological parsimony of the box representations, with only boxes, corresponding to individuals and the single whole scene, and one-place predicates, corresponding to the attributes assigned to the individuals and to the scene. In particular, in the box notation, the property of AGENThood is attributed to Brutus, within the scope of the overall stabbing scene.

AGENT here is our shorthand for a bundle of perceptual features distinctive of animate objects in vigorous action, including 'biological motion'. Biological motion is recognized by human infants (Johansson, 1973), and by day-old chicks (Vallortigara et al., 2005). All that is claimed in using the term AGENT is that the observer of a scene can tell which of the participants is "doing something".

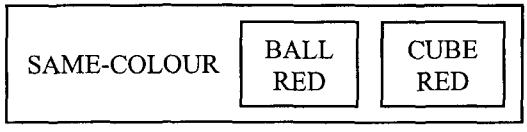
A claim made here, the most radical departure from FOPL, is that **all** prelinguistic predicates were one-place. Some justification for this is given in the next section.

8. One-Place Predicates over Scenes and Objects

Predicates traditionally treated as two-place or three-place can be treated as one-place, given an implicit **relativizing function** of the outer scene-representing box. Only a few examples can be given here. The following diagram represents a scene consisting of a fly on a ceiling.



Here is a diagram representing a scene of which it is judged that two objects are of the same colour.



It is hypothesized that predicates appearing in the outer box are delivered by global attention.

The ordering of elements within all boxes is immaterial. It will be seen that the proposal that all predicates are one-place is, though strictly true, not as radical a departure from tradition as may be thought. Some predicates, though applying to single scenes (thus technically one-place) only apply to scenes which contain a specified number of objects. Thus, for example, the judgement represented by SAME-COLOUR cannot be arrived at with a scene containing less than two objects. Some predicates, often corresponding to those to which an emerging syntax will associate transitive verbs and prepositions, only apply to scenes with more than one object. But there is no exact correspondence between later-emerging syntactic categories and particular types of prelinguistic predicates. When public labels get attached to such internal private representations, for purposes of communication, other purely public considerations, such as Topic and Comment, play a part in the structuring of sentences.

References

Ballard, D.H., Hayhoe, M.M., Pook, P.K. & Rao, R.P.N. (1997), Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences,* 20(4), 723-742.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences,* 24 (1), 87-114.

Culham, J.C., Brandt, S.A., Cavanagh, P., Kanwisher, N.G., Dale, A.M. & Tootell, R.B.H. (1998). Cortical fMRI Activation Produced by Attentive Tracking of Moving Targets. *Journal of Neurophysiology,* 80, 2657-2670.

Dehaene, S. (1997). *The Number Sense.* New York: Oxford University Press.

Dennett, D. (1996). *Kinds of Minds.* London: Weidenfeld and Nicolson.

Herrnstein, R. (1991). Levels of categorization. In G. Edelman, W. Gall, & W. Cowan (Eds.), *Signal and Sense*, pp. 385–413. Somerset, NJ: Wiley-Liss.

Hurford, J. R. (1999). Individuals are abstractions. *Behavioral and Brain Sciences,* 22(4) 620-621.

Hurford, J. R. (2001). Protothought had no logical names. In J.Trabant & S.Ward (Eds.), *New Essays on the Origin of Language*. pp. 119-132. Berlin: de Gruyter

Hurford, J. R. (2003). The Neural Basis of Predicate-Argument Structure, *Behavioral and Brain Sciences*, 26(3),261-283.

Johansson, G. (1973), Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics,* 14, 201–211.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 63, 81-97.

Parsons, T. (1990). *Events in the Semantics of English: a Study in Subatomic Semantics.* Cambridge, MA: MIT Press.

Vallortigara, G., Regolin, L. & Marconato, F. (2005), Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biology,* 3(7), e208. (Public Library of Science, online journal).

Zuberbühler, K., Cheney, D.L., & Seyfarth, R.M. (1999). Conceptual semantics in a non-human primate: *Journal of Comparative Psychology,* 113(1), 33-42.

CONVEX MEANINGS AND EVOLUTIONARY STABILITY

GERHARD JÄGER

Department of Language and Literature, University of Bielefeld,
PF 10 01 31, D-33501 Bielefeld, Germany
Gerhard.Jaeger@uni-bielefeld.de

Gärdenfors (2000) argues that natural denotations of natural language predicates are convex regions in a conceptual space. Using techniques from evolutionary game theory, the paper shows that this convexity criterion is a consequence of the evolutionary dynamics of language use.

Evolutionary game theory (EGT) is a mathematical framework to model the consequences of interaction between individuals for the evolutionary dynamics of a population. Suppose a certain type of interaction between members of a population has an impact on their fitness. Furthermore, the outcome of the interaction depends on heritable traits ("strategies") of the individuals involved. Under these conditions, the interaction can be modeled as a strategic game, where utility can be identified with fitness.

One of the appealing features of this model is the fact that the attractor points of the ensuing dynamics (the "evolutionarily stable strategies") can be characterized purely in terms of the utility matrix. This is particularly straightforward in the case of asymmetric games. There, exactly the **strict Nash equilibria** are evolutionarily stable (as proved by Reinhard Selten in 1980). A pair of strategies is a strict Nash equilibrium if each strategy is the unique best response to the other strategies.

This model is applicable both to biological and to cultural evolution. Applied to the cultural evolution of language, the individuals involved are speakers/hearers-in-an-utterance, strategies are linguistic constructions, and fitness is the likelihood of a construction to be imitated. We expect to find natural languages in evolutionarily stable states most of the time. The notion of evolutionary stability is thus a way to reduce linguistic universals to the evolutionary dynamics of language use.

Gärdenfors (2000) argues the semantic domains that natural language deals with have a geometrical structure. He gives evidence that simple natural language adjective usually denote natural properties, where a natural property is a **convex** region of such a "conceptual space". In this paper I will argue that under very

natural assumptions about the utility function, convexity of meanings falls out as a consequence of evolutionary stability.

Imagine a simple communication game. The game leader, "Nature", shows one player (the sender S) a point in some continuous Euclidean space. S can send one out of a finite set of signals to the receiver R. R in turn has to guess the point that Nature showed to S.

The problem has the shape of a signaling game. Nature chooses some point in the meaning space, according to some fixed probability function p_i. S and R have the joint goal to maximize the similarity between Nature's choice and R's choice.

Formally, we say that M is the set of meanings (points in the meaning space), S is a function from M into some finite set F of forms, and R is a function from F to M. The utility of the communicators can be defined as

$$u(S, H) = \sum_{m \in M} p_i(m) \cdot sim(m, R(S(m))) \tag{1}$$

where *sim* is a function that measures the similarity between two points. Similarity between two points is a monotonically decreasing function of their distance in the conceptual space. By convention, similarity is always positive, and every point has a similarity of 1 to itself. The interests of S and R are completely identical. Also, the signals themselves come with no costs.

Suppose S knows R, the interpretation function of the receiver. What would be the best coding strategy then? For each possible signal f, S knows R's interpretation, namely $R(f)$. So for a given choice m of Nature, S should choose the signal f that maximizes $sim(m, R(f))$. In other words, each form f corresponds to a uniqe point $R(f)$, and each input meaning m is assigned to the point which is most similar to m, or, in other words, which minimizes the distance to m. This kind of partitioning of the input space is called a **Voronoi Tesselation**. It is easy to show that the Voronoi tessellation based on a Euclidean metric always results in a partioning of the space into convex regions.

Since the meaning space is continuous, there may be pairs of functions from M to F that, though different, differ with a probability 0. (For instance if they only differ with respect to a single point.) If we identify sender strategies with equivalence classes of such functions rather than with such functions itself, there is actually a unique best response to each receiver strategy—the equivalence class of the Voronoi tessellation that is induced by that receiver strategy.

The best response of R to a given coding function S is given by.

$$R(f) = \arg\max_m \int_{S^{-1}(f)} p_i(m') sim(m, m') dm' \tag{2}$$

The precise nature of such an optimal sender's response depends on the details of Nature's probability function p_i, the similarity function, and the geometry of the

underlying space. Suffice it to say that for a closed space and continuous p_i and *sim*, the existence of such a best response is always guaranteed.

From these considerations it follows that in each game of the described class (with a finite Euclidean meaning space and continuous similarity functions and probability distributions), there are evolutionarily stable states, and **in each evolutionarily stable state, the meaning of each form is a continuous region of the meaning space.**

For the purpose of illustration, I did a few computer simulations of the dynamics described above (using a discrete approximation of a continuous space). The meaning space was a set of squares inside a circle. The similarity between two squares is inversely related to its Euclidean distance. All meanings were assumed to be equally likely. The experiments confirmed the evolutionary stability of Voronoi tesselations. The graphics in Figure 1 show stable states for different numbers of forms. The shadings of a square indicates the form that it is mapped to by the dominant sender strategy. Black squares indicate the interpretation of a form under the dominant receiver strategy.

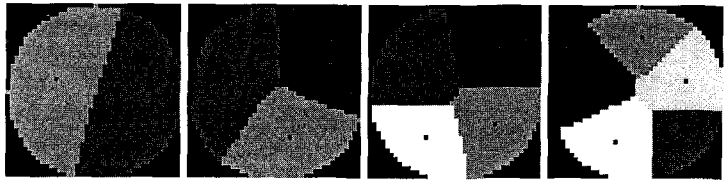


Figure 1. Evolutionarily stable states of the signaling game with a uniform probability distribution over meanings

In the previous simulations, I assumed a uniform probability distribution over the meaning space. This leads to an infinity of evolutionarily stable states. In the remainder of the paper I will illustrate with an example that a skewed probability distribution may reduce the number of equilibria quite drastically.

Let us assume that the meaning space contains finitely many, in fact very few, small regions that are highly frequent, while all other meanings are so rare that their impact on the average utility is negligible. For the sake of concreteness, let us suppose that the meaning space forms a circle, and that there are just four meanings that are frequent. Let us call them Red, Green, Blue and Yellow. Inspired by the middle plane of the color space (supressing the brightness dimension), I assume that all four prominent meanings are close to the periphery, and they are arranged clockwise in the order Red, Yellow, Green, and Blue. Their positions are not entirely symmetric. Rather, they are arranged as indicated in Figure 2.

Since the similarity between two points is inversely related to their distance, it follows that Blue and Green are more similar to each other than Red and Yellow, which are in turn more similar than the pair Green/Yellow and the pair Red/Blue.

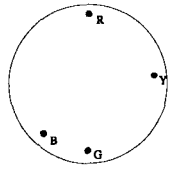


Figure 2. Schematic arrangement of the four prominent meanings in the example

The pairs Blue/Yellow and Red/Green are most dissimilar.

We finally assume that the probabilities of all four prominent meanings are close to 25%, but not completely equal. For the sake of concreteness, let us assume that

$$p_i(\text{Red}) > p_i(\text{Green}) > p_i(\text{Blue}) > p_i(\text{Yellow}) \tag{3}$$

Now suppose the sender has just two forms at her disposal. What are the strict Nash equilibria of this game?

A Voronoi tesselation induces a partition of the set {Red, Yellow, Green, Blue}. The partition {Red, Green}, {Blue,Yellow} is excluded because it is not convex. This leaves seven possible partitions:

1. {**Red, Yellow, Green, Blue**}/{} This is a weak Nash equilibrium; i.e., one of the strategies involved is not the *unique* best response to the other player's strategy. It is thus not evolutionarily stable.

2. {**Red**}/{**Yellow, Green, Blue**} If the sender strategy partitions the meaning space in this way, the best response of the receiver is to map the first signal to Red and the second one to the point that maximizes the average similarity to the elements of the second partition. If the probabilities of Yellow, Green and Blue are almost equal and all other points have a probability close to 0, the average similarity to Yellow, Green and Blue is a function with three local maxima, that are located close to Yellow, Green and Blue respectively. So if the sender uses this partition, the best response of the receiver is to map the first form to Red and the second to the point with the highest average similarity to Yellow, Green and Blue. This is one of the mentioned three local maxima. Since, by assumption, Green is more probable than Yellow and Blue, the maximum close to Green is the global maximum. But this entails that the sender strategy is not the best response to the receiver, and thus this partition does not lead to evolutionary stability.

3. {**Yellow**}/{**Red, Green, Blue**} For similar reasons as in the previous case, this partition thus does not correspond to a Nash equilibrium either.

4. {**Green**}/{**Red, Yellow, Blue**} Since Blue is closer to Green than to Red, this partition does not correspond to an equilibrium for analogous reasons.

5. {**Blue**}/{**Red, Yellow, Green**} This case is analogous because Green is closer to Blue than to Red.

6. {**Red, Yellow**}/{**Green, Blue**} The best response of the receiver here is to map the first form to Red and the second to Green. The best response of the sender to this strategy in turn is to map Red and Yellow to the first form, and Green and Blue to the second. So this partition creates a strict Nash equilibrium.

7. {**Red, Blue**}/{**Yellow, Green**} The best response of the receiver here is to map the first form to Red and the second to Green. The best response of the sender in turn would be to to map Red and Yellow to the first form, and Green and Blue to the second. Hence this partition does not induce a Nash equilibrium.

So it turns out that with two forms, only the bipartition {Red, Yellow}/{Green, Blue} is evolutionarily stable.

Let us turn to the analogous game with three forms. Each sender strategy in this game creates a tripartition of the meaning space. We only have to consider convex tripartitions. All convex bipartitions are trivially also tripartitions, with an empty third cell. It is also immediately obvious that such a partially trivial partition cannot give rise to a strict Nash equilibrium. Besides, there are four more, non-trivial convex tripartitions:

1. {**Red**}/{**Yellow**}/{**Green, Blue**} The best response of the receiver is to map the first signal to Red, the second to Yellow, and the third to Green. The best response of the sender to this strategy is to use the above-mentioned partition, so this leads to a strict Nash equilibrium.

2. {**Yellow**}/{**Green**}/{**Blue, Red**} This does not correspond to a Nash equilibrium because the best response of the receiver is to map the third form to red, and since Blue is closer to Green than to Red, the best response of the sender would be to switch to the previous partition.

3. {**Green**}/{**Blue**}/{**Red, Yellow**} The best response of the receiver is to map the three forms to Green, Blue, and Red respectively, and the best response of the sender in turn is to use the Voronoi tessellation that is induced by these three points. This is exactly the partition in question, so it does lead to a strict Nash equilibrium.

4. {**Red**}/{**Blue**}/{**Yellow, Green**} Since Yellow is closer to Red than to Green, this does not lead to a Nash equilibrium either.

So in this game, we have two partitions that are evolutionarily stable, namely {Red}/ {Yellow}/{Green, Blue} and {Green}/{Blue}/{Red, Yellow}. There is an asymmetry between the two equilibria though. Recall that the evolutionary model assumes that the strategy choice of the players is not fully deterministic but subject to some perturbation. Suppose the system is in one of the two evolutionarily stable

states. Unlikely though it may be, it is possible that very many mutations occur at once, and all those mutations favor the other equilibrium. This may have the effect of pushing the entire system into the other equilibrium. Such an event is very unlikely in either direction. However, it may be that such a switch from the first to the second equilibrium may by more likely than in the reverse direction. This would have the long term effect that in the long run, the system spends more time in the second than in the first equilibrium. Such an asymmetry grows larger as the mutation rate gets smaller. In the limit, the long term probability of the first equilibrium converges to 0 then, and the probability of the second equilibrium to one. Equilibria which have a non-zero probability for any mutation rate in this sense are called *stochastically stable*.

Computer simulations indicate that for the game in question, the only stochastically stable states are those that are based on the partition {Red}/{Yellow}/{Green, Blue}. In the simulation, the system underwent 20,000 update cycles, starting from a random state. Of these 20,000 "generations", the system spent 18,847 in a {Red}/{Yellow}/{Green, Blue} state, against 1,054 in a {Green}/{Blue}/{Red, Yellow} state. A switch from the first into the second kind of equilibrium did not occur.

Figure 3 visualizes the stable states for the game with two, three and four different forms. As in Figure 1, the shade of a point indicates the form to which the sender maps this point, while the black squares indicate the preferred interpretation of the forms according to the dominant receiver strategy. The circles indicate the location of the four focal meanings Red, Yellow, Green and Blue.

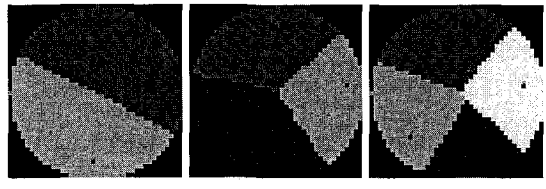


Figure 3. Evolutionarily stable states of the signaling game with focal points

The last example sketched a hypothetical scenario which induces the kind of implicative universals that are observed in natural languages with regard to color universals. My choice of parameters was largely stipulative. However, it is hoped that psycholinguistic research can supply empirically justified values for them. Evolutionary consideration could then establish a link between psycholinguistics and typology.

References

Gärdenfors, P. (2000). *Conceptual spaces.* Cambridge, Mass.: The MIT Press.

NATURAL-LANGUAGE "CHEAP TALK" ENABLES COORDINATION ON A SOCIAL- DILEMMA GAME IN A CULTURALLY HOMOGENEOUS POPULATION

MARK JEFFREYS
Behavioral Science Dept., Utah Valley State College, University Pkwy., Orem, UT, 84103, USA

ChickenHawk is a social-dilemma game in which the only way to win is to play "Hawk" against "Chicken." The purpose of the game is to distinguish between uncoordinated and coordinated self-sacrifice. In a test of four signaling conditions with players who belong to a culturally homogeneous population, a "cheap talk" condition led to efficient coordination, whereas signaling opportunities engaging social reputation and allowing eye-contact without speech yielded poorly coordinated altruistic behavior. The implications are: (1) without language, mere willingness to cooperate on a social dilemma is insufficient for coordinating intentions, and (2) given a sufficiently cohesive social group, language can coordinate inequitable, altruistic sacrifices of modest but real material incentives, even where fully anonymous defection is an option.

1. Motivation

In their coda to Christiansen and Kirby's (2003) *Language evolution*, Komarova and Nowak appeal for applications of game-theoretic studies of nonkin cooperation to research on the evolution of language, asserting gnomically that, "we speak because we cooperate, we cooperate because we speak" (p. 336). Taking up the challenge of that koan-like chiasmus, this study game-tested the extent to which successful cooperation depends on speech in a modern population, given a naturalistic social environment but a material conflict of individual interest. The idea was to pit natural language against "costly signaling" on a clear-cut social dilemma where only defectors could gain rewards, but any reward required partners to exactly match one partner's defection to the other's sacrifice. For any resources at all to be extracted required both cooperation and coordination. Two questions arose. First, how much coordination of intention to cooperate can face-to-face signaling effect without speech? Second, how much cooperation can verbal agreements sustain given both an incentive and anonymous opportunity to defect? Answers might clarify how "we cooperate because we speak," or, instead, suggest that language is akin to a flashy feature bolted on to an already functional system, enabling new tricks.

2. Theory

If language is a "discrete-combinatorial system" like DNA or binary code, as Pinker (1994, 1998) has described it, then the content of its signals cannot be considered inherently self-guaranteeing in the sense spelled out by the Zahavian (1997) theory of "handicap" or "costly" signal selection. There are no inherent

falsehood production costs. Drop the word "no," and the previous sentence becomes both false and easier to produce. Indeed, this feature of language is part of universal folk wisdom, encapsulated in myriad adages. "Talk is cheap." "Don't just talk the talk; walk the walk." "Names will never hurt me."

The prior assumption of this study, then, both from costly-signaling theory and from experimental evidence of indirect recipocity and reputation anxiety shown by nonkin human participants in economic games (e.g., Brandt et al., 2003, Kings-Casas et al., 2005, Milinski et al., 2002, Nowak & Sigmund, 1998), was that discrete-combinatorial language could not have bootstrapped the evolution of uniquely human habits of layered, opportunistic groupishness, often involving nonkin cooperation. These behaviors require costly signaling of quality and intent. Their core signal is always, "Invest in me; don't mess with me."

Language can encode this message, but can it vouch for it?

3. Population

With apologies to Sartre, the human evolutionary environment of adaptation is other people. Geography and food-gathering methods matter less than cultural familarity and the sense of shared membership in determining levels of nonkin cooperation. For that reason, this study sampled a modern population that, despite living in a postindustrial consumerist society, reasonably approximates the typically close-knit ancestral social context.

None of the participants identified themselves as close genetic kin, but all were adult college students. More than 80% preferred to identify themselves as active members of the Church of Jesus Christ of Latter-Day Saints (Mormons). This percentage is roughly equal to that of active Mormons in the college's student body and in the local county. A majority of participants had lived most of their lives in Utah and been born within 100 miles of the college. Many of the men had already served their 2-year Mormon missions. All surveyed participants chose the cultural identity "American" and most chose to describe their ethnic identity as "White" or "Caucasian."

In short, the participants in this study came from a loosely defined cohort within a cohesive cultural group. They had many common experiences of major rituals, group educational hurdles and life-history stages, as well as a strong sense of their shared, distinct sociocultural identity and of a history rooted in their religious geography, which incorporates familiar details of their surrounding landscape. In that sense, they formed a sample of an entirely "natural" human population, however modern, field-tested in their home habitat.

4. Methods

4.1. ChickenHawk Game Design

Pairs played a two-person, one-shot game for an entry in a $100 prize drawing. The choices were to demand or cede the entry. If both players demanded or both

ceded, neither won anything. The only way to win an entry, therefore, was to demand and hope that one's partner ceded that sole opportunity. Players made their choices using preprinted, numbered tickets as ballots: a large, red ticket constituted a demand; a small, blue ticket meant the player had opted to cede the opportunity. (See Fig. 1.)

	Column Player	
	Blue	Red
Blue	0, 0	0, 1
Row Player		
Red	1, 0	0, 0

Figure 1. 2x2 ChickenHawk Matrix.

The simplicity of the game and the fact that to defect is each individual's resource-maximizing strategy makes the game easy to teach. The single-shot format removes the possibility of reciprocally exchanged payoffs. The lottery shields defectors' anonymity, and, as the maximum possible entries in each lottery is quite small, anyone who gains an entry has a good chance of winning. Spite is also in play, as a defection can deny an entry.

4.2 Basic play

A small group of participants (6-12) gather in the same room. All are volunteers and have been paid an honorarium of $10US. No attempt is made to prevent participants from interacting with each other normally before or after the experiment. Participants receive instruction in their game, are quizzed, practice the game at least once and discuss the outcome. They open their coded envelopes, remove their red and blue tickets, and play one game. In all of the treatments, the winner receives a private check in the mail.

4.3 Public Balloting

Two of the treatment conditions use a public ballot box, which is a small box carried to each participant. In these treatments, the participants stand in a circle, facing each other, and must stuff their red or blue ticket into the box in view of the group. In these treatments, the valid entries are determined at the time of the experiment, and all are sealed in identical envelopes. The envelopes are shuffled in a bag. A participant who ceded draws. The experimenter marks the envelope. The envelope is set aside, check to be mailed later.

4.3.1. The reputation condition

In this treatment, none of the partners are identified to each other before they make their choices, but they know that the box will be opened and the partners matched by checking their codes against a list, immediately after play. Participants are not allowed to converse or gesture once they have opened their envelopes, removed their tickets, and circled around to start play. They may, of course, exchange eye contact and facial expressions. After 20 seconds pause, the experimenter counts down from 10, and then brings the ballot box around.

4.3.2. The honest condition

After opening their envelopes and removing their tickets, players must hold one ticket in each hand. Partners are pointed out to each other. They may not speak or gesture to each other, except by raising, lowering, or waving either or both tickets. Again, they may exchange eye contact and facial expressions. After 20 seconds and then the countdown from 10, each player is committed to whichever color of ticket she or he holds aloft.

4.4. Secret Balloting

The other two treatments employ a two-compartment ballot box with two armholes in the front and a lidded, fastenable top. In these treatments players practice with both the lid up and the lid fastened so that they learn how to place their choice in the upper compartment and discard their other ticket in the bottom one. In these conditions, the drawing is done by a research assistant later. As in all treatments, the winner gets a check in the mail.

4.4.1. The no-signaling condition

This is the control treatment. After instruction and practice, participants face outward in a circle and after 30 seconds the box is brought around. There is no talking or eye contact, and no player is given any information about partners or the votes cast. As soon as the box has completed the round, the game is done.

4.4.2. The "cheap talk" condition

Partners are identified as in the "honest" condition. They have 30 seconds to confer. The experimenter counts down the last 10, then they must form an outward-facing circle. The box comes around and when they have all voted, again the game is done.

4.5 Replications

Colleagues donated classroom time for the first 2 replications. The participants in those 2 classroom groups (N = 22 and N = 12), each played the game under all

four conditions, but with a different, randomly assigned partner each time. The two groups played conditions in reverse order. As it was possible that these repetitive-play circumstances triggered a sense of reciprocity even where there was no material opportunity for it, the experiments were then replicated as true one-shot trials, outside the student-center cafeterias. The participants in these 7 smaller "cafeteria" groups (N = 56) played only one game apiece. These trials replicated each of the 4 signaling conditions twice, once with a group of 6 players and once with a group of 8.

5. Expectations by condition

Decades of social-dilemma game experiments, reviewed in Sally (1995) and Camerer (2003), show that we should rarely expect all participants to play the self-interested strategy, and especially not in one-shot games nor experiments in which participants meet each other. Nonetheless, the no-signaling condition of ChickenHawk should logically produce the most defections, as it affords no chance to exploit language, knowledge of who one's partner is nor any social preferences for cooperation or enhancing reputation.

Conversely, in the reputation condition it would reasonable to expect the most frequent altruism, as the condition affords an opportunity to display a sacrificial gesture but not to identify or coordinate with one's partner.

Reasoning from the theoretical distinction between "cheap talk" and "costly signaling" systems outlined above, the prior expectation in this study was that, given the homogeneity of the participant population and their prior acquaintance, they would be able to coordinate their pairwise choices with the greatest collective efficiency in the "honest" condition in which partners could see each others faces and signal to each other, at least semaphorically, their options on a simple, binary decision.

On the other hand, the temptations of anonymity, as well as the complex social emotions that arise in face-to-face conversation, justified an expectation of greater defection in the "cheap talk" condition, out of self-interest or a desire to deny someone else an entry he or she may have forcefully negotiated.

6. Results and discussion

As expected, under no conditions did all participants defect. And also as expected, they defected at by far the highest rate in the "no signaling" condition. Individual participants were, equally unsurprisingly, much more likely to cede their opportunity in the "reputation" condition, so much so that the collective efficiency of resource extraction in those treatments was lowered by the surplus of altruists matched to each other.

Two results ran sharply contrary to expectations. The first surprising result was that, even when participants had both the motive of reputation and the

possibility of arranging their choices by pre-play signaling through holding up their ballots and exchanging eye-contact and facial expressions with their partners, they were incapable of extracting resources at a rate significantly better than would be expected from a random voting pattern. Nor was this due to excess of self-interested demands for entries. As in the "reputation" condition, too many of these players ceded. They cooperated, but they failed to coordinate. (See Table 1.)

Table 1. Summary results, all treatments

Condition	Pairs' Success Rate	Pair-Type Skew	Participants' Altruism Rate
No signal	0.33	DD**	0.21** (-)
Reputation	0.46	CC*	0.69* (+)
Honest	0.42	CC*	0.69* (+)
Cheap Talk	0.92**	CD*	0.5

N = 24 pairs (48 subj.) for each condition: 90 unique participants, 96 unique pairings, 192 ballots.
* $p<0.05$, with random play as the null hypothesis; ** $p<0.01$
Significance measured by 95%CI & 99.7%CI of random binomial distribution for pairs' success & participants' altruism rates.
Significance measured by χ^2 with 1df for pair-type skew.
(+/-) indicates direction of deviation from expectation

The second surprising result was that any temptations to renege on deals made in a mere thirty seconds of pre-play conversation in the "cheap talk" condition did not derail the effectiveness of talking itself for coordinating strategies. Deals between partners were honored, judging by their 92% efficiency at extracting the one available entry per pair.

This latter result may simply derive from the fact that this is an exceptionally prosocial cultural group, but it seems difficult to imagine that the population sampled by this study, atypically homogeneous though it may be today, would be more interdependent than most ancestral human groups. Nor can the caveat of a peculiarly prosocial culture cover the curious *lack* of better coordination on the "honest" condition. If anything, those results would imply that too much eagerness to display one's commitment to self-sacrifice could interfere with a group's collective success. This is not implausible: extravagant individual costly-signaling displays, even of group loyalty, could be counterproductive for the group's collective efficiency.

So perhaps we do "cooperate because we speak." Or at least speech enabled us to leverage cooperation into much more efficient coordination. We need not abandon the argument that language, as discrete-combinatorial system, could not do so. Both the theoretical expectations and empirical findings of this study fit with Baker's (2004) suggestion that language differences function as necessary barriers to out-groups. One needs commitment to master any in-group's

language. Where all the linguistic markers of identity comport, as for the participants in this study, the in-group status marked by speech may itself be the costly signal that bolsters even the cheapest talk.

Acknowledgements

Thanks to my undergraduate research assistants, Jeffrey Barth and Carrie Robinson.

References

Baker, M. C. (2003). Linguistic difference and language design. *TRENDS in Cognitive Sciences, 7*, 349-353.

Brandt, H., Hauert, C., & Sigmund, K. (2003). Punishment and reputation in spatial public goods games. *Proceedings of the Royal Society of London B, 270*, 1099-1104.

Camerer, C. F. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. Princeton: Princeton University Press.

Kings-Casas, B., Tomlin, D., Anen, C, Camerer, C. F., Quartz, S. R., & Montague, P. R.. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science, 308*, 78-83.

Komarova, N. L., & Nowak, M. A. (2003) Language, learning, and evolution. In M. H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 317-337). Oxford: Oxford University Press.

Milinski, M., Semmann, D., & Krambeck, H.-J. (2002). Reputation helps solve the "tragedy of the commons." *Nature, 415*, 424-426.

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*, 573-577.

Pinker, S. (1994). *The language instinct*. NY: Harper Collins.

-----. (1998). Words and rules. *Lingua, 106,* 219-242.

Sally, D. (1995). Conversation and cooperation in social dilemmas. A meta-analysis of experiments from 1958-1992. *Rationality and Society, 7*, 58-92.

Zahavi, A., & Zahavi, A. (1997). The Handicap Principle. Oxford: Oxford University Press.

CONSTRAINING THE TIME WHEN LANGUAGE EVOLVED

SVERKER JOHANSSON
School of Education and Communication, University of Jönköping, Box 1026, Jönköping, SE-551 11, Sweden
lsj@hlk.hj.se

The precise timing of the emergence of language in human prehistory cannot be resolved. But the available evidence is sufficient to constrain it to some degree. This is a review and synthesis of the available evidence, leading to the conclusion that the time when speech became important for our ancestors can be constrained to be not less than 500,000 years ago, thus excluding several popular theories involving a late transition to speech.

1. Introduction

That modern humans have language, and that our remote ancestors did not, are two incontrovertible facts. But there is no consensus on *when* the transition from non-language to language took place, nor any consensus on whether it was a sudden jump or a gradual process. In this paper, I will explore to what extent the timing of the transition to language can be constrained by fossil, archaeological, genetic, and other evidence. A fuller discussion of this and related issues can be found in Johansson (2005).

2. Upper limits

Very little can be said about upper limits on the age of language. Our closest relatives today do not have language in any reasonable sense, at least not in the wild. A reasonable inference from the lack of native ape language is that the last common ancestor of us and the other apes also lacked language. The alternative, that language evolved earlier and was subsequently lost in the chimpanzee lineage, appears implausible — why would something as useful as language be lost by a species heavily engaged in social communication?

The last common ancestor of humans and chimps almost certainly lived less than 10 million years ago (mya), and conceivably as recent as 4 mya. The current best estimate from molecular data is around 5 mya (Pääbo, 2003).

The oldest known fossils that with some confidence can be assigned to the human line are those in the genus *Ardipithecus*, from 4 - 6 mya, found in Ethiopia (Haile-Selassie, Suwa, & White, 2004). Other possible contenders for the earliest

proto-human fossil are *Sahelanthropus tchadensis*, a skull found recently in Chad (Brunet et al., 2002), from about 6-7 mya with a puzzling mixture of features making it difficult to classify, and *Orrorin tugenensis* (Senut et al., 2001) just below 6 million years old, found in Kenya.

The discoverers of *Ardipithecus*, *Sahelanthropus* and *Orrorin* all consider their own fossil to be a human ancestor, and the others to be side branches (Cela-Conde & Ayala, 2003). I find the case for *Ardipithecus* somewhat more compelling, but the jury remains out. All these fossils have sufficient similarities with both humans and other apes that they are likely to be quite close to the branching point in the family tree. This would give an estimate from fossils of the most likely age of the last common ancestor somewhere in the vicinity of 6 mya, consistent within the uncertainties with the molecular estimate.

But beyond this common ancestor there is no way to place any more stringent upper limit on the time of language emergence. It may appear unlikely that e.g. australopithecines possessed language 2-3 mya, but there is no hard evidence that excludes it. It has been argued both that a minimum brain size is needed for language, and that the presence of language implies a human-like material culture. Both of these arguments are somewhat plausible, but it is by no means established that they exclude language in Lucy.

Thus, the best firm upper limit that can be placed on the time of language emergence remains a bit beyond 5 million years ago.

3. Lower limits

All modern human populations have language, obviously. Given that language has at least some biological substrate (if not necessarily an innate grammar) this implies that the most recent common ancestor of all modern humans had language.

The molecular data strongly support a common origin for all extant humans somewhere around 100,000 – 200,000 years ago (Ayala & Escalante, 1996; Wood, 1997; Bergström et al., 1998; Cavalli-Sforza & Feldman, 2003). The so-called 'Mitochondrial Eve' (Cann, Stoneking, & Wilson, 1987; Saville, Kohli, & Anderson, 1998; Cavalli-Sforza, 1998), the putative common ancestress of all women, was at the forefront of this molecular wave, but she has since been joined by a corresponding 'Y-chromosome Adam' (Fu et al., 1996; Dorit, Akashi, & Gilbert, 1995; Pääbo, 1995), as well as by data from non-sex-linked genes (Fischman, 1996), and from X-chromosomes (Disotell, 1999). It follows that the origins of language cannot possibly be more recent than 100,000 years ago. This conclusion is consistent with archaeological evidence of the spread of *Homo sapiens* out of Africa and the peopling of various continents, notably Australia more than 50,000 years ago.

This 100,000-year lower limit already excludes theories that connect the origins of language with the apparent 'cultural revolution' in the archaeological record around 40,000 years ago, discussed in the next section.

For at least some aspects of language, stronger time limits are possible. Our habitual use of speech is reflected in certain aspects of our anatomy, that can be studied in fossils. Speech adaptations can potentially be found in our speech organs, hearing organs, brain, and in the neural connections between these organs.

• *Speech organs.* The shape of the human vocal tract, notably the lowered larynx, is a clear speech adaptation. The vocal tract itself is all soft tissue and does not fossilize, but its shape is connected with the shape of the surrounding bones, the skull base and the hyoid. Already *Homo erectus* had a near-modern skull base (Baba et al., 2003), but the significance of this is unclear (Fitch, 2000; Spoor, 2000). Hyoid bones are very rare as fossils, as they are not attached to the rest of the skeleton, but one Neanderthal hyoid has been found (Arensburg et al., 1989), very similar to the hyoid of modern *Homo sapiens*, leading to the conclusion that Neanderthals had a vocal tract similar to ours (Houghton, 1993; Boë, Maeda, & Heim, 1999).

• *Hearing organs.* Some fine-tuning appears to have taken place to optimize speech perception, notably our improved perception of sounds in the 2-4 kHz range. The sensitivity of ape ears has a minimum in this range, but human ears do not, mainly due to minor changes in the ear ossicles, the tiny bones that conduct sound from the eardrum to the inner ear. This difference is very likely an adaptation to speech perception, as key features of some speech sounds are in this region. The adaptation interpretation is strengthened by the discovery that a middle-ear structural gene has been the subject of strong natural selection in the human lineage (Olson & Varki, 2004). According to Martínez et al. (2004), these changes in the ossicles were present already in the 400,000-year-old fossils from Sima de los Huesos in Spain, well before the advent of modern *Homo sapiens*. These fossils are most likely Neanderthal ancestors, that Martínez et al. (2004) attribute to *Homo heidelbergensis*.

• *Brain.* Only the gross anatomy of the brain surface is visible as imprints on the inside of well-preserved fossil skulls. In principle, the emergence of e.g. Broca's area could be pinpointed this way. But other apes have brain structures with the same gross anatomy as both Broca and Wernicke (Gannon et al., 1998; Cantalupo & Hopkins, 2001), so the imprints of such areas in the skulls of protohumans tell us nothing useful about language. Nor is there any clearcut increase in general lateralization — chimp brains are not symmetric.

• *Neural connections.* Where nerves pass through bone, a hole is left that can be seen in well-preserved fossils. Such nerve canals provide a rough estimate of the size of the nerve that passed through them. A thicker nerve means more neurons, and presumably improved sensitivity and control. The hypoglossal canal, leading to the tongue, is sometimes invoked in this context, but the fossil evidence is contradictory (Kay, Cartmill, & Balow, 1998; DeGusta et al., 1999). A better case can be made for the nerves to the thorax, presumably for breathing control. Both modern humans and Neanderthals have wide canals here, whereas *Homo er-*

gaster have the narrow canals typical of other apes (MacLarnon & Hewitt, 1999).

In conclusion, the fossil evidence indicates that at least some apparent speech adaptations were present in Neanderthals. None of these anatomical details is compelling on its own, but their consilience strengthens the case for Neanderthal speech in some form.

The presence of speech in Neanderthals sets a lower limit for the age of speech at the time of the last common ancestor of us and the Neanderthals (unless one postulates, implausibly, the independent evolution of the same set of adaptations in both lineages). It has long been a controversial issue whether the Neanderthals actually were a separate lineage, or just a subspecies of *Homo sapiens*, but genetic evidence from Neanderthal fossils clearly demonstrates their separateness, and indicates that the last common ancestor lived at least half a million years ago (e.g. Krings et al., 1999; Höss, 2000; Beerli & Edwards, 2002; Knight, 2003; Caramelli et. al., 2003)[a]. The fossil evidence points in the same direction, with the earliest modern humans in Europe more resembling Africans than Neanderthals (Tyrrell & Chamberlain, 1998).

There is no consensus on the taxonomy of the transitional fossils from around the time of our common ancestor with the Neanderthals. The names *Homo heidelbergensis, rhodesiensis, antecessor, helmei* and others are all in current use. It is, however, quite well established that all of these have their roots in *Homo erectus (sensu lato)*, so I will use the name *erectus* for our last common ancestor.

4. The revolution that wasn't

The archeological record has frequently been invoked as support for the late, sudden appearance of language, due to the perception of a technological and creative revolution around 40,000 years ago (e.g. Binford, 1989; Li & Hombert, 2002).

Language use in itself is not archeologically visible, but other forms of sign use may be, and may be used as indicators that some level of semiotic abilities has been reached. Invoking ancient art, including pigments and personal ornaments, as indicators that the artists were capable of symbolic thought, or even as an indicator that language had evolved, is fairly common (Mellars, 1998).

The supposedly sudden appearance of advanced art and advanced tools in the caves of Europe about 40,000 years ago is taken as evidence of a cognitive leap. However, the appearance of a sudden dramatic 'cultural revolution' around 40,000 years ago, has turned out to be largely an illusion caused by the predominance of European sites in the documented archeological record, and possibly some Eurocentrism among archeologists (Henshilwood & Marean, 2003). *Homo sapiens*

[a]Some studies of genetic statistics in modern humans support some degree of admixture, e.g. Eswaran, Harpending, and Rogers (2005), but this must carry less weight than the direct evidence from fossil Neanderthal mtDNA, which is consistently distinct from either recent or fossil *Homo sapiens* mtDNA.

did indeed invade Europe rather suddenly about 40,000 years ago, bringing along an advanced toolkit — but that toolkit had been developed gradually in Africa over the course of more than 200,000 years (McBrearty & Brooks, 2000; Van Peer et al., 2003). Recent discoveries of works of abstract art (Henshilwood et al., 2002), pigment use (Barham, 2002), and personal ornaments (Henshilwood et al., 2004), all substantially older than 40,000 years, add further support to the long timescale of McBrearty and Brooks (2000). The debate over the supposed revolution is reviewed by Bar-Yosef (2002) or Henshilwood and Marean (2003).

Additional evidence has been uncovered recently that appears to show that simple art may actually have predated the appearance of anatomically modern *Homo sapiens* (Bahn & Vertut, 1997; Keys, 2000; Bednarik, 2003), in the context of *Homo heidelbergensis* or possibly even *Homo erectus*. Objects that can reasonably be interpreted as art have been found associated also with Neanderthals (Appenzeller, 1998; d'Errico et al., 2003; Wynn & Coolidge, 2004). While these finds are much simpler than the figurative art of later *Homo sapiens* (e.g. Bahn & Vertut, 1997; Conard, 2003), they nevertheless push back the origin of the biological capacities needed for simple art at least to the common ancestor of Neanderthals and us, some 500,000 years ago. And given that the symbolic capacities needed for art are also needed for language, and are interpreted by some as indicative of the presence of language, this adds support to the time limit inferred from anatomy in the previous section.

5. Conclusions

Fossil evidence indicates that speech optimization of our vocal apparatus got started well before the emergence of *Homo sapiens*, almost certainly more than half a million years ago, probably in *Homo erectus*. As the speech optimization, with its accompanying costs, would not occur without strong selective pressure for complex vocalizations, presumably verbal communication, this implies that *Homo erectus* already possessed non-trivial language abilities.

There is no real evidence indicating just how complex language *erectus* had. It must have been complex enough to require fine-grained vocal distinctions, but this need not imply anything like modern grammar. They may have been at a holophrastic stage, or they may have had nearly full human language — it is difficult to imagine any way to tell. On one hand, *erectus* is the first hominid with a brain size approaching the modern human range — there are modern humans alive today with *erectus*-sized brains and excellent language skills — and they were also the first to spread out to many different habitats on different continents. But on the other hand their comparatively simple, static culture argues against their having modern human cognitive skills. In particular, it is quite clear that they lacked the cumulative cultural evolution that is so characteristic of modern humans. Given that they are different from modern humans in such fundamental ways, their having full modern human language appears unlikely.

Language need not have started in a spoken modality; sign language may have been the original language (e.g. Corballis, 2002), likely building on mimesis (Donald, 1997; Zlatev, Persson, & Gärdenfors, 2005). This means language may be older than speech — but hardly younger. A lower age limit on speech remains a firm lower limit on the age of language at the *erectus* level, if not necessarily on full modern grammar.

Modern humans, after parting company with the Neanderthals perhaps half a million years ago, would have acquired the remaining features of modern language in parallel with acquiring modern human anatomy. Both aspects were finished before modern humans started spreading over the world, at least 100,000 years ago. The last common ancestor of all humans today, probably living in Africa not so long before this exodus, is the likely speaker of Proto-World, the common ancestor of all the modern language families, and the earliest language which we may have any remote hope of ever reconstructing. But there is no reason to believe that this Proto-World was the *first* language spoken — as discussed above, our ancestors may have had language for a million years already. The details of those earlier proto-languages are likely to remain opaque for the foreseeable future.

References

Appenzeller, T. (1998). Art: evolution or revolution? *Science 282:1451-1454.*

Arensburg et al. (1989). A middle paleolothic human hyoid bone. *Nature 338:758-760.*

Ayala, F. J., & Escalante, A. A. (1996). The evolution of human populations: a molecular perspective. *Mol Phyl & Evo 5:188-201.*

Baba et al. (2003). Homo erectus calvarium from the pleistocene of java. *Science 299:1384-1388.*

Bahn, P. G., & Vertut, J. (1997). *Journey through the ice age, 2nd ed.* London: Weidenfeld & Nicolson.

Barham, L. S. (2002). Systematic pigment use in the Middle Pleistocene of South-Central Africa. *Current Anthropology 43:181-190.*

Bar-Yosef, O. (2002). The upper paleolithic revolution. *Ann Rev Antropology 31:363-393.*

Bednarik, R. G. (2003). A figurine from the African Acheulian. *Current Anthropology 44:405-438.*

Beerli, P., & Edwards, S. V. (2002). When did neanderthals and modern humans diverge? *Evolutionary Anthropology, Suppl 1:60-63.*

Bergström et al. (1998). Recent origin of HLA-DRB1 alleles and implications for human evolution. *Nature Genetics 18:237-242.*

Binford, L. R. (1989). Isolating the transition to cultural adaptations: An organizational approach. In E. Trinkaus (Ed.), *The emergence of modern humans: Biocultural adaptations in the Later Pleistocene.* Cambridge: Cambridge University Press.

Boë, Maeda, & Heim. (1999). Neandertal man was not morphologically handicapped for speech. *Evolution of Communication 3:49-77.*

Brunet et al. (2002). A new hominid from the Upper Miocene of Chad, Central Africa. *Nature 418:145-151.*

Cann, Stoneking, & Wilson. (1987). Mitochondrial DNA and human evolution. *Nature 325:31-36.*

Cantalupo, C., & Hopkins, W. D. (2001). Asymmetric Broca's area in great apes. *Nature 414:505.*

Caramelli et. al. (2003). Evidence for a genetic discontinuity between neandertals and 24,000-year-old anatomically modern europeans. *Proc Nat Acad Sci 100:6593-6597.*

Cavalli-Sforza, L. L. (1998). The DNA revolution in population genetics. *Trends Gen 14:60-65.*

Cavalli-Sforza, L. L., & Feldman, M. W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nature Genetics Supplement 33:266-275.*

Cela-Conde, C. J., & Ayala, F. J. (2003). Genera of the human lineage. *Proc Nat Acad Sci 100:7684-7689.*

Conard, N. J. (2003). Paleolithic ivory sculptures from southwestern Germany and the origins of figurative art. *Nature 426:830-832.*

Corballis, M. C. (2002). *From hand to mouth: the origins of language.* Princeton: Princeton University Press.

DeGusta et al. (1999). Hypoglossal canal and hominid speech. *Proc Nat Acad Sci 96:1800-1804.*

d'Errico et al. (2003). Archaeological evidence for the emergence of language, symbolism, and music — an alternative multidisciplinary perspective. *J World Prehistory 17:1-70.*

Disotell, T. R. (1999). Origins of modern humans still look recent. *Curr Bio 9:R647-R650.*

Donald, M. (1997). Precis of origins of the modern mind: Three stages in the evolution of culture and cognition. *Behav & Brain Sci 16:737-791.*

Dorit, Akashi, & Gilbert. (1995). Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science 268:1183-1185.*

Eswaran, V., Harpending, H., & Rogers, A. R. (2005). Genomics refutes an exclusively african origin of humans. *J Hum Evo 49:1-18.*

Fischman, J. (1996). Evidence mounts for our African origin — and alternatives. *Science 271:1364.*

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends Cogn Sci 4:258-267.*

Fu et al. (1996). Estimating the age of the common ancestor of men from the ZFY intron. *Science 272:1357-1365.*

Gannon et al. (1998). Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. *Science 279:220-222.*

Haile-Selassie, Suwa, & White. (2004). Late Miocene teeth from Middle Awash, Ethiopia, and early hominid dental evolution. *Science 303:1503-1505.*

Henshilwood, C. S., & Marean, C. W. (2003). The origin of modern human behavior. *Current Anthropology 44:627-651.*

Henshilwood et al. (2002). Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science 295:1278-1280.*

Henshilwood et al. (2004). Middle Stone Age beads from South Africa. *Science 304:404.*

Höss, M. (2000). Neanderthal population genetics. *Nature 404:453-454.*

Houghton, P. (1993). Neandertal supralaryngeal vocal tract. *Am J Phys Anthro 90:139-146.*

Johansson, S. (2005). *Origins of language — constraints on hypotheses.* Amsterdam: Benjamins.

Kay, Cartmill, & Balow. (1998). The hypoglossal canal and the origin of human vocal behavior. *Proc Nat Acad Sci 95:5417-5419.*

Keys, D. (2000). Earliest paint shows that man wasn't the first artist. *http://www.independent.co.uk/News/UK/Science/2000-05/paint010500.shtm.*

Knight, A. (2003). The phylogenetic relationship of neandertal and modern human mitochondrial dnas based on informative nucleotide sites. *J Hum Evo 44:627-632.*

Krings et al. (1999). Dna sequence of the mitochondrial hypervariable region ii from the neanderthal type specimen. *Proc Nat Acad Sci 96:5581-5585.*

Li, C. N., & Hombert, J.-M. (2002). On the evolutionary origin of language. In M. I. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language.* Amsterdam: Benjamins.

MacLarnon, A. M., & Hewitt, G. P. (1999). The evolution of human speech: the role of enhanced breathing control. *Am J Phys Anthro 109:341-363.*

Martínez et al. (2004). Auditory capacities in middle pleistocene humans from the sierra de atapuerca in spain. *Proc Nat Acad Sci 101:9976-9981.*

McBrearty, S., & Brooks, A. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J Hum Evo 39:453-563.*

Mellars, P. A. (1998). Neanderthals, modern humans and the archaeological evidence for language. In N. G. Jablonski & L. C. Aiello (Eds.), *The origin and diversification of language.* San Francisco: California Academy of Sciences.

Olson, M. V., & Varki, A. (2004). The chimpanzee genome — a bittersweet celebration. *Science 305:191-192.*

Pääbo, S. (1995). The Y chromosome and the origin of all of us (men). *Science 268:1141.*

Pääbo, S. (2003). The mosaic that is our genome. *Nature 421:409-412.*

Saville, Kohli, & Anderson. (1998). mtDNA recombination in a natural population. *Proc Nat Acad Sci 95:1331-1335.*

Senut et al. (2001). First hominid from the Miocene (Lukeino formation, Kenya). *C R Acad Sci Paris, Sciences de la Terre et des Planetes 332:137-144.*

Spoor, F. (2000). Balance and brains: evolution of the human cranial base. *Anthroquest News http://www.leakeyfoundation.org/n9_spr2000_3.html.*

Tyrrell, A. J., & Chamberlain, A. T. (1998). Non-metric trait evidence for modern human affinities and the distinctiveness of neanderthals. *J Hum Evo 34:549-554.*

Van Peer et al. (2003). The Early to Middle Stone Age transition and the emergence of modern human behaviour at site 8-B-11, Sai Island, Sudan. *J Hum Evo 45:187-193.*

Wood, B. (1997). Ecce Homo - behold mankind. *Nature 390:120-121.*

Wynn, T., & Coolidge, F. L. (2004). The expert Neandertal mind. *J Hum Evo 46:467-487.*

Zlatev, J., Persson, T., & Gärdenfors, P. (2005). Bodily mimesis as "the missing link" in human cognitive evolution. *Lund University Cognitive Studies 121.*

WORKING BACKWARDS FROM MODERN LANGUAGE TO PROTO-GRAMMAR

SVERKER JOHANSSON

School of Education and Communication, University of Jönköping, Box 1026, Jönköping, SE-551 11, Sweden
lsj@hlk.hj.se

The possibilities for a stepwise evolution of grammar are evaluated through an analysis of which components of modern human grammar are removable, and in what order, while still leaving a functional communication system. It is found that recursivity is a prime candidate for being a late evolutionary addition, with flexibility and hierarchical rules coming next. Furthermore, it is argued that recursivity need not be the unitary infinite-loop concept of formal grammars, but can evolve in several smaller steps.

1. Introduction

Communication is certainly possible without grammar, as shown by the communicative abilities of both agrammatic patients and children at the one-word stage of language acquisition, and our remote ancestors certainly lacked grammar, as they lacked language in any form. This means that grammar, and syntax in particular, must have emerged during the course of human evolution.

It is sometimes argued that modern human grammar is a monolithic system that cannot be built piece by piece (e.g. Chomsky, 1972). But I will argue that such a stepwise construction of grammar (or deconstruction, from the perspective of modern grammar) is perfectly possible if the structure of grammar is looked at from an appropriate perspective, and the pieces are added in the right order — not all aspects of grammar are totally interdependent (cf. Jackendoff, 1999; Pinker & Jackendoff, 2005). Certainly, if any component of modern human language is removed, what is left is not equal to modern human language — but it may still be a functional language, if not as rich and expressive as what we're using today. A language with proto-syntax, missing one or two principles of modern grammar, may not be adequate to write this paper — but may nevertheless be adequate for the daily life of proto-humans.

2. Syntax precursors?

Useful exaptations or precursors for syntax among the capacities likely to be present among our pre-linguistic ancestors are difficult to identify, and even more

difficult to verify, though more or less speculative ideas abound.

The cognitive capacity needed to handle relational concepts would appear necessary. Several authors seek to base this relational capacity in social interactions or 'social scripts', e.g. Worden (1996), Tomasello (1999), Aiello (1998), or Calvin and Bickerton (2000). Others invoke tool-making in a similar role (Greenfield, 1991; Ambrose, 2001; Wildgen, 2004).

Armstrong et al. (1995) invoke the temporal structure of sign sequences as the roots of syntax, similar to the model for syntax origins proposed by Condillac (1746), an intriguing possibility as an iconic sign sequence describing mimetically an action indeed naturally possesses a rudimentary structure that might be a reasonable syntax precursor.

Nevertheless, none of these or other proposed syntax precursors takes us very far along the road towards modern human grammar, and there is little direct evidence supporting any of them.

3. Working backwards from modern grammar

What happens if we look at the problem from the other end, not at possible syntax precursors, but at modern human grammar, and contemplate which components of modern grammar might be removable? Removability of a component from a modern grammar reasonably entails its addability to a proto-grammar.

A definitive analysis of the removability of different components or principles of grammar requires that we know what these principles are. However, as noted by Edelman and Pedersen (2004):

> ...we have, as yet, no comprehensive, psychologically real and neurobiologically grounded process model for language, and with a descriptive model there is a distinct possibility that the features we believe to be important are in fact immaterial (p. 399).

The Chomskian paradigm, e.g. (1995), is one such descriptive model, popular with many linguists. But there are several competing theories of grammar that remain viable, both other generative grammars (e.g. Bresnan, 1982; Pollard & Sag, 1994) and e.g. functional-cognitive grammars (Van Valin & LaPolla, 1997; Givón, 1997; Halliday, 2004). Instead of assuming a specific theory of grammar, a slightly different perspective will be adopted here, focusing on a few features that modern human languages incontrovertibly possess, regardless of the details of grammatical theory, and seeing which of those features may be removable, and in which order.

Human languages are universally:

1. *Structured* in the sense that an utterance is not just a random juxtaposition of words, but in some way indicates the relations between words. The structure indicators may be linear order, or morphological markers, or whatever.

2. *Hierarchical* in the sense that there are levels of structures within structures.

3a. *Flexible* in the transformational sense that there are many different ways to express the 'same' meaning by moving around words and restructuring sentences according to certain rules.

3b. *Recursive* in the sense that the same rules and structures may recur at different levels in the hierarchy, so that a structure may contain a substructure that is another instantiation of the same structure, in theory repeated *ad infinitum.*

The features are listed with the most fundamental first, and the most easily removable at the end, as discussed below. *Flexible* and *Recursive* are independently removable, and are thus at the same level.

None of the other features of language requires the *Recursive* feature, whereas *Recursive* certainly requires both *Hierarchical* and *Structured.* It is quite possible to have a language with only partial recursivity, or even none at all, supported by e.g. the fact that some children with SLI (Specific Language Impairment) apparently lack recursivity (Bloom, 1999). Therefore, *Recursive* is a prime candidate for being a late evolutionary addition to human grammar, a possibility further discussed in the next section. Hauser & Chomsky & Fitch (2002) similarly places recursion as the final step in the emergence of language, in that recursion is the sole component of their FLN, i.e., the only component of the language faculty that is narrowly language-specific, with everything else being used also for non-linguistic purposes. Hurford (2003) likewise proposes that the earliest languages lacked subordinate clauses, and thus presumably lacked (at least non-trivial) recursion.

Flexible, like *Recursive*, appears to be an optional feature that can be removed without fatal effects. *Flexible* definitely requires *Structured* to be meaningful, and may require *Hierarchical*, but none of the other features require *Flexible*. Depending on the exact grammatical theory, there may be a lot of obligatory moving around of constituents in a sentence — but in modern human grammar there are also lots of optional movement possibilities, constituents that can be appear in different places at the discretion of the speaker (topicalization is one example; see e.g. Box 4 of Jackendoff (1999) for others). Language would be perfectly functional, if less rich and nuanced, without these discretionary movements. Thus, *Flexible* is also a candidate for being a late evolutionary addition to human grammar.

A *Hierarchical* language must be *Structured*, but need not be either *Recursive* or *Flexible*. The main breakthrough in achieving *Hierarchical* may be the grouping of words into headed units, and the application of structural rules to headed units as a whole, rather than to individual words (Jackendoff, 1999). It is possible to have a hierarchy of structures, without the same structure being allowed to recur in infinite recursive chains.

The *Structured* feature, finally, is fundamental — it makes little sense to talk about syntax at all for a non-structured language, and all the other features presup-

pose *Structured*. But it is perfectly possible to imagine a language that is *Structured* without possessing any of the other features — the two-word stage of child language may be an example, and some pidgin languages may be clearer examples. This means that *Structured* must be the first syntax feature to emerge. Given that alternative ways of indicating structure are typically the product of grammaticalization processes, requiring a pre-existing syntax, it may be argued that linear word order is most plausible as the original implementation of *Structured* (cf. Hurford, 2003).

This adds up to an allowed sequence of successive grammar elaboration, that may be a candidate evolutionary sequence:

1. One-word stage — basic semantics with no syntax
2. Two-word stage — *Structured*, but with none of the other features.
3. Hierarchical structure, much like a basic phrase structure grammar, but with no recursivity. This means a language without subordinate clauses and other forms of embedding.
4. Recursive syntax (alternatively, *Flexible* may be added before *Recursive*, since they are largely independent of each other).
5. Full modern human grammar.

Each step in this sequence corresponds to a functional communication system, if not as elaborate and rich as the modern human system. And none of the gaps that need to be bridged when going from one step to the next looks anything like the huge chasm commonly pictured between non-syntax and syntax. The steps roughly resemble the stages of child language acquisition, where both recursivity (Goldin-Meadow, 1982) and flexibility (Håkansson, 1994) are fairly late additions.

Jackendoff (1999) presents a similar sequence, with a similar number of syntactic steps (he has more steps in total, but many of them do not concern syntax, and are not covered here[a]). Jackendoff's syntactic steps are (i) Concatenation of symbols, (ii) Symbol position significant [i.e. *Structured*], (iii) Hierarchical structure, and (iv) Phrasal syntax conveying semantic relations. The main differences appear to be that Jackendoff concatenates symbols into longer strings than two words before adding *Structured*, and that he does not make a clear distinction between *Hierarchical* and *Recursive*.

4. More on recursion

Depending on exactly what the underlying grammar looks like, it is not inconceivable that some of these steps, notably *Recursive*, can be subdivided even further,

[a]See Johansson (2005) for a full discussion of other aspects of language origins.

with e.g., simple additive 'tail recursion' being added before central-embedding 'true recursion'.[b] Expressed in phrase structure rules:

$$S \rightarrow NP\,VP \quad (1)$$

$$NP \rightarrow Adj\,NP \quad (2)$$

$$NP \rightarrow NP\,Comp\,S \quad (3)$$

Both rule 2 and rule 3 are recursive. But rule 2, which adds very little complexity in either production or comprehension, may well be an earlier development than rule 3, which is much more difficult to handle even for adult modern humans.

Recursivity is plausible as a late addition in phylogeny as well as ontogeny, also because adults have trouble with it — comprehension is poor on sentences with multi-level recursion (Christiansen & Chater, 1999), such as the following sentence built by triple application of rule 3 above: *The rat the cat the dog the man hit chased caught squeaked.* In theory, such sentences are grammatical — but they are commonly judged as ungrammatical (Christiansen & Chater, 1999), are difficult and time-consuming to parse (Bates, Devescovi, & D'Amico, 1999), and are exceedingly rare in natural language. The central-embedding recursion discussed here appears to be the worst case for our language processing, which breaks down with no more than three or four nested levels (Marcus, 2004), but with a sufficient number of levels of embedding most people find other types of recursion non-trivial to parse as well, as soon as the recursion amounts to more than simple concatenation.

Recursion was invoked by Chomsky (1957) as a language feature that was impossible to learn without an innate grammar, which may be true for infinite recursion. However, as noted above, human language does not in practice allow recursion more than a few levels deep. And recursion to the same depth that humans can handle, has been 'learned' by a connectionist network (Christiansen & Chater, 1999).

In real language production, we often do not build recursive sentences top-down, as they are typically presented in formal grammars. Instead, recursion is more often a matter of embedding a central clause in more and more layers of predication, commonly as a result of social interactions (Harder, 2004):

1. *Mary to Jack:* George is impossible!
2. *Jack to Joe:* Mary thinks that George is impossible.
3. *Joe to Jack:* Are you sure that Mary thinks that George is impossible?

[b]The computational requirements for 'tail' and 'true' recursion are quite distinct (Pinker & Jackendoff, 2005).

4. etc...

In this way, recursion can emerge from our ability to handle predicate logic in social interactions. The 'viewpoint chain' of Langacker (2001) provides a very similar grounding of recursion, as may others of his 'paths of mental access', particularly causal chains. There are also other ways for recursion to be an emergent property of language without a language-specific biological basis of its own, e.g. through the semiotic constraints of Deacon (2003).

5. Discussion

The first step towards syntax, getting started on the sequence at all, might be thought the most difficult — but since we have compelling evidence that stage 1 and possibly stage 2 are within reach of enculturated non-human apes (e.g. Savage-Rumbaugh et al., 1993), that step involves nothing but activating already existing capabilities, which cannot be an insurmountable problem. It can also be noted in this context that Fitch and Hauser (2004) managed to train monkeys to master a simple structured grammar in a toy language, but failed with a recursive grammar.

The postulated social scripts of Worden (1996) and Aiello (1998) have the features *Structured* and *Hierarchical*, and even rudimentary recursivity. This is a consequence of the structure and complexity of observed social behavior, making it plausible that apes had a structured and hierarchical conceptual system available as a language exaptation. If the cognitive machinery of the social scripts was available and could be used for language, we would reach stage 3 in the evolutionary sequence above. Byrne (2000) postulates more general cognitive structures for dealing with structured, hierarchical problems, which would be eminently exaptable to syntax processing, whereas Bickerton (2000) invokes hierarchical semantic structures as an exaptation for syntax.

Pidgin languages, or 'The Basic Variety' (BV) of Jackendoff (1999), with their highly simplified grammatical structure, can also be invoked here, at the very least as examples of functional languages without all the features of full modern human grammar (McMahon, 1994), and possibly as a modern-day example of what an intermediate stage in the evolution of syntax might have looked like (Bickerton, 1995). Pidgins and BV commonly lack *Recursive*, and may lack *Hierarchical*, consisting of just a linear structure of words.

For the last stages of the origin of modern grammar, Hurford (2003) invokes the observed unidirectionality of grammaticalization processes. Since grammaticalization is a process of delexification, going from lexical stems to function words and inflections, it appears plausible that the earliest languages lacked function words and inflections. This would be a functional, if pidgin-like, language, that could then evolve into our present languages through normal processes of diachronic language change, without any further need for biological evolution.

This also means that morphology need not be an issue in discussions of language origins, since it is largely the product of grammaticalization. This "morphologization" may have occurred either before or after the emergence of recursivity.

In conclusion, the gradual evolution of modern human grammar through several functional intermediate stages, appears perfectly possible. The exact path of evolution is speculative, due to the dearth of data on the structure of actual proto-languages, but no insurmountable obstacles are visible. Some of the required transitions can actually be observed today, either in ontogeny, or in e.g., transitions from pidgin to creole language.

References

Aiello, L. C. (1998). The foundations of human language. In N. G. Jablonski & L. C. Aiello (Eds.), *The origin and diversification of language.* San Francisco: California Academy of Sciences.

Ambrose, S. H. (2001). Paleolithic technology and human evolution. *Science 291:1748-1753.*

Armstrong et al. (1995). *Gesture and the nature of language.* Cambridge: Cambridge University Press.

Bates, Devescovi, & D'Amico. (1999). Processing complex sentences: a cross-linguistic study. *Lang & Cogn Proc 14:69-123.*

Bickerton, D. (1995). *Language and human behavior.* Seattle: UWP.

Bickerton, D. (2000). How protolanguage became language. In Knight, Studdert-Kennedy, & Hurford (Eds.), *The evolutionary emergence of language.* Cambridge: Cambridge University Press.

Bloom, P. (1999). Language capacities: Is grammar special? *Curr Bio 9:R127-R128.*

Bresnan, J. (1982). *The mental representation of grammatical relations.* Cambridge: MIT Press.

Byrne, R. W. (2000). Evolution of primate cognition. *Cogn Sci 24:543-570.*

Calvin, W. H., & Bickerton, D. (2000). *Reconciling Darwin and Chomsky with the human brain.* Cambridge: MIT Press.

Chomsky, N. (1957). *Syntactic structures.* Cambridge: MIT Press.

Chomsky, N. (1972). *Language and mind* (2nd ed.). New York: Harcourt, Brace & World.

Chomsky, N. (1995). *The minimalist program.* Cambridge: MIT Press.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cogn Sci 23:157-205.*

Condillac, E. B. (1746). *Essai sur l'origine des conaissances humaines.* Paris.

Deacon, T. (2003). Universal grammar and semiotic constraints. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution.* Oxford: Oxford University Press.

Edelman, S., & Pedersen, B. (2004). Review of Briscoe 'Linguistic evolution through language acquisition: formal and computational models'. *Journal of Linguistics 40:396-400.*

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science 303:377-380.*

Givón, T. (1997). *Grammatical relations: A functionalist perspective.* Amsterdam: Benjamins.

Goldin-Meadow, S. (1982). The resilience of recursion: a study of a communication system developed without a conventional language model. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: the state of the art* (p. 51-77). Cambridge: Cambridge University Press.

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behav & Brain Sci 14:531-595.*

Håkansson, G. (1994). Verb-initial sentences in the development of Swedish. *Working Papers 42:49-65, Dept of Linguistics, University of Lund, Sweden.*

Halliday, M. A. K. (2004). *Introduction to functional grammar* (2nd ed.). Oxford: Oxford University Press.

Harder, P. (2004). *Recursion as a culturally grounded phenomenon.* presented at "Language, Culture, and Mind", 18-20 July 2004, Portsmouth, http://www.port.-ac.uk/departments/academics/psychology/lcmconference2004/.

Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve. *Science 298:1565-1566.*

Hurford, J. R. (2003). The language mosaic and its evolution. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution.* Oxford: Oxford University Press.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends Cogn Sci 3:272-279.*

Johansson, S. (2005). *Origins of language — constraints on hypotheses.* Amsterdam: Benjamins.

Langacker, R. W. (2001). Dynamicity in grammar. *Axiomathes 12:7-33.*

Marcus, G. F. (2004). Before the word. *Nature 431:745.*

McMahon, A. M. S. (1994). *Understanding language change.* Cambridge: Cambridge University Press.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition 95:201-236.*

Pollard, C., & Sag, I. (1994). *Head driven phrase structure grammar.* Stanford: CSLI Publications.

Savage-Rumbaugh et al. (1993). Language comprehension in ape and child. *Monographs Soc Res Child Dev, Serial No. 233, vol 58, nos 3-4 (http://www.gsu.edu/~wwwlrc/monograph.html).*

Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge: Harvard University Press.

Van Valin, R., & LaPolla, R. (1997). *Syntax: Structure, meaning, and function.* Cambridge: Cambridge University Press.

Wildgen, W. (2004). *The evolution of human language: Scenarios, principles, and cultural dynamics.* Amsterdam: Benjamins.

Worden, R. P. (1996). Primate social intelligence. *Cogn Sci 20:579-616.*

LANGUAGE CO-EVOLVED WITH THE RULE OF LAW

CHRIS KNIGHT

School of Social Sciences, Media and Cultural Studies, University of East London, Docklands Campus, London E16 2RD, England

Many scholars assume a connection between the evolution of language and that of distinctively human group-level morality. Unfortunately, such thinkers frequently downplay a central implication of modern Darwinian theory, which precludes the possibility of innate psychological mechanisms evolving to benefit the group at the expense of the individual. Group level moral regulation is indeed central to sexual, social and political life in all known hunter-gatherer communities. The production of speech acts would be impossible without such regulation. The challenge, therefore, is to explain on a Darwinian basis how life could have become subject to the rule of law. Only then will we have an appropriate social framework in which to contextualize our models of how language may have evolved.

1. Language and the rule of law

Let me begin with a self-evident point, perhaps too often taken for granted. When academics participate in conferences and debates, we find ourselves operating under the rule of law. Protocols exist. We must keep to agreed time limits, disclose our sources, accept criticism and renounce any temptation to use threats, material inducements or force. There is status competition, certainly. But status is determined on an intellectual basis by peer evaluation alone; we compete to demonstrate relevance (Sperber and Wilson 1995 [1986]; Dessalles 1998) in one anothers' eyes.

What applies in academic life applies wherever language is used. Protocols exist. Compared with academic discourse, informal gossip may be livelier, more relaxed, less abstract and more intimately bound up with non-linguistic modes of expression. But despite such obvious differences, the same basic principles apply. Civilized discourse (Grice 1989: 22-40; Leech 1980; Brown and Levinson 1987 [1978]) is inseparably bound up with such things as tact, mutual face-saving and respect. What I am here terming 'the rule of law' is never just behavioral dominance exerted by certain individuals over others. It is contractual, valid only when based on mutual consent. The relevant contracts may be formal, informal or completely taken for granted (and correspondingly invisible). But only once such understandings are in place can any of us 'do things with words' (Austin 1978 [1955]).

This being the case, speech acts are never behavioral. On the level which matters – that is, on the level which is distinctively human – we proceed as if

playing an abstract game. As in chess, moves are made within a hallucinatory world, each intervention digitally specified and effective insofar as its intention is socially recognized (Grice 1989: 86-116). To emit a signal – even to transform the entire state of play – no physical investment need be made. In a language game (Wittgenstein 1968), interventions need not be materially costly. In principle, a nod or wink might suffice. What constitutes a move is the fact that it is agreed to be one – nothing more. Abstract digital information is then transmitted – with intuitive mind reading filling in the gaps and guessing at the communicative intention (Sperber and Wilson 1986).

Language evolved among hunters and gatherers. Hunter-gatherer communities are at least as civilized as any others (Lee 1988). Artificial kinship structures are designed to ensure that sex, for example, falls under the rule of law (Morgan 1871). Ritual taboos surround matters such as incest or menstruation (Durkheim 1963 [1898]). Protocols and ritual injunctions facilitate sexual circulation – rather in the way that traffic signals in a modern city ensure traffic flow (Lévi-Strauss 1969). Humans need rules if they are to transcend primate-style conflict. The law exists to uphold behavior of a civilized kind.

Language is inseparable from 'society' in this sense. Should law and order collapse, the social magic of words (Bourdieu 1990) would evaporate. Linguists implicitly recognize this when they make one of their most characteristic simplifying assumptions – that of a 'homogenous speech community'. The expression is Chomsky's (1965: 3), but the underlying idea is traceable to Saussure (1974 [1915]). Saussure assumed civilized, democratic conditions – what cricket-players might term 'a level playing field'. This being the case, his concept of the default state for society had nothing in common with Darwin's 'survival of the fittest'. Questions of sex, power, violence and so forth fall outside the framework of rule-governed intercourse – and, therefore, have no place in language as a system. Sending a message, in Saussure's conceptual universe, is very like going to the post office: one must affix the correct stamp. Anticipating Wittgenstein, Saussure likened language in its systematic aspect to a board game such as chess.

Applied to non-human primates, not one fragment of Saussure's semiotic paradigm is relevant in any way. Chimpanzees may co-operate, particularly if related. For much of the time, group members foraging in proximity to one another may well be relaxed and sociable. Periodically, however, a ferocious fight breaks out. This may be settled eventually, but it would be stretching credulity to apply the term 'civilized' to political arrangements of this kind. Apes are under no pressure to operate within the law. They are not like human hunters and gatherers. They do not have formal kinship systems. They do not sustain institutional facts – facts dependent for their existence on communal belief. Whether they compete or co-operate, they inhabit only an unregulated – that is, a

straightforwardly Darwinian – world. In such a world, language is not theoretically possible.

2. In defense of Darwinism

Human society is not Darwinian; yet Darwinism must be used to explain it. In this paper, I am exploring the possibility that language evolved as an *internal feature* of life under the rule of law. The Darwinian thinker John Maynard Smith advanced a proposal along these lines in his 1995 book co-authored with Eörs Szathmáry – *The Major Transitions in Evolution.* Within this theoretical framework, 'society' became established when group-level contracts became evolutionarily stable. It was difficult to achieve this, since the work of enforcement through rewards and punishment requires specific costly investment. Contracts are vulnerable to free-riders – individuals seeking the benefits of co-operation without paying their share of the costs. Somehow, however – in at least one ancestral population – this problem was dealt with. The founding social contract survived (Maynard Smith and Szathmáry 1995).

The rule of law creates a climate of trust (Sztompka 1999). Although competition is not abolished, it assumes a different form – it must proceed within the law. Hunter-gatherers in general are self-governing egalitarians (Lee 1988). Where a community is self-governing, laws are conceptualized in moral terms. Pantheons of other worldly beings – spirits of various kinds – may be employed to represent contractual authorities (Durkheim 1965 [1912]). Alliances committed to enforcement share a common moral purpose and perspective (Alexander 1987). Competition is for moral standing within the community or alliance. Status in this context is freely accorded, not extracted by threat or coercion. Speakers gain social standing by being perceived by their peers as relevant (Dessalles 1998). It is not possible to gain such status by being secretive, devious or inscrutable. Individuals come under pressure to co-ordinate personal perspectives with wider social aims, adopting what might be termed a 'god's eye' view of the world (Boyer 2001). Once having passed through the relevant rite of passage (Van Gennep 1960 [1909]) – once having paid the costs of entry into the ritual community – adults are expected to assume responsibility in this way (Knight 1998, 1999).

When individuals trust one another and share the same perspective, they may communicate effectively without language. A barely perceptible wink can 'speak volumes' on one condition: shared understandings must already be well entrenched. A nod may be regarded as a rudimentary 'up/down' switch; a wink is closer to 'on/off'. It is immediately clear that digital switches of this kind presuppose trust: two parties engaged in a violent argument could not conceivably communicate in so effortless a way. On the other hand, a network of trusting conspirators might well extend their repertoire of switches, recruiting additional parts of the body such as hands, tongue or lips.

Many animals can produce subtle movements of their hands, tongues and lips. Indeed, since mastication and ingestion depends on such mobility, it would be strange if they could not. The odd thing about humans is that such simple motions, evolved for non-communicative functions, have been exapted to serve purposes in communication. To grasp how unnatural and unlikely this is, let us imaginatively enter an unregulated world. Imagine being a motorist caught in a traffic jam at a city intersection where the signals have failed. Using your vehicle's right or left indicator no longer works. Even in a vehicle buzzing with digital equipment it would make no difference whatsoever. Nobody takes the slightest notice; nobody makes way. Your skills in signaling have become quite irrelevant. The explanation is not that the world has become cognitively deficient. It is simply that civilization has temporarily collapsed. The only solution is to use the whole vehicle, pushing, shoving and nudging your way through. Unless you are prepared to risk colliding – a method persuasive only insofar as costs are incurred – you may remain for hours in the same place.

There are no traffic signals or other moral regulators in nature (Dawkins 1989). When chimpanzees vocalize in a public space, they therefore use the whole vehicle. Pant hoots and waa barks are 'body language' in this sense, the audible component inseparable from visually salient and in other ways unmistakable signs. In these as in other indexical displays, the costs of signaling may rise or fall in accordance with risks of exploitation or deceit (Searcy and Nowicki 2005). But in a Darwinian context, no signal can be completely free of charge. Each must pass a test of quality and reliability, evaluated on an analog scale. An inevitable consequence is that evolution in the direction of language is ruled out.

But if ape signals are body language, can we define speech by contrast as 'head language'? Certainly, no other species communicates so narrowly in and through the head. Chomsky is surely right on this point: language is 'internal', offering a window into the autonomous realm of 'mind'. But Chomsky's natural science approach to the study of mind is not the only possible one. It may be better to do social science. In his book, *The Cultural Origins of Human Cognition,* Michael Tomasello (1999) identifies 'mind' with a particular capacity for social cognition. No ape is capable of bi-directional mind reading or continuously maintained joint attention. Apes don't point at things; humans from an early age do. In short: the human brain is wired up to *correlate perspectives*; ape brains are not. Since an ape cannot view itself from another's perspective, it is unable to represent to itself its own mental states. Apes, in short, are incapable of what Tomasello calls 'intersubjectivity'.

For apes, this is adaptive because they inhabit a Darwinian world – one lacking in moral regulation. Each strives to extract information from its rivals while divulging as little as it can in return. The extent to which facial inscrutability is positively selected in primates has been demonstrated by a

remarkable comparative study. The structure of non-human primate eyes – when compared with their human counterparts – is clearly adaptive. Humans divulge direction of gaze thanks to a dark iris rendered salient against the background of a brilliant white sclera. With their dark irises in equally dark surrounds, by contrast, the eyes of mature primates are evidently not designed to give much away. Whereas the human design facilitates openness and transparency, that of adult primates appears specifically designed to *avoid* divulging direction of gaze to nearby conspecifics (Kobayashi and Kohshima 2001).

The structure of the eyes must surely be connected to the fact that although chimpanzees co-operate, their status is not determined by the positive esteem of peers. It is not freely allocated. On the contrary, it is determined by establishing dominance at others' expense, either individually or by forming an alliance. Alliance partners are valued for their fighting ability and talents at Machiavellian scheming. Admittedly, strategies that humans might regard as socially creative do have a place, as De Waal (1996) has been at pains to show. But a consistently honest, open, trusting and socially responsible ape would not get far.

By nature, humans are competitive or co-operative according to need. But at least on a face-to-face level, status competition among humans assumes non-Darwinian forms. Admittedly, class societies – capitalism in particular – appear at least superficially to embody Darwinian principles. But humans in general do not inhabit a Darwinian world. Egalitarian hunter-gatherers long ago transcended that level of complexity. Under what I am here calling 'the rule of law', status must be gained by making contributions valued by the community as a whole.

3. Conclusion

The 'human revolution' culminated in something remarkable – the establishment of the rule of law (Dunbar, Knight and Power 1999). Apes do not recognize group level social contracts. When they signal, consequently, they must intervene in physical and biological reality. There is no other kind of reality in which they *can* intervene. Each animal can make a difference only physically, only with its body – with signals inseparable from the body. Chimpanzee waa-barks and pant-hoots are examples of this. They are indices of bodily and emotional states – direct behavioral interventions, with direct behavioral impact and consequences.

By contrast, a human linguistic utterance – a 'speech act' – is an intervention in a different kind of reality, known by philosophers as 'institutional reality'. 'Institutional facts' exist because it is collectively believed that they do. Examples are money, marriage, football scores and underworlds (Searle 1996). A speech act, like a move in a board game, is internal to reality of this kind (Saussure 1974 [1915]; Wittgenstein 1968). Signals need produce no physical or behavioural impact – only a shift in perspective. A shift of this kind is an alteration within a peculiar domain, neither objective nor subjective in the

ordinary sense. Things are 'seen' or 'judged' differently, and *to that extent* reality (the world as jointly constructed) has changed (Searle 1996). When human life became subject to the rule of law, participation in this kind of reality became possible for the first time. Language first emerged in the form of cryptic 'nods' and 'winks' – mutually understood signs – for navigating within this novel domain. Such entities were no longer evolving in a Darwinian biological environment. The normal laws of signal evolution therefore no longer applied. Wastefulness in signals (Zahavi 1991) was no longer under positive selection. The paradoxes of language evolution correspondingly dissolved.

Today, no-one doubts that language has a genetic component. But respect for the law cannot be explained by this or that instinct, cognitive module or postulated gene. Where behavioural strategies provoke counter-strategies, contrasting strategic outcomes must necessarily emerge on the basis of identical genes. Only once a new strategy has become evolutionarily stable does it in turn alter selection pressures, modifying the trajectory of genetic evolution accordingly. The law cannot be enforced partially or half-heartedly – no such strategy will prove evolutionarily stable. In that sense, our hunter and gatherer ancestors cannot have won the revolution gradually, making progress only through 'descent with modification'. At a certain point, once conditions were ripe, they established the rule of law in the only way they could. For the individuals concerned, it would have been all or nothing. In one small population at least, their resistance to lawlessness succeeded – and turned the world upside down (Mellars and Stringer 1989; Knight 1991; Boehm 2001).

Language is dependent on civilized, rule-governed behavior. This cannot be assumed; it must be explained (Knight, Power and Watts 1995; Power and Aiello 1997; Dunbar, Knight and Power 1999). Ancestral humans surely had good Darwinian reasons to band together in enforcing the rule of law. Any explanation must therefore be in terms of standard Darwinian theory.

References

Alexander, R. D. (1987). *The Biology of Moral Systems.* New York: Aldine de Gruyter.

Austin, J. L. (1978 [1955]). *How to Do Things with Words.* Oxford: Oxford University Press.

Boehm, C. (2001). *Hierarchy in the Forest. The evolution of egalitarian behavior.* Cambridge, MA: Harvard University Press.

Boyer, P. (2001). *Religion Explained.* New York: Basic Books.

Bourdieu, P. (1991). *Language and Symbolic Power.* Cambridge: Polity Press.

Brown, P. and S. C. Levinson (1987 [1978]). *Politeness: some universals in language usage.* Cambridge: Cambridge University Press. Reprinted from: Goody, E. (Ed.), *Questions and politeness: strategies in social interaction* (pp. 56-324). Cambridge: Cambridge University Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, Mass.: MIT Press.

Dawkins, R. (1989). *The Selfish Gene,* 2nd edn. Oxford: Oxford University Press.

Dessalles, J.-L. (1998). 'Altruism, Status and the Origin of Relevance', in J. R. Hurford, M. Studdert-Kennedy and C. Knight (Eds.), *Approaches to the Evolution of Language* (pp. 130-47). Cambridge: Cambridge University Press.

De Waal, F. B. M. (1996). *Good Natured. The origins of right and wrong in humans and other animals.* Cambridge, MA: Harvard University Press.

Dunbar, R., C. Knight and C. Power (Eds.) (1999). *The Evolution of Culture.* Edinburgh: Edinburgh University Press.

Durkheim, E. (1963 [1898]). La prohibition de l'inceste et ses origines. *L'Année Sociologique, 1, 1-70.* Reprinted as *Incest. The nature and origin of the taboo,* trans. E. Sagarin. New York: Stuart.

Durkheim, E. (1965 [1912]). *The Elementary Forms of the Religious Life.* New York: Free Press.

Durkheim Knight, C., C. Power and I. Watts (1995). The human symbolic revolution: A Darwinian account. *Cambridge Archaeological Journal, 5(1),* 75-114.

Grice, H. P. (1989). *Studies in the Way of Words.* Cambridge, MA: Harvard University Press.

Knight, C., (1991). *Blood Relations. Menstruation and the origins of culture.* New Haven and London: Yale University Press.

Knight, C., (1998). Ritual/speech coevolution: a solution to the problem of deception. In J. R. Hurford, M. Studdert-Kennedy and C. Knight (Eds.), *Approaches to the Evolution of Language: Social and cognitive bases* (pp. 68-91). Cambridge: Cambridge University Press.

Knight, C., (1999). Sex and language as pretend-play. In R. Dunbar, C. Knight and C. Power (Eds.), *The Evolution of Culture* (pp. 228-247). Edinburgh: Edinburgh University Press.

Kobayashi, H., and S. Kohshima (2001). 'Unique Morphology of the Human Eye and Its Adaptive Meaning: Comparative Studies on External Morphology of the Primate Eye'. *Journal of Human Evolution, 40/5,* 419-35.

Lee, R. B. (1988). 'Reflections on Primitive Communism', in T. Ingold, D. Riches, and J. Woodburn (Eds.), *Hunters and Gatherers, I. History, Evolution and Social Change* (pp. 252-268). Chicago: Aldine).

Leech, G. (1980). Language and tact. In: *Explorations in semantics and pragmatics. Pragmatics and Beyond 4* (pp. 79-117). Amsterdam: John Benjamins.

Lévi-Strauss, C. (1969), *The Elementary Structures of Kinship.* London: Eyre and Spottiswoode.

Maynard Smith, J. and E. Szathmáry (1995). *The Major Transitions in Evolution.* New York: W. H. Freeman.

Mellars, P. A. and C. Stringer (Eds.), (1989). *The Human Revolution: Behavioural and Biological Perspective in the Origins of Modern Humans.* Edinburgh: Edinburgh University Press.

Morgan, L. H. (1871). *Systems of Consanguinity and Affinity of the Human Family.* Washington: Smithsonian Institution.

Power, C. and L. C. Aiello (1997). Female proto-symbolic strategies. In L. D. Hager (Ed.), *Women in Human Evolution* (pp. 153-171). New York and London: Routledge.

Saussure, F. DE (1974 [1915]). *Course in General Linguistics,* trans. W. Baskin. London: Fontana/Collins.

Searcy, W. A. and S. Nowicki (2005). *The Evolution of Animal Communication.* Princeton and Oxford: Princeton University Press.

Searle, J. R. (1996). *The Construction of Social Reality.* London: Penguin.

Sperber, D. and D. Wilson (1986). *Relevance. Communication and cognition.* Oxford: Blackwell.

Stompka, P. (1999). *Trust.* Cambridge: Cambridge University Press.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition.* Cambridge, MA: Harvard University Press.

Van Gennep, A. (1960 [1909]). *The Rites of Passage.* London: Routledge and Kegan Paul.

Wittgenstein, L. (1968). *Philosophical Investigations,* trans. G. E. M. Anscombe. Oxford: Blackwell.

Zahavi, A. (1991). On the definition of sexual selection, Fisher's model, and the evolution of waste and of signals in general. *Animal Behaviour, 42,* 501-3.

A SALTATIONIST APPROACH FOR THE EVOLUTION OF HUMAN COGNITION AND LANGUAGE

SUSAN J. LANYON

Cognitive Science Program, School of History and Philosophy of Science, University of New South Wales, Sydney, NSW 2052, Australia
susanlanyon@bigpond.com

The debate over the evolution of an innate language capacity seems to divide into two principle camps. The neo-Darwinian approach generally argues that human psychological modules, including the language faculty, must have arisen gradually and incrementally having been honed by natural selection. Thus Pinker, when theorizing about language evolution, sees "no reason to doubt that the principle explanation is the same as for any other complex instinct or organ, Darwin's theory of natural selection" (Pinker, 1994, 333). However, as Knight et al. (2000) have pointed out, little attention has been paid by the neo-Darwinian approach to address the causes of the emergence of novelty. The saltationist approach gleans much of its evidence from the archaeological and paleontological record, which is interpreted as unsupportive of the neo-Darwinian paradigm. Jackendoff (1999) accuses those who do not accept that language arose gradually through natural selection as having been "forced to devalue evolutionary argumentation". Jackendoff's concern seems to stem from the view that there is only one way that evolution can proceed, through gradual change driven by natural selection. My concern is for the neglect of the vast amount of evidence supporting the theory that modern humans did not emerged in a gradual, step-wise fashion, so there is no reason to believe that cognition and language evolved in this manner. Here I argue that hominins evolved through major evolutionary leaps, which may have numbered only two or three significant mutation "events". Neoteny (the retention of infant or juvenile growth rates) appears to have been a major force in the evolution of our primate ancestors and this process can explain the sudden emergence of many of the traits that define what it means to be human. Further evidence from the fossil and archaeological record supports a "sudden" emergence of human cognition and language.

1. Evolution - Gradual vs Saltational

Evolution by natural selection in the Darwinian tradition is the most commonly attested theory for the evolution of modern humans. This theory generally assumes that human evolution, including the evolution of the brain, has progressed incrementally as a trend toward complexity. Regardless of the "unit of selection", be it the phenotype or an individual gene, the neo-Darwinian model emphasizes gradualism. Numerous neo-Darwinian evolutionary accounts have proposed that traits like bipedalism, a large brain, reduced dentition, language, tool use and other characteristics of hominins, must have evolved together while maintaining some sort of feedback mechanism that has "fine tuned" or "gradually" adapted each

characteristic feature to a changing environment.

Thus Pinker has argued that since the emergence of bipedalism a few million years ago, and with the hands freed, subsequent species ratchet upward, click by click, in the features that distinguish us, like the size of the brain and the sophistication of tools. Pinker has interpreted the fossil record under a gradualist scenario with a reduction in the brow ridges, the teeth and the jaw because "tools and technology have taken over from teeth" (Pinker, 1997, 200). However, the reduction in size of the ape-like "dagger-like" canines in the first hominins, can be explained by the process of neoteny. The canines are the last teeth to fully form in juvenile chimpanzees and appear to have not reached their full potential in australopithecines, who appear to have partly skipped the juvenile growth spurt that is part of the life-history of other apes. Although there appear to be some subtle differences in developmental pathways between australopithecines and chimpanzees (Kuykendall 2003), it is generally accepted that the life history of this genus was generally ape-like. Moreover, the first stone tools appear around 4 million years *after* bipedalism and small canines emerged, hardly a "click by click" scenario. The brain of australopithecine did not increase in size from that of a chimpanzee over this vast period of time. It should therefore not be surprising that we cannot find any form of human-like behaviour during this era of hominin evolution.

Following the "gradualist" school, Jackendoff (1999) claims to be able to decompose modern language into partially ordered steps that have evolved incrementally to be finally integrated into the larger combinatorial system. These modules are assumed to be innate, localized and part of our genetic inheritance. His claim that the language faculty exhibits a degree of modularity is not disputed. Jenkins (2000, 145), who also questions the gradualist approach, nevertheless accepts that the language faculty seems to contain modules such as a lexicon, a computational component, semantics, morphology, the phonological component, and phonetics. Jackendoff's intuition that any increase in explicit expressive power of the communication system is adaptive, whether for cooperation in hunting, gathering, defense, or for social communication such as gossip, is also not disputed. But, as Bickerton points out, "there is simply not one scintilla of evidence: simply a blind faith that, if evolution is gradual, and we are where we are, we must have got here, far as it may seem, in a series of incremental steps" (Bickerton, 2002, 104). Rather, the saltationist approach to the evolution of the human ancestral line matches closely with the paleontological evidence, whereas the neo-Darwinian approach, is quite contradictory to that evidence.

2. Evolution through Developmental Change

Rather than evolving in gradual, incremental steps from an ancestral primate, the hominin clade appears to have arisen due to the heterochronic variation in form during the early stages of development producing the novelty that we see in the first hominins as well as modern humans. Lovejoy et al. (1999) have argued that

whilst systematic assembly mechanisms for the formation of various functional anatomical parts of the mammalian *bauplan* appear to be highly conserved, subtle changes in developmental rate or timing can produce profound morphological changes in phenotype. They have shown how a simple dimensional change in the development of the chimpanzee pelvis could have produced the anatomical differences found in australopithecines. Also, the neck of the hominid femur together with the human knee can be explained by a simple heterochronic change in conjunction with cartilage plasticity and habitual bipedal gait.

The affinity of the juvenile form of both chimpanzees and ancestral hominins, as well as the likeness of some *H. erectus* juvenile hominins and modern humans, shows that most of the skeletal traits of the hominin line arose primarily due to changes in developmental timing in the early stages of ontogeny. Neoteny has been a major factor in the evolution of ancestral hominins. It is interesting to note that the first discoveries and classification of *Australopithecus africanus* by Dart in the first half of the twentieth century, were rejected by many, claiming that these remains were those of a juvenile gorilla.

Chimpanzees are born with the skull positioned on the top of the spinal column, just like human infants. The newborn chimpanzee is bipedal albeit with a bent-knee, bent-hip gait similar to that posited for australopithecine and even perhaps Neanderthals. The skull of the chimpanzee is also similar to that of a human with a flat face, without brow ridges, without cranial crests, same position of orbits, thinness of bone and a high domed forehead. The infant chimpanzee also has a small human-like jaw and palate. All of these human-like features change dramatically during the juvenile growth spurt of the chimpanzee. Penin et al. (2002) have recently confirmed the hypothesis for the neotenic emergence of the human skull shape. Their results show that the adult skull in humans reaches a size-related *shape* that is equivalent to that of a juvenile chimpanzee but nevertheless grows to a size similar to that of an adult chimpanzee. Further Antón and Leigh (2003) have found that the cranium of anatomically modern *H. sapiens* can be considered to be paeodomorphic as it retains the same shape of the juvenile *H. erectus* (Mojokerto child).

Most of the skeletal traits that separate us from the apes, namely bipedalism (with the associated modifications of the spine, pelvis, femur and knee), skull and face shape, and dentition, can be explained by the process of neoteny. The "final" mutation, which seems to have occurred in an ancestral species living in Africa, profoundly modified the neural architecture of the human brain, which gave rise to the sudden appearance of our modern symbolic thought and language capacity. Although we do not know precisely how this apparent massive increase in cognitive ability emerged in the human brain, we can determine certain allometric departures in the human brain. Rilling and Seligman (2002) have shown that although the frontal lobes of the human brain are not larger than expected for an ape of our size, the pre-frontal sector is disproportionately large. The human

temporal lobe white matter volume is significantly expanded to produce a different shape and size with many more connections linking temporal and prefrontal cortex. This extra-allometric expansion appears to be specifically involved with speech production and comprehension necessary for language. Most of this expansion occurs postnatally in humans and a second cycle of development in this region of the brain begins around two years of age in human infants (Pollack, 1994, 165).

3. The first "Sudden Appearance"

The earliest fossil hominins dating to around 6-7 million years ago are found in Africa where they existed up until 1.2 million years ago (Lahr & Foley, 2004). The defining features of these australopithecines were bipedalism, small stature (1-1.5 metres tall) and reduced sized canines. The australopithecines show a mosaic of ape-like and human-like features including tooth size and shape, face shape, brow ridges, brain case size, femur shape, pelvic size and shape and various configurations of foot bones that may mean full bipedalism or some retention of the ability to retain an aboreal, climbing lifestyle.

The most plausible account for the emergence of the first hominin is for a simple heterochronic mutation that changed the developmental pathway of a chimpanzee/s that in particular, did not go through the normal juvenile growth spurt thereby retaining the bipedalism, skull shape and reduced canines of its infancy. This creature, which we can assume had the same, or very close, genetic make-up of its progenerator, may have interbred with its original chimpanzee group to produce most of the variation that we now find in the fossil record over a five million year time span. It is interesting to note that these hominins retained chimpanzee-sized brains throughout their approximately 5 million year history (Lahr & Foley, 2004). Importantly, there are no stone tools associated with australopithecines.

The implications of this stasis for the evolution of human-like cognition and language are quite profound. It is generally agreed that a fundamental requirement for the use of language is the capacity for symbolic thought. Sterelny (2003) stresses the fact that we rely on the symbolic meaning in each of our utterances to be understood by our audience. We need a rich theory of mind for language to operate successfully. Despite many years of experimentation with Great Apes, there is little evidence that they have rich ways of representing the minds of others. We therefore have no reason to suppose that the australopithecines, during their approximately 4-5 million years of existence as a species, had made any major cognitive advance over the Great Apes and were probably not capable of using any sort of proto-language. This conclusion is difficult to reconcile with the notion that language has evolved gradually and incrementally, as Jackendoff and Pinker claim. The more compelling evidence from the paleontological and archaeological record reveals an astonishing stasis for this great expanse of time.

4. *Homo erectus*

Around 2 million years ago, after around 4-5 million years of stasis, *H. erectus*, a generally much larger hominin than the australopithecines and the first tool-maker[a], first appears in Africa. Shortly after their emergence, they migrated to most other accessible continents. The stone tools found with this species were effectively identical to the first stone tools that appeared in the record 500,000 years earlier (Tattersall, 1997). This suggests that although this early *H. erectus* population was larger in physical size and cranial capacity than *H. habilis*, it had not developed any notable increase in cognitive ability.

After extensive analysis of the fossil record of most of the *H. erectus* remains covering Africa, China, Europe and South East Asia, Rightmire (1990) has concluded that this species had maintained a conservative morphology throughout the Pleistocene. *H. erectus* features vary, but these variations, especially in brain size, were not on a scale greater than those found within contemporary modern human populations. These phenotypic differences can be found within geographically widespread groups of *H. erectus* and cannot be placed within a specific time-line of size or complexity, arguing against a gradual change within groups or arising from intergroup competition. Despite the obvious increase in cranial capacity around 2 million years ago, Walker & Shipman (1996) argue that any difference in brain size from *H. habilis* to *H. erectus* was simply due to an increase in body size. *H. erectus* had a body size slightly larger than most modern humans, but nevertheless had only a brain the size of a one year old human.

Pinker (1997) has asserted that an increase in the sophistication of stone tools is one of the keys to understanding the gradual increase in human cognitive complexity. Around 1.4 million years ago, a relatively standardized and symmetrical stone tool was being made in Africa (Tattersall, 1997). We could concede that this probably represents an increase in cognitive ability for our hominin ancestors although even this proposed advance in planning ability has been challenged by some. For example, Ambrose (2001) has suggested that the Acheulian hand-axes are more likely to be the "unintended byproduct of nonstylistic factors rather than intended target types". When raw materials are particularly scarce, a cutting tool would need to be resharpened rather than discarded and we can see how a simple cleaver becomes a teardrop-shaped hand-axe after several bouts of resharpening. Moreover, this tear-drop shaped acheulian hand-axe, known for its distinctive shape, and associated with *H. erectus*, remained unchanged for one million years. It is found wherever *H. erectus* roamed, which covered a distance of 10,000 miles from Africa to Japan. Bickerton asks the question, "is it possible to think of any *sapiens* innovation that has traveled for the best part of 10 thousand miles without undergoing the slightest change?". He adds that "the mismatch between the

[a]The first stone tool makers *H. habilis*, were hardly bigger or more "advanced" in their body skeletons than the australopithecines.

fossil and archaeological records forms an acute embarrassment for those who believe that human cognitive capacities, including language, developed gradually" (Bickerton, 2002, 106).

Pinker and Jackendoff argue that the larynx was recently adapted for speech. MacLarnon & Hewitt (2004) note that shape of the basicranium of *H. erectus* indicates that the larynx and hyoid were both situated in the same position as in apes. However Pinker and Jackendoff overlook this fact when claiming that the descended larynx is part of a suite of vocal-tract modifications that were evolutionarily shaped to "subserve" language. There is no gradual evolution of the morphology of the vocal tract that is being "fine-tuned" to serve language.

A further mismatch between advanced anatomy and cognitive behaviour has been found with recent excavations of a 800,000 year old cave site in Atapuerca, Spain. A juvenile hominin fossil has the face of a totally modern human, and Arsuaga (1999) argues that this face may explain how modern humans emerged due to a mutation that caused the retention of juvenile morphology (neoteny) of this ancient species, which he has named *Homo antecessor*. However, the tools found alongside *H. antecessor* were archaic for their time, resembling the 2.5-2.6 Myr old Oldowan tradition from Africa.

5. The Appearance of Modern Humans

The first appearance of anatomically modern humans emerged in Africa by around 120,000 years ago (Stringer, 2003). The fossil record in Africa closely matches the latest genetic evidence for this recent emergence. Pearson (2004) has pointed out that MtDNA and the X and Y chromosomes in all modern humans show a very low diversity, especially compared to Apes. It appears that an anatomically modern population emerged in Africa and subsequently replaced all other hominin populations living there at that time. Pearson suggests that these first modern humans slowly grew in population over a period of 60,000 years, but remained in Africa, explaining the greater genetic diversity that is found there today. After 60,000 years, these modern peoples then dispersed to the Near East, South East Asia and Australia, and then on to Europe and the Far East.

Pearson's findings are matched by clear evidence from the archaeological record, that most of what we recognize as fully modern human intentional and symbolic behavior, arose suddenly in southern Africa, around 90,000 to 100,000 years ago (Mellars, 2005). New forms of skin working technology appear along with highly specialized geometric blades, used as insets in multi-component hafted tools, together with intricately shaped barbed bone points. High quality raw materials had been deliberately transported from a distance of at least 20 km. Many sites have an abundance of red ochre showing signs of scraped surfaces indicating their use as coloring pigments. Other large pieces of ochre have been deliberately incised with repetitive geometrical patterns from which we can infer some sort of symbolic or ceremonial activity (Mellars, 2005). Large amounts of

red ochre are also associated with ceremonial burial. We find an abundance of carefully perforated shells, which again had been imported from long distances. Their use as personal ornaments is confirmed by unambiguous indications of elaborate ceremonial burials associated with a range of perforated seashells (Mellars, 2005). We cannot say whether or not these first humans had language, but we can be sure that they were engaging in symbolic behaviour, making it highly likely that some form of language was at least possible.

6. Conclusion

I propose that the archaeological and paleontological record does not support theories that argue for gradual change for the evolution of the hominin clade. Rather, we see the sudden appearance of the first hominins around 6-7 million years ago due to a neotenous mutation of an ancestral ape, followed by an astonishing stasis in bodily form and brain size for at least 4 million years when the first crude stone tools appear. We then find the sudden appearance of *H. erectus* who again shows little variability in either physiology or tool technology for 2.5 million years. It has been shown that any variability within all of these ancestral hominin groups is no more than we find today among modern humans. The final sudden appearance produced anatomically modern humans (*H. sapiens*), which emerged in Africa around 120,000 years ago with the cognitive architecture to support an extraordinary array of symbolic behaviour not seen before in any ancestral species. I therefore contend that arguments from the gradualist school of thought claiming that our mind and language faculty are highly modularised due to the gradual accretion of functionally specific components, which have evolved gradually over evolutionary time, are not tenable.

References

Ambrose, S. H. (2001). Paleolithic Technology and Human Evolution. *Science, 291*, 1746-1753.

Antón, S., & Leigh, S. (2003). Growth and life history in Homo *erectus*. In Thompson (2003) (p. 219-245). Cambridge: Cambridge University Press.

Arsuaga, J., Martinez, I., Lorenzo, C., & Gracia, A. (1999). The Human Cranial Remains from Gran Dolina Lower Pleistocene Site. *Journal of Human Evolution, 37*, 431-457.

Bickerton, D. (2002). From protolanguage to language. In Crow (2002) (p. 103-120). UK: Oxford University Press.

Crow, T. (2002). *The speciation of modern homo sapiens.* UK: Oxford University Press.

Jackendoff, R. (1999). Possible Stages in the Evolution of the Language Capacity. *Trends in Cognitive Sciences, 3*, 272-279.

Knight, C., Studdert-Kennedy, M., & Hurford, J. (2000). *The evolutionary emergence of language.* Cambridge, UK: Cambridge University Press.

Kuykendall, K. (2003). Reconstructing australopithecine growth and development: What do we think we know? In Thompson (2003) (p. 191-218). Cambridge: Cambridge University Press.

Lahr, M. M., & Foley, R. (2004). Palaeoanthropology: Human Evolution Writ Small. *Nature, 431*, 1043-1044.

Lovejoy, C., Cohn, M., & White, T. (1999). Morphological Analysis of the Mammalian Postcranium: A Developmental Perspective. *Proc. Natl. Acad. Sci. USA, 96*, 13247-13252.

MacLarnon, A., & Hewitt, G. (2004). Increased Breathing Control: Another Factor in the Evolution of Human Language. *Evolutionary Anthropology, 13*, 181-197.

Mellars, P. (2005). The Impossible Coincidence. A Single-Species Model for the Origins of Modern Human Behavior in Europe. *Evolutionary Anthropology, 14*, 12-27.

Pearson, O. (2004). Has the Combination of Genetic and Fossil Evidence Solved the Riddle of Modern Human Origins? *Evolutionary Anthropology, 13*, 145-159.

Penin, X., Berge, C., & Baylac, M. (2002). Ontogenetic Study of the Skull in Modern Humans and the Common Chimpanzees: Neotenic Hypothesis Reconsidered with a Tridimensional Procrustes Analysis. *American Journal of Physical Anthropology, 118*, 50-62.

Pinker, S. (1994). *The language instinct.* NY: William Morrow & Co.

Pinker, S. (1997). *How the mind works.* Australia: Penguin.

Pollack, R. (1994). *Signs of life.* Boston, US.: Houghton Mifflin.

Rightmire, G. (1990). *The evolution of homo erectus.* Uk: Cambridge University Press.

Rilling, J., & Seligman, R. (2002). A quantitative morphometric comparative analysis of the primate temporal lobe. *Journal of Human Evolution, 42*, 505-533.

Sterelny, K. (2003). *Thought in a hostile world.* UK: Blackwell.

Stringer, C. (2003). Out of Ethiopia. *Nature, 423*, 692-694.

Tattersall, I. (1997). Out of Africa Again ... and Again? *Scientific American, April*, 60-67.

Thompson, J., Krovitz, G., & Nelson, A. (2003). *Patterns of Growth and Development in the Genus* Homo. Cambridge: Cambridge University Press.

Walker, A., & Shipman, P. (1996). *The wisdom of the bones.* NY: Alfred A. Knopf.

INTERACTION OF DEVELOPMENTAL AND EVOLUTIONARY PROCESSES IN THE EMERGENCE OF SPOKEN LANGUAGE

JOHN L. LOCKE

Department of Speech-Language-Hearing Science, Lehman College, City University of New York, 250 Bedford Park Blvd. West, Bronx, New York 10468, USA

Evolution is a two-stage process (West-Eberhard 2003). In the first stage, a plastic phenotype responds to environmental variation, producing novel forms that vary genetically. In the second stage, selection acts on the variants. From Mivart (1871) and Garstang (1922) to the new evolutionary developmental biologists (Northcutt, 1990), it has been argued that the first stage---the origin of novel characters---can only occur in development. While vigorously supporting this view, the new Evolutionary Developmental Biology (termed "evo-devo") has also demonstrated reverse order ("devo-evo") effects of evolution on development (Gilbert, 2003; Hall, 2000; Jablonka & Lamb, 1998; Wagner et al., 2000). Indeed, for some, each of the two processes---evolution and development---is an *explanandum* as well as an *explanans* (Robert, 2002).

Since traits that emerge in development frequently do so as the result of a change in behavior (Bateson, 1988; Gottlieb, 2002), it is essential that theories of linguistic evolution place a high valuation on developing behaviors. Recently, in collaboration with Barry Bogin, I have argued that evolution produced a new ontogenetic stage and remodeled an existing one; and that developmental processes, operating in the new and remodeled stages, produced novel behaviors that were naturally or sexually selected (Locke & Bogin, in press). The net effect of these evo-devo and devo-evo changes, as proposed, was an increase in the frequency of genes supporting precursors to, and ultimately the components of, spoken language.

An example of a revision of an existing ontogenetic stage is the effect of bipedalism (and antecedent events) on infancy. It has been suggested that this change redesigned the hominid nervous system (Eccles, 1989) and introduced new ways of interacting (Jablonski et al., 2002). Bipedalism also narrowed the pelvis (Leutenegger, 1980). This caused an "obstetrical dilemma" for the expectant mother and her large-headed fetus, a dilemma that was solved by shifting skull and brain development into the postnatal period. This shift increased helplessness, extending the period and intensity of interactions between offspring and their vigilant parents.

New levels of care and sibling competition would have escalated conflict

between the infant and its parents (cf. Trivers, 1972). One solution was for infants to signal for care more strategically. I propose that infants who issued more effective care signals were more likely than others to receive care and live to reproductive age (Locke, in press). I also propose that infants who cooed and babbled at appropriate intervals were unusually likely to engage with adults, to receive more sophisticated forms of care as infancy progressed, and to generate and learn complex phonetic patterns. These benefits would have accrued particularly to infants who monitored adult reactions (Chisolm, 2003) and adjusted their vocal output accordingly.

At some point, it became possible for infants to "cry wolf," that is, manipulate their voices in such a way as to appear more needy or worthy than they really were (Hauser 1986). If mothers wanted to devote more time to other infants and tasks, they would have had to monitor their infants' vocalizations more carefully, and learn to discriminate the sound of tactical signals from sincere ones. Thus, I suggest that the increase in infants' helplessness would also have enhanced parental ability to interpret infant vocalizations (Brockway, 2003; Locke & Bogin, in press).

An example of a new ontogenetic stage is childhood, a uniquely human stage that entered the *Homo* line about two million years ago (Bogin, 2001; 2003). Coterminous with weaning, chimpanzee infancies last five years. During this period, maternal lactation suppresses ovulation, limiting the rate of population growth. Hominid mothers weaned their infants earlier, decreasing inter-birth spacing and increasing the number of possible offspring. The years liberated by an earlier weaning created a short, two-year childhood, with different characteristics than infancy and the juvenile stage that follows. These characteristics would have favored the invention of vocal and symbolic behaviors by the young (Locke & Bogin, in press), conferring benefits that contributed to the extension of childhood, additionally, to its present four-year duration.

New childcare pressures would also have increased reliance on *surrogate* parents, or babysitters (Hrdy, 1999). In traditional societies---the social arrangements most closely resembling the environments of evolutionary adaptedness---cooperative breeding is essential. A possible linguistic benefit was decontextualization---a special feature of human language (Hockett, 1977)---inasmuch as infants would have encountered a wider range of individuals who (a) knew less about them, (b) operated on broader and less certain schedules of caregiving, and (c) felt less responsibility for them than the mother would have. A natural result, presumably, was an added measure of vocal and communicative

flexibility, and increased ability to manipulate, and read, caregiver intentions.

The effects of an altered infancy and new childhood would also have jacked up the value of parental instruction. In human societies, the young are exposed to a range of potentially dangerous objects and conditions, new risks emerging with the development of walking and other motor functions. Those who pointed and vocalized in response to visual attractions would have learned more about their environment, and negotiated those environments safely and successfully (cf. Caro & Hauser, 1992). It is proposed that the joint use of manual and vocal signals increased fitness over the course of numerous ancestral ontogenies. A secondary effect was increased command of vocal behavior, for there is evidence in modern infants that manual activity increases the frequency and syllabicity of vocalization (Ejiri & Masataka, 2001; Iverson & Fagan, 2004), and it has been reported recently that manual-vocal combinations are unusually likely where the infant's intentions are communicative (Locke, in submission). Thus it is claimed that motoric and referential functions jointly contributed to our species' volubility, and roused the mandibular and articulatory systems responsible for the production of speech-like sounds.

Of course, sounds without symbols only go so far (Hurford, 2004). Fortunately, the developments discussed above would also have favored the evolution of sound-meaning relationships. This issue was partially addressed by Fitch (2004), who argued that parents would have benefited by communicating accurate information to their offspring. The better informed young, according to Fitch's functional proposal, would be more likely to survive into adulthood, passing on to their own offspring genes associated with the improved system of communication. Continuity being a hallmark of human development, infants and children who achieved effective use of sound-meaning signals would have carried some form of the relevant control behaviors into juvenility and adolescence. In those stages, I claim, vocal and verbal skills facilitated the quest for status and sex (Locke, 2000; Locke & Bogin, in press), selection automatically strengthening---in a second hit---precursive behaviors that persisted, in some form, from earlier stages.

There is evidence across all the stages of life history to the effect that speech attracts attention, that attention raises status, and that highly vocal individuals enjoy higher status than less vocal individuals. Dominance hierarchies begin to form as early as five years (Strayer & Trudel, 1984), and these are based largely on vocal and verbal behavior (Hold-Cavell & Borzutsky, 1986). Among the status-enhancing developments, depending on the culture, are riddling (McDowell, 1979), joking (McGhee, 1979; Shultz & Horibe, 1974), and

dueling rhymes (Dundes et al., 1970). In 3- to 5-year old African American boys, Wyatt (1995; 1999) has observed an elementary form of "the dozens," a duel---usually "fought" by adolescents and young adult males---and a clear case of "rap."

I suggest that in traditional (oral) societies, the vocal and verbal abilities that enabled adolescents and young adults to engage and compete with others, and to perform in public arenas, indexed fitness and were selected. If so, there may be a genetic basis for verbal expressivity and dominance in modern humans, a prospect supported by several adoptive twin studies (Gangestad & Simpson, 1993; Lykken, 1982; see also Snyder, 1987).

My claim, then, is that insertion of new or remodeled ontogenetic stages into human life history produced new developmental processes that fashioned novel communicative behaviors, and that these increased fitness. If so, reciprocal action by evolutionary and developmental mechanisms---specifically an "evo-devo-evo" sequence---may have played a major role in the evolution of language.

References

Bateson, P. (1988). The active role of behaviour in evolution. In Ho, M.-W., & Fox, S. W. (Eds.), *Evolutionary processes and metaphors*. London: Wiley.

Bogin, B. (2001). *The growth of humanity*. New York: Wiley-Liss.

Bogin, B. (2003). The human pattern of growth and development in paleontological perspective. In Thompson, J. L., Krovitz, G. E. & Nelson, A. J. (eds.) *Patterns of growth and development in the Genus* Homo. Cambridge: Cambridge University Press.

Brockway, R. (2003). Evolving to be mentalists: the "mind-reading mums" hypothesis. In Sterelny, K., & Fitness, J. (Eds.), *From mating to mentality: evaluating evolutionary psychology*. New York: Psychology Press.

Caro, T. M., & Hauser, M. D. (1992). Is there teaching in nonhuman animals? *Quarterly Review of Biology, 67,* 151-174.

Chisolm, J. S. (2003). Uncertainty, contingency, and attachment: a life history theory of theory of mind. In Sterelny, K., & Fitness, J. (Eds.), *From mating to mentality: evaluating evolutionary psychology*. New York: Psychology Press.

Dundes, A., Leach, J. W., & Özkök, B. (1970). The strategy of Turkish boys' verbal dueling rhymes. In Gumperz, J. J., & Hymes, D. (Eds.), *Directions in sociolinguistics: the ethnography of communication*. New York: Holt, Rinehart and Winston.

Eccles, J. C. (1989). *Evolution of the brain*. London: Routledge.

Ejiri, K., & Masataka, N. (2001). Co-occurrence of preverbal vocal behavior

and motor action in early infancy. *Developmental Science, 4*, 40-48.

Fitch, W. T. (2004). Kin selection and "mother tongues": a neglected component in language evolution. In Oller, D. K., & Griebel, U. (Eds.), *The evolution of communication systems: a comparative approach*. Cambridge, MA: MIT Press.

Gangestad, S. W., & Simpson, J. A. (1993). Development of a scale measuring genetic variation related to expressive control. *Journal of Personality, 61*, 133-158.

Garstang, W. (1922). The theory of recapitulation: a critical re-statement of the biogenetic law. *Journal of the Linnaean Society (Zoology), 35*, 81-101.

Gilbert, S. F. (2003). Evo-devo, devo-evo, and devgen-popgen. *Biology and Philosophy, 18*, 347-352.

Gottlieb, G (2002). *Individual development and evolution: the genesis of novel behavior*. Mahwah, NJ: Erlbaum. Originally published by Oxford University Press in 1992.

Hall, B. K. (2000). Evo-devo or devo-evo---does it matter? *Evolution & Development, 2*, 177-178.

Hauser, M. D. (1986). Parent-offspring conflict: care-elicitation behaviour and the 'cry-wolf' syndrome. In Else, J. G., & Lee, P. C. (Eds.), *Primate ontogeny, cognition and social behaviour*. Cambridge: Cambridge University Press.

Hockett, C. F. (1977). *The view from language: selected essays 1948-1974*. Athens, GA: University of Georgia Press.

Hold-Cavell, B. C. L., & Borsutzky, D. (1986). Strategies to obtain high regard: longitudinal study of a group of preschool children. *Ethology & Sociobiology, 7*, 39-56.

Hrdy, S. B. (1999). *Mother nature: natural selection and the female of the species*. London: Chatto & Windus.

Hurford, J. R. (2004). Language beyond our grasp: what mirror neurons can, and cannot, do for the evolution of language. In Oller, D. K., & Greibel, U. (Eds.), *Evolution of communication systems: a comparative approach*. Cambridge, MA: The MIT Press.

Iverson, J. M., & Fagan, M. K. (2004). Infant vocal-motor coordination: precursor to the gesture-speech system? *Child Development, 75*, 1053-1066.

Jablonka, E., & Lamb, M. J. (1998). Bridges between development and evolution. *Biology and Philosophy, 13*, 119-124.

Jablonski, N. G., Chaplin, G., & McNamara, K. J. (2002). Natural selection and the evolution of hominid patterns of growth and development. In Minugh-Purvis, N., & McNamara, K. J. (Eds.), *Human evolution through developmental change*. Baltimore, MD: Johns Hopkins University Press.

Leutenegger, W. (1980). Encephalization and obstetrics in primates with particular reference to human evolution. In Armstrong, E., & Falk, D. (Eds.), *Primate brain evolution: methods and concepts*. New York:

Plenum.
Locke, J. L. (2000). Rank and relationships in the evolution of spoken language. *Journal of the Royal Anthropological Institute, 7,* 37-50.
Locke, J. L. (in press). Parental selection of vocal behavior: crying, cooing, babbling, and the evolution of language. *Human Nature.*
Locke, J. L. (in submission). Bimodal signaling in infancy: motor behavior, reference, and the evolution of spoken language.
Locke, J. L., & Bogin, B. (in press). Language and life history: a new perspective on the evolution and development of linguistic communication. *Behavioral and Brain Science.*
Lykken, D. T. (1982). Research with twins: the concept of emergenesis. *Psychophysiology, 19,* 361-373.
McDowell, J. H. (1979). *Children's riddling.* Bloomington, IN: Indiana University Press.
McGhee, P. E. (1979). *Humor: its origin and development.* San Francisco, CA: W. H. Freeman.
Mivart, St. G. (1871). *On the genesis of species.* London: Macmillan.
Northcutt, R. G. (1990). Ontogeny and phylogeny: a re-evaluation of conceptual relationships and some applications. *Brain, Behavior and Evolution, 36,* (2-3), 116-140.
Robert, J. S. (2002). How developmental is evolutionary developmental biology? *Biology and Philosophy, 17,* 591-611.
Shultz, T. R., & Horibe, F. (1974). Development of the appreciation of verbal jokes. *Developmental Psychology, 10,* 13-20.
Snyder, M. (1987). *Private realities/public appearances: the psychology of self-monitoring.* New York: Freeman.
Strayer, F. F., & Trudel, M. (1984). Developmental changes in the nature and function of social dominance among young children. *Ethology and Sociobiology, 5,* 279-295.
Trivers, R. L. (1972). Parental investment and sexual selection. In Campbell, B. (Ed.), *Sexual selection and the descent of man, 1871-1971.* Chicago, IL: Aldine.
Wagner, G. P., Chiu, C-H., Laubichler, M., & Hansen, T. F. (2000). Developmental evolution as a mechanistic science: the inference from developmental mechanisms to evolutionary processes. *American Zoologist, 40,* 819-831.
West-Eberhard, M. J. (2003). *Developmental plasticity and evolution.* Oxford: Oxford University Press.
Wyatt, T. A. (1995). Language development in African American English child speech. *Linguistics and Education, 7,* 7-22.
Wyatt, T. A. (1999). An Afro-centered view of communicative competence. In Kovarsky, D., Duchan, J. F., & Maxwell, M. (Eds.), *Constructing incompetence: disabling evaluations in clinical and social interaction.* Mahwah, NJ: Erlbaum.

LABELS FACILITATE LEARNING OF NOVEL CATEGORIES

GARY LUPYAN
Carnegie Mellon University Department of Psychology,
Center for the Neural Basis of Cognition
342C Baker Hall,
Pittsburgh PA , 15217, USA

A major feature that sets language apart from other communication systems is the use of categorylabels—words. In addition to providing a means of communication, there is growing evidence that category labels play a role in the formation and shaping of concepts. If verbal labels help humans acquire or use category information, one can ask whether it is easier to learn labeled categories compared to unlabeled ones. Normal English-speaking adults participated in a category-learning task in which categories were labeled or unlabeled. The presence of labels facilitated the learning of unfamiliar categories and resulted in more robust category representations. The advantage for acquiring named categories was observed even though the category labels did not convey any additional information and all participants had equivalent experience categorizing the stimuli. This work provides empirical support for the idea of labels as conceptual anchor points (Clark, 1997).

1. Introduction

Humans are the only animals to associate conceptual categories with words. Clearly, this association allows linguistic communication. Additionally, words have been argued to stabilize abstract ideas in working memory (Clark, 1997) and make them available for inspection (Clark & Karmiloff-Smith, 1993; Rumelhart, Smolensky, McClelland, & Hinton, 1986; Vygotsky, 1962). These hypotheses are seductive, in part, because they suggest that words may have evolved, in part, as a cognitive aid rather than just as a means of communication. At present, no empirical evidence exists showing that learning names for categories facilitates their acquisition and representation. The present work provides just such evidence.

The question of the effect of labels on category-learning is not new and has been investigated extensively in the context of children's language acquisition. For instance, Waxman and Markow have (1995) have shown that superordinate labels such as "vehicle" can serve as "invitations" for children to form the appropriate category. One may rightfully point out that it is through words that we come to know what categories are relevant and learning words can guide our attention to the meaningful aspects of our environment. The present question, then, is not whether labels facilitate category-formation because they point out

the relevant categories (i.e., the label "chair" suggests that chairs are a useful and relevant category of objects), but whether named categories are easier to acquire *because* they have a name. Given a set of new objects organized, and equal experience categorizing them, are people who learn names for the categories better able to categorize the objects?

In support of Clark's notion that labels may act to stabilize abstract ideas, recent cross-cultural findings have provided evidence that language is closely linked to the human ability to appreciate exact numerosities (Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004) such that entertaining the concept "exactly 15" may rely on having a discrete label "15" (or the ability to create an exact label by iterative means). There is also evidence that in anomia and aphasia more generally—neuropsychological conditions affecting language and the ability to produce names—the ability to form and act based on category knowledge may be severely compromised (De Renzi, Faglioni, Savoiardo, & Vignolo, 1966; Hjelmquist, 1989; Roberson, Davidoff, & Braisby, 1999). While such evidence is suggestive of cognitive roles played by words, a controlled comparison of category-learning with and without labels in normal English-speaking adults is currently lacking. If labels enable humans to form conceptual representations that would otherwise be more difficult or impossible to form by serving as "conceptual anchors," (Clark, 1997) one can ask whether it is easier to learn labeled categories compared to unlabeled ones. Crucially, because the present work relies on participants learning novel categories, it is possible to control stimulus familiarity (something not possible in the types of cross-cultural arithmetic investigations cited above) such that all participants have equal experience categorizing them, but some also learn names for the categories, while others do not.

2. Method

The task required participants to learn to classify 16 "aliens" (Fig. 1) into two categories—those to be approached, and those to be avoided—by responding with the appropriate direction of motion. These behavioral categories were chosen specifically because they exemplify typical categories learned by non-human animals. Participants were randomly assigned to a *label* or *no-label* condition. After each response, auditory feedback consisting of a buzz or a bell indicated whether the response was correct. In addition to the feedback, participants in the *label* condition saw one of two nonsense words ("grecious" or "leebish" depending on the alien category) appear next to the alien. The word appeared only after a response was made and was unrelated to the response.

Participants in the *label* group learned that aliens that look a certain way should be avoided, *and* that they are *leebish*, while participants in the *no-label* group learned only that aliens looking a certain way should be avoided.

2.1. *Participants*

Forty-four Carnegie Mellon University undergraduates participated in the experiment for course credit. Data from 4 participants were excluded for failure to follow instructions. Data for the testing part of the task were available only for the last 20 participants.

2.2. *Materials*

The stimuli were a subset of the YUFO stimulus set (Gauthier, James, Curby, & Tarr, 2003). See Figure 1 for sample stimuli. The category distinction involved subtle differences in the configuration of the "head" and "body" of the creatures. The stimuli were presented on a 17" CRT screen. Responses were collected using a Gravis gamepad controller.

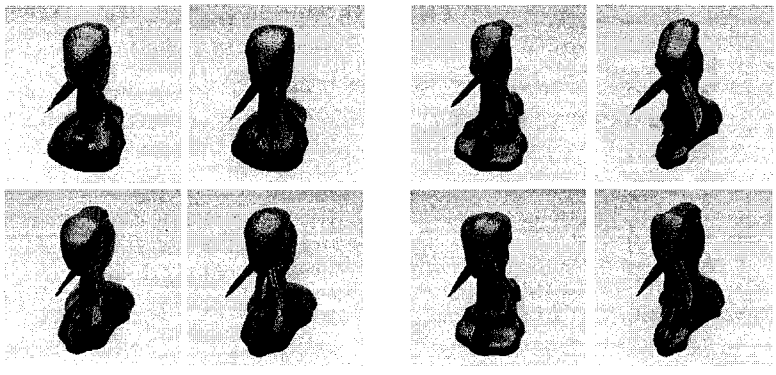


Fig 1. Sample stimuli.

2.3. *Training*

Participants were told that they were acting as an explorer on another planet learning about alien lifeforms. Their task was to figure out which aliens they should approach and which they should move away from. On each training trial, one of the 16 aliens appeared in the center of the screen. After a 500 msec delay, an outline of a character in a spacesuit (the "explorer") appeared in one of four quadrants—left, right, top, or bottom of the alien. Participants were instructed to respond with the appropriate direction key depending on the category of the

alien. For instance, if the explorer appeared to the top of the alien, and the participant thought the correct response was to move towards, she would press the 'down' key, which would move the explorer down—towards the alien. Auditory feedback—a buzz or a bell—sounded 200 ms after each response. In the *label* condition, a printed label then appeared next to the alien. After another 1500 ms, the stimulus was erased and a fixation cross marked the start of the next trial. The total duration of each trial was equal in the two conditions. The pairing of the label ("leebish" or "grecious") with the category (move away, move towards), and the perceptual stimuli (*family 1*, *family 2*), was counterbalanced across participants. There were 144 training trials. Note that all participants received the same number of categorization trials, the only difference being in the presence or absence of category labels accompanying each response.

2.4. *Testing*

After completing the training trials, participants were instructed that in the upcoming part, no feedback (or labels) would be provided. On each trial, one of the aliens appeared in the center of the screen and remained visible until a response was made. Participants were instructed to press one button if they thought the alien belonged to the *escape* category and another button for *approach*. There was no mention of the names learned in the *label* condition. To determine whether participants were responding based on category knowledge or memory for specific examples, the testing part included novel stimuli from the learned categories in addition to previously categorized stimuli.

3. Results

A full treatment of the results will be reported elsewhere. All tests reported as significant were significant at the $p < .05$ level. Participants in the label condition performed significantly better overall and learned the categories significantly faster than participants in the no-label condition. After completing the learning phase, participants' knowledge of the categories was tested without feedback or labels. Since no reinforcement or correction was provided in the testing phase, category knowledge may deteriorate over time. It was found that participants in both groups did well above chance on the generalization trials, so the reported data collapses across both types of testing stimuli, a total of 144 trials. Participants who learned the categories with labels retained their category knowledge throughout the testing phase, whereas those who learned the categories without labels displayed decreasing accuracy over time as revealed by

a significant block by condition interaction (Figure 1B). There was no difference in latencies between the two groups in any part of the experiment, suggesting that the advantage in the label condition did not arise from greater time spent on task.

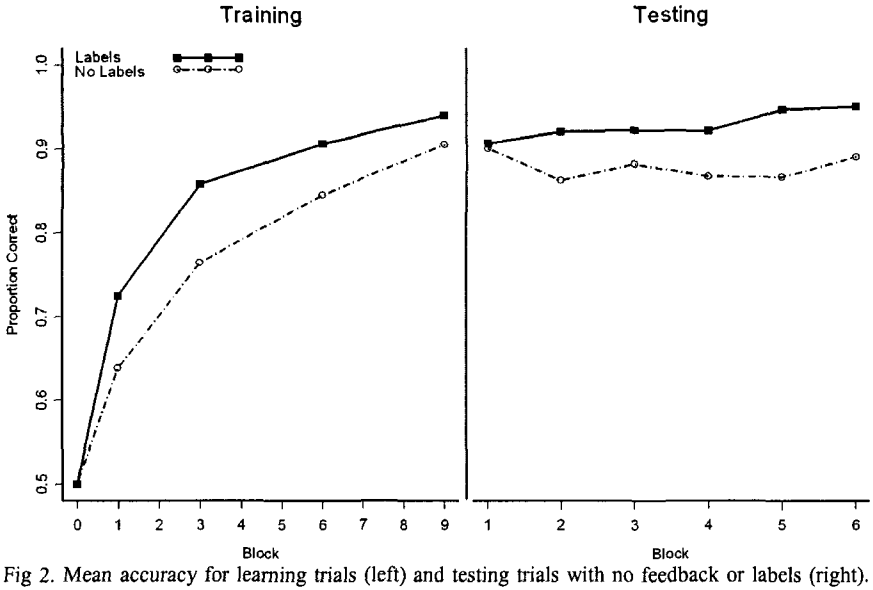


Fig 2. Mean accuracy for learning trials (left) and testing trials with no feedback or labels (right). Block 0 is included for illustration only and is not included in the analyses.

4. Discussion

Participants who learned names for novel categories showed faster learning of the categories compared to participants who categorized the items without learning names. Once the categories were learned by both groups, participants in the *label* group showed more robust category knowledge as evidenced by continued good performance when feedback and labels were removed.

The mounting evidence of effects of language on thought has been argued to be an artifact of cultural differences that limit exposure to certain types of cognitive tasks rather than arising from linguistic experience (Gelman & Butterworth, 2005; Li & Gleitman, 2002). The present task contrasted adults from similar backgrounds and controlled the exposure and categorization practice with the experimental stimuli. Despite undergoing the same amount of experience categorizing perceptual stimuli into behavioral categories (move away or move towards), the presence of labels helped adults to represent the stimuli in terms of the appropriate behavioral categories.

Humans are unique in their ability to habitually associate words with mental categories. Most notably, this ability allows for linguistic communication. The present results suggest that learning category labels also provides for facilitated representation of the labeled concepts. Words or other discrete modes of representation may be necessary for entertaining certain abstract concepts like large exact numerosities. However, learning to associate words even with more ordinary perceptually-based categories such as those used here, facilitates their acquisition and results in more robust subsequent knowledge.

The categories used in the current experiment were by no means incommensurate with linguistic coding. Indeed, many participants in the *no-label* condition reported self-generating labels during the learning task. It remains to be seen whether labeling can produce an even greater difference in learning unfamiliar categories for which the self-generation of labels is more difficult or impossible.

4.1. *Why did the labels help?*

The current results do not provide an unambiguous explanation of the mechanism involved. However, we can rule out several explanations. Labels did not encourage participants to work harder, at least as evidenced by a lack of differences in the reaction times of the two groups. Alternatively, it might be argued that the presence of the written labels directed the participants' attention to a certain part of the stimulus which made it easier to find the features relevant to the categorization task. It was found, however, that performance was identical whether the labels were presented in the visual (written) or purely auditory modality. We have preliminary evidence showing that mere exposure to the labels does not provide the benefit. They need to be learned. The minority of participants who did not learn the mapping between the labels and the stimuli, or indicated on a post-study questionnaire that the labels were not helpful were comparable in their performance to those in the *no-label* condition.

What was it about the labels then that facilitated category-learning? The current task required participants to learn a subtle and difficult-to-verbalize distinction based on experience with individual category exemplars. In this context, the labels can be thought to provide perceptually-simple correlates to an otherwise perceptually-complex task. The labels seem to allow participants to represent the category distinction in terms of the categorical distinction: "leebish" vs. "grecious," rather than the fuzzier perceptual distinction: "slightly more rounded and larger" vs. "less rounded and smaller."

4.2. *Are words special?*

Do labels need to be words or can any discrete information correlated with categories result in facilitated learning? A follow-up experiment was conducted that substituted word-labels with hieroglyphic symbols. Surprisingly, it was found that once the mapping between the symbols and categories was learned, they appeared to facilitate category-learning nearly as much as the verbal labels. It is not clear, however, whether the participants thought of the symbols as verbal labels. On the other hand, not all correlated information seems to have the facilitating effect. In another study, participants were told that the two groups of aliens lived in different places, and after making a response the alien would move towards one of these places (up or down). Learning to associate the motions with the aliens did not appear to facilitate category-learning. On the contrary, there is evidence that being exposed to this extra correlation actually made the task more difficult.

To summarize, we do not believe that only words can serve as discrete category markers. The labeling advantage may extend to other discrete environmental cues that are highly correlated with the categories. Words are special in that they normally constitute these discrete environmental cues, and so participants may find it more natural to use verbal labels as cues to categories compared to other types of information (e.g., lives above vs. lives below).

5. Conclusions

Providing redundant labels facilitated learning of novel categories and resulted in more stable category representations. Despite undergoing the same amount of experience categorizing perceptual stimuli into behavioral categories (*move away* or *move towards*), the presence of labels (nonsense words: "leebish"/"grecious") helped normal English-speaking adults to represent the stimuli in terms of the appropriate behavioral categories and resulted in more robust category representations. The current work shows that learning to name items facilitates placing them into appropriate behavioral categories, supporting Clark's (1997; 1998) view of words as "conceptual anchors" that facilitate negotiating perceptually-complex environments.

Acknowledgements

Thanks to Jay McClelland and David Rakison for useful discussion, Mike Tarr for kindly providing the YUFO stimuli, and Brian Mathias and Ashleigh Molz

for help with data collection. This work was supported in part by a NSF Graduate Fellowship to the author.

References

Clark, A. (1997). *Being There: Putting brain, body, and world together again.* Cambridge, MA: MIT Press.

Clark, A. (1998). Magic Words: How Language Augments Human Computation. In P.Carruthers & J. Boucher (Eds.), *Language and Thought: Interdisciplinary themes* (pp. 162-183). Cambridge University Press.

Clark, A. & Karmiloff-Smith, A. (1993). The Cognizer's Innards: A Psychological and Philosophical Perspective on the Development of Thought. *Mind & Language, 8,* 487-519.

De Renzi, E., Faglioni, P., Savoiardo, M., & Vignolo, L. A. (1966). The influence of aphasia and of the hemispheric side of the cerebral lesion on abstract thinking. *Cortex, 2,* 399-420.

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology, 20,* 507-523.

Gelman, R. & Butterworth, B. (2005). Number and language: how are they related? *Trends in Cognitive Sciences, 9,* 6-10.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306,* 496-499.

Hjelmquist, E. K. E. (1989). Concept-Formation in Non-Verbal Categorization Tasks in Brain-Damaged Patients with and Without Aphasia. *Scandinavian Journal of Psychology, 30,* 243-254.

Li, P. & Gleitman, L. (2002). Turning the tables: language and spatial reasoning. *Cognition, 83,* 265-294.

Pica, P., Lemer, C., Izard, W., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306,* 499-503.

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition, 71,* 1-42.

Rumelhart, D. E., Smolensky, D. E., McClelland, J. L., & Hinton, G. E. (1986). Parallel Distributed Processing Models of Schemata and Sequential Thought Processes. In J.L.McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing Vol II* (pp. 7-57). Cambridge, MA: MIT Press.

Vygotsky, L. S. (1962). *Thought and Language.* Cambridge, MA: MIT Press.

Waxman, S. R. & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29,* 257-302.

EMERGENCE OF COMMUNICATION IN TEAMS OF EMBODIED AND SITUATED AGENTS

DAVIDE MAROCCO STEFANO NOLFI

Institute of Cognitive Science and Technologies, CNR, Via San Martino della Battaglia 44, Rome, 00185, Italy

In this paper we will describe the results of an experiment in which an effective communication system arises among a collection of initially non-communicating agents through a self-organization process based on an evolutionary process. Evolved agents communicate by producing and detecting five different signals that affect both their motor and signaling behavior. These signals identify features of the environment and of the agents/agents and agents/environmental relations that are crucial for solving the given problem. The obtained results also indicate that individual and social/communicative behaviors are tightly co-adapted.

1. Introduction

The development of embodied agents able to interact autonomously with the physical world and to communicate on the basis of a self-organizing communication system is a new exciting field of research (Steels and Vogt, 1997; Cangelosi and Parisi, 1998; Steels, 1999; Marocco, Cangelosi and Nolfi, 2003; Quinn et al, 2003; for a review see Kirby, 2002; Steels, 2003; Wagner et al., 2003; Nolfi, in press). The objective is to identify methods of how a population of agents equipped with a sensory-motor system and a cognitive apparatus can develop a grounded communication system and use their communication abilities to solve a given problem. Such communication systems may have similar characteristics to animal communication or human language.

In this paper we will describe the results of an experiment in which an effective communication system arises among a collection of initially non-communicating agents through a self-organization process based on artificial evolution. Unlike other experimental setup in which the interaction between agents or the motor behavior of agents is pre-determined and fixed (e.g. Steels, 1999; Marocco, Cangelosi, and Nolfi, 2003) evolving agents have to autonomously determine: (a) their individual behavior (i.e. how they behave on the basis of their sensory information when signals produced by other agents cannot be detected), (b) their communicative behavior (i.e. when and how many signals are produced, the context in which signals are produced, the type and number of signals produced, the effect of signals detected on the individual motor and signaling behavior, the modalities with which agents communicate).

2. Experimental set-up

A team of four simulated robots that "live" in the same environment (i.e. a white arena of 270x270cm surrounded by white walls containing two grey target areas, Figure 1, left) are evolved for the ability to solve a collective navigation problem. Robots are provided with simple sensory-motor capabilities that allow them to move, produce signals with varying intensities, and to gather information from their physical and social environment (including signals produced by other agents). The control system of the robots is an artificial neural network.

The robots have a circular body with a radius of 11 cm and the robots' neural controllers consist of neural networks with 14 sensory neurons (that encode the activation states of the 8 infrared sensors that allow the robots to detect obstacles, 1 ground sensor that allow robots to detect the color of the floor, 4 communicative sensors that allow robots to detect the signals emitted by nearby robots, and 1 sensor that encode the activation state of the communication actuator at times t-1, i.e. each robot can hear its own emitted signal at the previous time step) directly connected to the three motor neurons that control the desired speed of the two wheels and the intensity of the communication signal produced by the robot. The neural controllers also include two hidden neurons that receive connection from the sensory neurons and from themselves and send connections to the motor and communicating neurons (Figure 1, right). The communication sensors can detect signals produced by other robots up to a distance of 100cm from four corresponding directions (i.e. frontal [315°-45°], rear [135°-225°], left [225°-315°], right [45°-135°]).

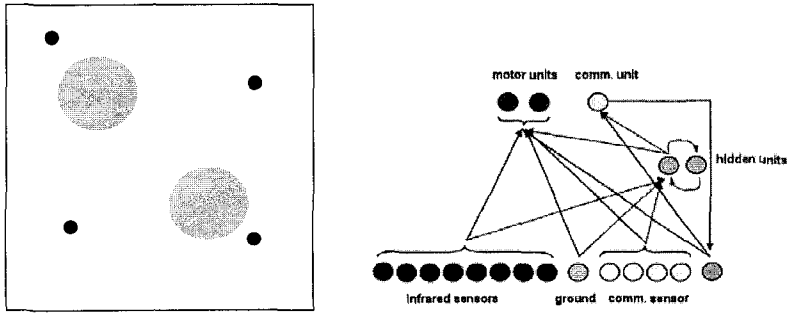


Figure 1. **Left**: The environment and the robots. The square represents the arena surrounded by walls. The two grey circles represent two target areas. The four black circles represent four robots. **Right**: The neural controller evolving robots. Internal neurons and recurrent connections are only included in one of the two experimental setting (see text).

Agents were evolved (Nolfi and Floreano, 2000) for the ability to find and remain in the target areas by subdividing themselves equally between the two areas. In particular, the fitness of the team of robots consists of the sum of 0.25 scores for each robot located in a target area and a score of -1.00 for each extra robot (i.e. each robot exceeding the maximum number of two) located in a target area. The total fitness of a team is computed by summing the fitness gathered by the four robots in each time step. The initial population consisted of 100 randomly generated genotypes that encoded the connection weights of 100 corresponding neural controllers. Each genotype is translated into 4 identical neural controllers that are embodied in the four corresponding robots. The evolutionary process lasted 100 generations. Each generation the 20 best genotypes were allowed to reproduce by generating five copies each, with 2% of their bits replaced with a new randomly selected value.

3. The evolved behaviour

By analyzing the behavior of one of the best team of evolved robots we can see that evolved robots are able to find and remain in the two target areas by equally dividing between the two. In the example shown in left side of the Figure 2, robots 2 and 3 quickly reach two different empty target areas. Later on, robot 1 and then robot 0 approach and enter in the bottom-right target area. As soon as the third robot (i.e. robot 0) enter in the area, robot 1 leaves the bottom-right target area and, after exploring the environment for a while, enters and remains in the top-left target area.

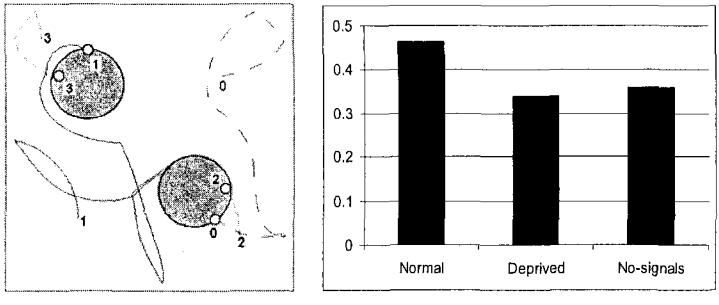


Figure 2. **Left**: The behavior displayed by the team of evolved robots of one of the best replications. The square and the gray circles indicate the arena and the target area respectively. Lines inside the arena indicate the trajectory of the four robots during a trial. The numbers indicate the starting and ending position of the corresponding robot (the ending position is marked with a white circle). **Right**: Average fitness of all teams of the last generations of 10 different replications of the experiment in a *Normal*, *Deprived*, and *No-signals* condition. In all cases, individuals have been tested for 1000 trials.

To determine whether the possibility to signal and to use other robots' signals is exploited by evolving robots we tested the evolved team in three conditions: a "Normal" condition, a "Deprived" condition in which robots evolved in a normal condition were tested in a control condition in which the state of communication sensors was always set to a null value, and a "No-signal" conditions in which robots were evolved and tested with their communication sensors always set to a null value (see Figure 2, right). The fact that performance in the "Normal" condition are better and statistically different ($p<0.001$) from the other two control conditions indicates that communication plays a role. Performance in the "Deprived" and "No-signals" conditions are not statistically different.

4. The communication system of evolved agents

By analyzing the communication system we observed that evolved agents produce different signals (as we said above, signals consists of single values ranging between [0,1]) and react to detected signals by modifying both their motor and signaling behavior.

For example, in one of the best replication of the experiment, evolved agents produce and use five different signals: a signal ***A*** with an intensity of about 0.42 produced by robots located outside the target areas not interacting with other robots located inside target areas; a signal ***B*** with an intensity of about 0.85 produced by robots located alone inside a target area; a signal ***C***, an oscillatory signal with an average intensity of 0.57, produced by robots located inside a target area that also contains another robot (i.e. when robots detect a signal produced by a robot also located in a target area); a signal ***D*** with an almost null intensity produced by robots outside target areas that are approaching a target area and are interacting with another robot located inside the target area; and a signal ***E***, an oscillatory signal with an average intensity of 0.33, emitted by robots located outside the target areas interacting with other robots also located outside target areas.

Detected signals affect the robots' motor and signaling behavior as follows: (1) robots located outside the target areas receiving signal ***E*** modify their motor trajectory so to reduce the time needed to reach a target area, on the average; (2) robots located outside target areas receiving signal ***B*** modify their motor behavior by approaching the robot emitting the signal (i.e. by approaching the target area in which the robot emitting the signal is located) and their signaling behavior (i.e. by producing signal ***D*** instead of signal ***A***); (3) robots located outside the target areas detecting the signal ***C*** modify their motor behavior so as

to tend to move away from the signal source; (4) robots located inside the target areas detecting the signal ***C*** modify their motor behavior so to increase their likeness to exit from the target area, (5) robots located outside the target areas detecting the signal ***A*** modify their signaling behavior by producing signal ***E*** instead of signal ***A***. The functionality of signals have been identified and demonstrated through experimental tests that we do not report in this paper for space reasons.

5. Relation between individual and social/communicative behavior

Since robots individual and social/communicative behavior are allowed to co-evolve we might wonder what the relation between these two forms of behaviors is and how the possibility to co-adapt them is exploited by evolved individuals.

The fact that agents tested in a condition in which signals produced by other agents cannot be detected (Figure 2, right, "Deprived" condition) perform similar to agents evolved and tested in this condition (Figure 2, right, "No-Signal" condition) indicates that evolved agents tend to optimize both their individual and social/communicative behavior. The adaptive pressure toward the development of an effective individual behavior can be explained by considering that signals produced by other agents are not always available since the signals that are produced and detected depends on the current position of the other agents that is partially unpredictable since agents start from randomly initialized positions and orientations.

Indeed, the analysis of the individual behavior (i.e. the behavior of agents that are not allowed to detect other agents signals) exhibited by evolved agents indicates that they are able to solve the navigation problem to a good extent. Indeed, by avoiding walls, by exhibiting curvilinear trajectories when far from walls, and by remaining in the target areas as soon as they enter into one of them, evolved agents are able to find and remain in the two target areas most of the times even in a "Deprived" condition. Figure 3 shows how evolved agents tested in a "Deprived" condition spend about 60% of their lifetime in the conditions in which the team gathers a positive fitness.

Communication is used by evolved agents as an additional mechanism, with respect to their individual capabilities, that allow them to correct mistakes produced by their individual behaviors (e.g. to exit from target area that contains more than two agents) and to improve some of their abilities that are accomplished through their individual behavior (e.g. by reducing the time required to reach target areas, on the average or the ability to directly move

toward a target area that contains a single agent by exploiting the signal produced by the agent already located in the target area). These improvements are reflected in the data shown in Figure 3 that, for example, indicates that agents spend much less time in target areas that already contains two or three other robots in a "Normal" rather than in a "Deprived" condition.

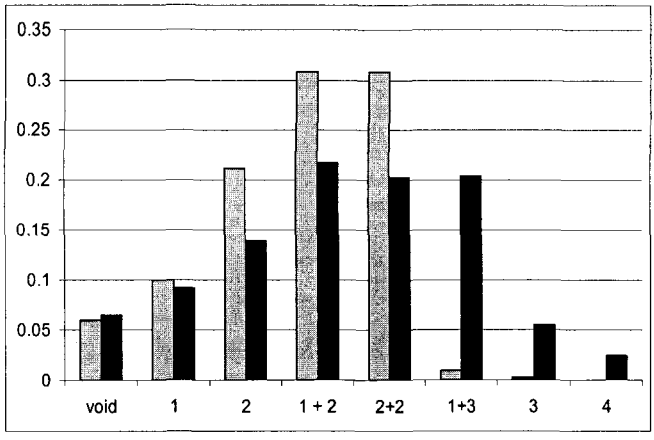


Figure 3. Percentage of lifecycles spent by a team of four agents (of the best evolved team) in the 8 possible different states tested in a "Normal" condition (gray bars) and in a "Deprived" condition in which agents are not allowed to detect other agents' signals. "Void" indicate the case in which all the four agents are located outside target areas (fitness = 0.0). "1" indicates the case in which only a single agent is located in a target area (fitness = 0.25). "2" indicates the case in which two agents are located in target areas (fitness = 0.5). "1+2" indicates the case in which one agent is located in a target area and two other agents are located in the other target area (fitness = 0.75). "2+2" indicates the case in which each of the two target area contains two agents (fitness = 1.0). "1+3" the case in which one target area contains one agent and the other three agents (fitness = 0.0). "3" indicates the case in which three agents are located in the same target area (fitness = -0.25). "4" indicates the case in which four agents are located in the same target area (fitness = -1.0). Average performance obtained by testing the agents for 1000 trials lasting 1000 cycles.

6. Discussion

In this paper we described the results of an experiment in which an effective communication system arises among a collection of initially non-communicating agents evolved for the ability to solve a collective navigation problem. With the methodology chosen, we observed that agents developed autonomously (i.e. without human intervention), first of all, an effective individual behavior that allow agents to cope the navigation problems without the collaboration of the other agents allowed by communication.

In addition, agents developed an effective communication system based on five different signals that correspond to crucial features of the environment, of the agents/agents relations, and agents/environmental relations (e.g. the relative location of a target area, the number of agents contained in a target area, etc.). These features, that have been autonomously discovered by the agents and that are grounded in agents' sensory-motor experiences, constitute the 'meanings' of the signals produced and detected by the agents. Used signals, therefore, do not only refer to the characteristics of the physical environment but also to those of the social environment constituted by the other agents and by their current state.

The analysis of the obtained results also indicate that individual and social/communicative behaviors are tightly co-adapted. In fact, since individual behavior in evolved agents are optimized as well as social/communicative behavior, detected signals act as a sort of additional mechanism that enhances individual behavior (when signals are available). On the other hand, individual behavior, in absence of useful signals, guarantees the maximum performance that can be achieved on the basis of the available sensory information.

The individual behavior also creates the basis for the exploitation of signaling capabilities. For example, the individual ability to reach and remain in a target area represents a pre-condition for the emergence of an ability to signal the relative position of a target area and to use that signal appropriately. Similarly the limits of the individual behavior, for instance the tendency to enter into a target area that already contains two other agents, represents a pre-condition for the development of communication abilities that allow agents to exit from target areas that contains more than two agents.

Interestingly, one can find interesting similarities between the communication systems observed in our experiments and forms of animal communication described in the literature. For instance, signals that refer to agent/environment interactions are similar to alarm calls or food calls in birds and primates that provide information about objects or events that are external to the animal that emits the signal (Hauser, 1996). Moreover, the coordinated oscillatory signals produced by two robots located in the same target area (that allow the robots to keep additional robots away while maintaining the couple of robots in the area) are similar to the synchronized communicative interactions known as vocal duetting produced by several animals. Indeed, as in the case of the robots described in this paper, in some birds duets play an important cooperative function since they allow the members of a couple of animals to defend their territory and/or to keep the pair bond (Langmore, 1998; Slater, 1997).

Acknowledgments

The research has been supported by the ECAGENTS project founded by the Future and Emerging Technologies programme (IST-FET) of the European Community under EU R&D contract IST-2003-1940.

References

Cangelosi, A., & Parisi, D. (1998) The emergence of a 'language' in an evolving population of neural networks. *Connection Science*, 10: 83-97.

Hauser, M. D. (1996) *The evolution of communication*, Cambridge MA: MIT Press.

Kirby, S. (2002). Natural Language from Artificial Life. *Artificial Life*, 8(2):185--215.

Langmore, N. E. (1998) Functions of duets and solo songs of female birds, *Trends in Ecology and Evolution*,13: 136-140.

Marocco, D., Cangelosi, A., & Nolfi, S. (2003), The emergence of communication in evolutionary robots. *Philosophical Transactions of the Royal Society London - A*, 361: 2397-2421.

Nolfi, S. (in press). Emergence of Communication in Embodied Agents: Co-Adapting Communicative and Non-Communicative Behaviours. *Connection Science.*

Nolfi, S., & Floreano, D. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines.* Cambridge, MA: MIT Press/Bradford Books.

Quinn, M., Smith, L., Mayley, G., & Husbands, P. (2003) Evolving controllers for a homogeneous system of physical robots: Structured cooperation with minimal sensors. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 361:2321-2344.

Slater, P. J. B. (1997) Singing in the rain forest: the duets of bay wrens, *Trends in Ecology and Evolution*, 12: 207-208.

Steels, L. (1999). *The Talking Heads Experiment*, Antwerpen, Laboratorium. Limited Pre-edition.

Steels, L. (2003) Evolving grounded communication for robots. *Trends in Cognitive Science.* 7(7): 308-312.

Steels, L., & Vogt, P. (1997) Grounding adaptive language games in robotic agents. In: P. Husband & I. Harvey (Eds.), *Proceedings of the 4th European Conference on Artificial Life*, p. 474-482. Cambridge MA: MIT Press.

Wagner, K., Reggia, J.A., Uriagereka J., & Wilkinson, G.S. (2003). Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1):37-69.

A LANGUAGE EMERGENCE MODEL PREDICTS WORD ORDER BIAS

JAMES W. MINETT, TAO GONG & WILLIAM S-Y. WANG

Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong, China

The majority of extant languages have one of three basic word orders: SVO, SOV or VSO. Various hypotheses have been put forward to explain aspects of this bias, including the existence of a universal grammar, learnability imposed by non-linguistic-specific cognitive constraints, and the descent of the extant languages from a common ancestral proto-language. Here, we adopt a multi-agent model for language emergence that simulates the coevolution of a lexicon and syntax from a holistic signaling system. The syntax evolves through a process of categorization; local syntactic rules are constructed that assign a relative order (e.g., S before V) to the elements of the two categories to which each rule applies. We demonstrate that local syntax encoding the relative position of S and O are the most stable, allowing the coexistence of the global word order pairs SOV/SVO and VOS/OVS. The structure of the semantic space that the language encodes further constrains the global syntax that is stable.

1. Introduction

Declarative sentences involving a verb (V), a subject (S) and an object (O) have one of six logically possible word orders. Greenberg (1963) observed that only three of these basic word orders are common among the world's extant languages: SVO, SOV and VSO. A number of hypotheses have been put forward to explain this bias. Some researchers (e.g., Briscoe, 2000) have argued that the acquisition of word order is influenced by the parameters of a built-in universal grammar. Others emphasize the role of non-linguistic-specific cognitive constraints to learning word order (e.g., Lupyan & Christiansen, 2002). Recently, Gell-Mann and Ruhlen (2005) have enumerated the numbers of languages belonging to each of the world's extant language families (as classified by Ruhlen, 1991) that have the orders SVO, SOV and VSO. From these data, they have concluded that "if there was a language from which all modern languages derive it had the word order SOV". The general trend that they have observed in basic word order is for the gradual fronting of the verb from sentence final position to sentence initial position while maintaining the relative position of the subject before the object.

Here we investigate bias in basic word order by studying the behavior of a new simulation model for language emergence proposed by Gong et al. (2006) in this volume. In this model, word order is encoded locally by assigning a relative

order (e.g., S before V) to lexical items belonging to pairs of categories that emerge gradually during the simulation. Our aim is to examine the degree to which bias in basic word order can be explained by self-organization resulting from competition among such local syntax rules and the structure of the semantic space that can be cognized by language users.

2. The Emergence of Word Order

Gong et al. (2004, 2005) have developed a model for the phylogenetic emergence of language to show that a population of interacting language-capable individuals can acquire a common lexicon and word order as a result of a simple local learning algorithm that is based on the detection of recurrent patterns. However, the model suffers from the shortcoming that the semantic structures that the individuals can manipulate map directly to corresponding syntactic structures. The authors have now extended the model (Gong et al., 2006, in this volume) to model the emergence of the syntax itself. Speakers acquire lexical items by detecting recurrent patterns in both the meaning and the form of perceived utterances, just as in the previous model. However, in this new framework, the syntax evolves through a process of categorization: Lexical items are assigned to the same category when an individual observes that those items have the same relative order with respect to some other item. For example, an individual who observes that two predicates "eat" and "drink" both come before—but not necessarily immediately before—the agent "cat" may create a category containing both predicates. The knowledge that these predicates precede the agent "cat" (or a category containing it) is then stored as a *local syntax rule*. As the categorization process continues, so increasingly many lexical rules become assigned to categories, so allowing an increasing proportion of the language to become compositional under the regulation of local syntax. Eventually, the emergent syntax may evolve to the point that all lexical items in an utterance are consistent with a particular *global syntax* (by which we mean the syntax of an entire utterance, e.g., SOV) that results from the combination of several local syntax rules. Thus global syntax is viewed as merely an emergent property of the local syntax rules that emerge. For further details of the model, refer to the paper by Gong et al. in this volume.

Our aim here is to investigate the bias in word order that is predicted to emerge in a language whose syntax is encoded locally. We begin by describing the relative persistence of the word orders that can emerge. The semantic space of the model comprises a set of integrated meanings of two types: Type-I integrated meanings, which consist of a predicate that takes one argument, its

agent, and Type-II integrated meanings, which consist of a predicate that takes two arguments, an agent and a patient.

For utterances that encode Type-I integrated meanings, there are only two possible local word orders, SV and VS. If sufficiently many individuals acquire a preference for one particular order, then that order will tend to diffuse across the entire population, resulting in the emergence of a shared word order for Type-I integrated meanings. Thus a one-to-one mapping between the semantic structure and syntactic structure will have emerged.

For utterances that encode Type-II integrated meanings, however, the situation is more complex. There are three pairs of competing local word orders: SV and VS (as for Type-I utterances), SO and OS, and VO and OV. The global syntax of Type-II utterances emerges in several steps as each local syntax rule is formed. Let us assume, for example, that the SV/VS competition completes first. What resultant bias in the global syntax for Type-II utterances should be expected to emerge? Suppose, for the sake of argument, that it is SV that wins the competition with VS. This local syntax is invoked only for utterances having surface forms that are consistent with one of three global word orders: OSV, SOV and SVO. Two of these word orders are consistent with SO; only one is consistent with OS. Therefore, if individuals perceive utterances having each of these global orders with equal probability, SO will be preferred to OS. If SO is indeed the second local syntax rule to emerge, the global syntax will tend to fluctuate between SOV and SVO, both of which are consistent with the local syntax SV and SO. Similarly, emergence of OV will lead to fluctuation of the global syntax between OSV and SOV. We refer to such syntax as *imprecise*.

If, however, OS emerges second, the two local syntax rules, SV and OS, give rise to a single emergent global syntax: OSV. We call such a syntax *precise*. One might expect this syntax to persist due to the subsequent competition between VO and OV leading to VO being reduced in strength for being inconsistent with the global word order OSV, thereby reinforcing the global order OSV. However, SOV and SVO are more likely to persist in practice because of the greater expected frequency of utterances having surface form consistent with SO. Other combinations of local syntax rules influence global syntax similarly.

To summarize, combinations of local syntax rules that generate an imprecise syntax tend to persist, while those that generate a precise syntax tend not to persist.

3. The Observed Bias in Global Word Order

The previous section explains the bias in the persistence of global word order that we predict to result from different combinations of emergence of local syntax. So far, we have neglected the impact of the structure of the semantic space—for example, Type-I sentences and Type-II sentences might not be produced with equal frequency. In order to explore the validity of the predictions above, we have run the model for various sets of parameter values that control the structure of the semantic space and the initial language of the population.

The structure of the semantic space depends on the relative proportions of agents, patients and predicates that make up the integrated meanings that can be expressed. In the experiments that follow we generate the semantic space of integrated meanings from 4 agents/patients, 4 predicates taking one argument and 4 predicates taking two arguments. As a result, the semantic space consists of 16 Type-I integrated meanings and 48 Type-II integrated meanings. In each experiment, a population of 10 individuals is initialized with a shared compositional language comprising a complete set of lexical items and two local syntax rules by which the lexicon is regulated. In some experiments, the initial local syntax is precise, generating a single global word order. In other experiments, the initial local syntax is imprecise, generating two, competing global word orders. Each simulation is run for 5,000 communications. Each experiment is repeated 20 times.

For our first experiment, we set a three-fold token bias in favor of Type-I sentences. This models a situation in which individuals express the two sentence types with equal probability. In our first experiment, each individual in the population is initialized with a common compositional language having the precise local syntax SV+VO. The evolution of the syntax for one run is summarized in Figure 1: the top row of 3 panels shows the evolution of the local syntax; the second and third rows show the resultant evolution of the global syntax. Panels show the average proportions of integrated meanings for which individuals invoke the corresponding syntax in order to comprehend utterances.

The initial syntax SV+VO quickly evolves, VO falling in strength to be replaced by SO. As a result, the global syntax SOV starts to complete with the initial global syntax SVO. This syntax remains stable for about 1,200 communications at which time the competition between VO and OV shifts in favor of OV. This shift triggers SO to flip to OS, which in turn triggers SV to shift to VS. While this reorganization is ongoing, the global syntax OSV briefly invades the language. The effect of this reorganization of the syntax, however, is to shift the global syntax to OVS/VOS. Thus, two abrupt, synchronized

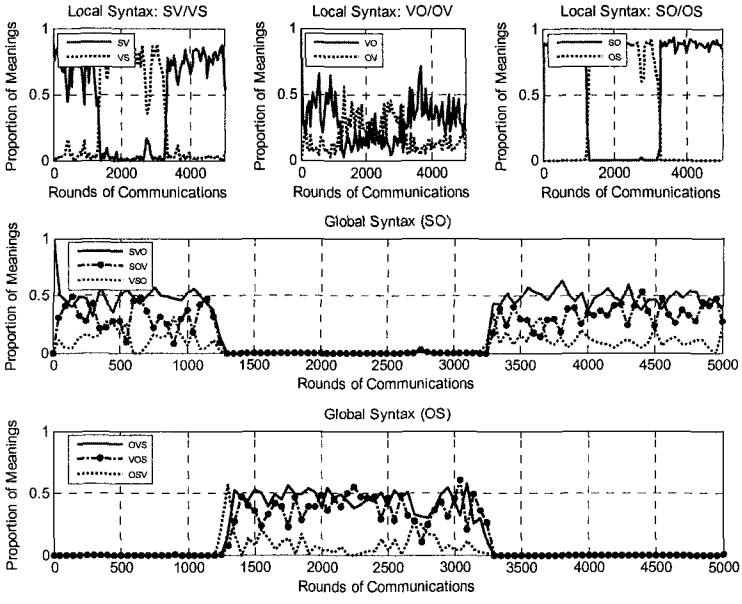


Figure 1. The evolution of syntax for an initial SV+VO compositional language.

transitions in the local syntax have brought about an inversion of the global syntax. For the next ~1,800 communications, the local syntax OS+VS remains stable, producing the two global word orders OVS and VOS. The system then undergoes another reorganization back to the original state of SO+SV.

The behavior just described is typical, not only of runs in which the initial syntax is SV+VO, but also for all other combinations of local syntax that we have examined: SV+SO, SV+OV, SO+OV, VS+SO, VS+OS. We conclude that, for a semantic space with a three-fold bias toward Type-I sentences, the system has only 4 qualitative states, as shown in Table 1.

In almost all runs, SO/OS is the most stable pair of local syntax rules. To understand why this should be the case, consider a syntax in which the relative order of S and O is not specified, e.g., SV+OV→SOV/OSV. When an individual hears an utterance encoded using this syntax, he is unable to use this syntax to distinguish the agent of the sentence from the patient. This combination of local syntax rules is therefore selected against, existing only as a transient state.

Table 1. Stable states of the local syntax and the resultant global syntax (Type-I : Type-II bias 3:1). The most frequently occurring syntaxes are also the most stable. The syntaxes marked as TRANSIENT occur only as transient states when the language reorganizes from one stable syntax to another. Other transient states occur only rarely, e.g. SO+VO→SVO/VSO. Precise grammars exist only as transient states. The stable syntaxes are all imprecise, as predicted in Section 2.

Local Syntax	Global Syntax	Stability
SO + SV	SOV and SVO	STABLE
OS + VS	VOS and OVS	STABLE
SO + VS	VSO	TRANSIENT
OS + SV	OSV	TRANSIENT

Some combinations of SV/VS and OV/VO, however, are not subject to this form of instability, e.g., the syntax SV+VO. This syntax is precise, specifying the unique global word order SVO, hence no such ambiguity arises. Nevertheless, because it is a precise syntax, it too exists only as a transient state. Precise syntaxes are quickly replaced because a third local syntax rule—SO in this case—can invade the language with no immediate reduction in communicative success. One of the two preexisting rules soon becomes redundant, causing the syntax to become imprecise (e.g., SV+SO). Change back to the precise syntax is then difficult because the imprecise syntax is now consistent with two global word orders, only one of which can be supported by the previous precise syntax. This same reasoning also explains the transient nature of the precise syntaxes SO+VS and OS+SV.

One further phenomenon of the evolution of syntax requires explanation. Why are the imprecise syntaxes SO+VO and OS+OV not also stable? To understand this we must consider the relative frequency of Type-I and Type-II sentences. Because of the three-fold bias in favor of Type-I sentences, which involve only S and V, the pair of local syntax rules SV and VS are reinforced frequently. They therefore stabilize before OV and VO, preventing the SO+VO and OS+OV syntaxes from occurring.

In a separate experiment, we set the Type-I : Type-II bias to 1:1. Individuals therefore make more frequent use (×3) of Type-II sentences than of Type-I sentences, reflecting a preference for more complex structures integrated meanings. This leads to the Type-II-specific local syntax rules being reinforced more quickly. As a result, there is no bias away from acquiring the OV and VO rules. The dominance of SO/OS, however, is maintained. Table 2 summarizes the stability of the syntax for this case.

Table 2. Stable states of the local syntax and the resultant global syntax (Type-I : Type-II bias 1:1).

Local Syntax	Global Syntax	Stability
SO + SV	SOV and SVO	STABLE
SO + VO	VSO and SVO	STABLE
OS + VS	VOS and OVS	STABLE
OS + OV	OSV and OVS	STABLE
SO + VS	VSO	TRANSIENT
SO + OV	SOV	TRANSIENT
OS + SV	OSV	TRANSIENT
OS + VO	VOS	TRANSIENT

In a final experiment, we set the Type-I : Type-II bias to 9:1 so that Type-I utterances were three time more frequent than Type-II sentences, reflecting a preference for simply structured integrated meanings. This semantic space generates the most stable syntaxes of all the cases that we have examined. In runs in which the initial compositional language is SV+SO, for example, the initial syntax is never replaced. The large preponderance of Type-I sentences means that the relative order of S and V are reinforced very often, preventing any new syntax from invading the language. We expect such stability to extend to other initial compositional languages.

4. Conclusion

The simulation results presented here demonstrate that bias in the word order of simple utterances is constrained primarily by the need to distinguish the agent and the patient, although the structure of the semantic space that the language encodes also has some effect. For all initial compositional languages that we have studied, the competition between SO and OS drives the evolution of global syntax. When SO is dominant, the local syntax SV tends to emerge also, generating the global syntaxes SOV and SVO. All other combinations of local syntax are transient. When OS is dominant, the prevailing global word orders are reversed: VOS and OVS tend to co-exist.

Empirical observations of the relative frequencies of each word order among the extant languages indicate that SOV and SVO are by far the most common word orders, with VSO being the next most frequent (Gell-Mann & Ruhlen, 2005). Our results are largely consistent with this empirical finding, although our

model predicts that OS+VS, generating the global syntax VOS and OVS is also common.

Acknowledgements

This work has been supported in part by grants 1224/02H and 1127/04H awarded by the Research Grants Council of Hong Kong to The Chinese University of Hong Kong.

References

Briscoe, E. J. (2000). Grammatical acquisition: inductive bias and coevolution of language and the language acquisition device. *Language, 76(2)*, 245–296.

Gell-Mann, M., & Ruhlen, M. (2005). The origin and evolution of word order. Unpublished manuscript.

Gong, T., Ke, J-Y., Minett, J. W., & Wang, W. S-Y. (2004). Language emergence: a self-organized model using indirect meaning transference. Paper presented at the Fifth International Conference on the Evolution of Language, Leipzig, Germany, March/April 2004.

Gong, T., Minett, J. W., Ke, J-Y., Holland, J. H., & Wang, W. S-Y. (2005). A computational model of the coevolution of lexicon and syntax. *Complexity, 10(6)*, 50–62.

Gong, T., Minett, J. W., & Wang, W. S-Y. (2006). Computational simulation on the coevolution of compositionality and regularity. Paper submitted to the Sixth International Conference on the Evolution of Language, Rome, Italy, April 2006.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. Universals of language. Greenberg, J. H. (ed.). Cambridge, MA.: MIT Press.

Lupyan, G., & Christiansen, M. H. (2002) Case, word order, and language learnability: insights from connectionist modeling. *Proceedings of 24th Annual Conference of the Cognitive Science Society* (pp. 596–601). Mahwah, NJ: Lawrence Erlbaum.

Ruhlen, M. (1991) *A Guide to the World's Languages, Vol. 1.* Stanford, CA: Stanford University Press.

TALKING TO ONESELF AS A SELECTIVE PRESSURE FOR THE EMERGENCE OF LANGUAGE

MARCO MIROLLI[1,2] DOMENICO PARISI[1]

[1] *Institute of Cognitive Sciences and Technologies, CNR, 44 Via San Martino della Battaglia, Roma 00185, Italy*

[2] *Philosophy and Social Sciences Department, University of Siena, 47 Via Roma, 53100, Siena, Italy*

Selective pressures for the evolutionary emergence of human language tend to be interpreted as social in nature, i.e., for better social communication and coordination. Using a simple neural network model of language acquisition we demonstrate that even using language for oneself, i.e., as private or inner speech, improves an individual's categorization of the world and, therefore, makes the individual's behavior more adaptive. We conclude that language may have first emerged due to the advantages it confers on individual cognition, and not only for its social advantages.

1. Introduction

1.1. *Talking-to-oneself and language evolution*

Human language has evolved. If we go back sufficiently in time we find ancestors of present-day humans who did not have language and then, some time later, ancestors who did have language. What is less clear is how and why language has evolved. If any capacity has evolved one can ask what adaptive advantages the possession of that capacity conferred to the individuals that possessed the capacity over other individuals that lacked it, even if this should not necessarily imply a pan-adaptivist view of evolution.

In the case of language it seems clear that language may have evolved because it conferred social advantages by dramatically improving the communicative capabilities of human beings which in turn improved coordination among individuals. The idea that language is just a very complex and powerful communication system can indeed be considered as the 'standard view' in the debate on language evolution (see, for example, Knight et al., 2000; Hurford, 2002; Pinker & Jackendoff, 2005).

Another less often explored possibility is that human language may have evolved because it made the cognitive functioning of single individuals more effective. Unlike animal communication systems human language can be used not only socially but also individually. It can be used to communicate with other individuals, by asking them information or providing them with information or

by asking them to do one thing or another, but it can also be used to talk with oneself, to comment on what one sees, to put one's predictions, explanatory hypotheses, and plans into words. These individual uses of language can result in more effective behaviours on the part of the individual even when the individual is acting alone, and this may have represented a reproductive advantage and a selective pressure for the emergence of language.

The idea that the main function of language is communication has been criticized by Bickerton (1990) and Chomsky (2002), who argue, from different standpoints, that language is fundamentally a system of mental representation and thought. Nevertheless, in the literature on language evolution the individual functions of language tend either to be ignored or to be thought to have appeared much more recently than it would be required if these uses acted as a selective pressure for the initial emergence of human language. It is assumed that human beings have learned to talk to themselves when their language was already completely developed and indistinguishable from the language spoken by present-day humans. The initial selective pressures for the emergence of language were social. When language was already evolved and fully modern, humans found that it could be usefully used to talk to oneself and not only to talk to others. But it is not clear that this is necessarily so. It is possible that even a very simple form of proto-language, consisting of words (or holophrases) that correlate with relevant experiences, can give important individual advantages if it is used not only for communication but also for talking to oneself. In this paper we use artificial neural network simulations for exploring the influence that a very simple form of proto-language can have on the categorization abilities of an individual when the language is used to talk to oneself, in private speech or in inner speech.

1.2. *Related work*

The idea that language is not only a system of communication but also a very powerful aid to cognition dates back at least to the 1930s, with the work of Russian psychologist Lev Vygotsky. Vygotsky (1978) considered language as a cognitive tool, whose development in the child and whose evolution in the species is the prerequisite to the development of all specifically human high-level psychological functions. Despite the fact that this idea has been somehow neglected in mainstream psychology, recently it has been raising increasing interest in the cognitive sciences (Gentner & Goldwin-Meadow, 2003) and is at the core of the interesting speculations on the evolution of human mind of such philosophers as Daniel Dennett (1991) and Andy Clark (2005). Indeed, there is

a growing body of empirical evidence demonstrating the importance of language for a number of cognitive functions including, for example, learning (Nazzi & Gopnik, 2001), memory (Gruber & Goschke, 2004), analogy making (Gentner, 2003) and problem solving (Diaz & Berk, 1992).

From the computational modeling point of view, there have been a few attempts to study the possible individual advantages that language can confer to those who possess it. Lupyan (2005) has demonstrated, with artificial neural networks, that learning to name perceived stimuli can facilitate category learning, especially the learning of those categories which are more difficult to learn. Using an artificial life framework, Cangelosi and Harnad (2000) have studied the adaptive advantage of what they call 'symbolic theft', that is the learning of new categories by hearsay from combinations of labels describing them. The results of their simulations show that organisms which employ symbolic theft outperform those which learn categories only by 'sensory-motor toil', that is, through trial-and-error and direct experience with objects. In previous work (Mirolli & Parisi, 2005b), we have shown that the use of a simple signalling system not only for communicative purposes but also for talking to oneself can facilitate the evolutionary emergence of the communication system itself. In particular, we have demonstrated that linguistic signals that benefit the hearer but not the speaker do not evolve if they are only used for communication while they do evolve if the hearer has to repeat the signals to himself or herself as an aid to memory. Finally, in a previous model on which the present work is based (Mirolli & Parisi, 2005a), we have shown that the coupling between the linguistic and the sensory-motor systems of an organism can improve categorization even after learning has taken place. In fact, we have shown that the internal representations of objects are improved if the objects are perceived together with the labels that designate them. In the previous model the linguistic input was always social, coming from another individual. In the present work we develop the model by studying the effects of language on categorization when the individual is all alone and he or she talks to herself either externally, with private speech, or internally, with inner speech.

2. A simple neural network model of language acquisition

We assume that we can divide this process of language acquisition in two stages. In the first stage the learner's brain functions as two functionally separated modules, the sensory-motor module and the linguistic module, which are modeled as two independent feed-forward three-layer neural networks (see figure 1a). In the sensory-motor module the input units encode the sensory

properties of perceived objects that can belong to four different categories[a]. In the first, pre-linguistic stage the sensory-motor module learns (by a standard back-propagation learning algorithm) to classify perceived objects by producing in the output units a specific activation pattern which represents the action appropriate to the category of the object. On the other hand, the input units of the linguistic module encode the acoustic properties of heard linguistic sounds and the linguistic module learns (by back-propagation) to imitate heard sounds, that is, to produce in its output units the same activation pattern that has been received in the input units[b]. Overall, the first stage of learning is meant to model the first year of life of the child, when the child acquires both non-linguistic sensory-motor mappings and the ability to imitate heard sounds but there still is no learning of the mapping between linguistic sounds and their meanings.

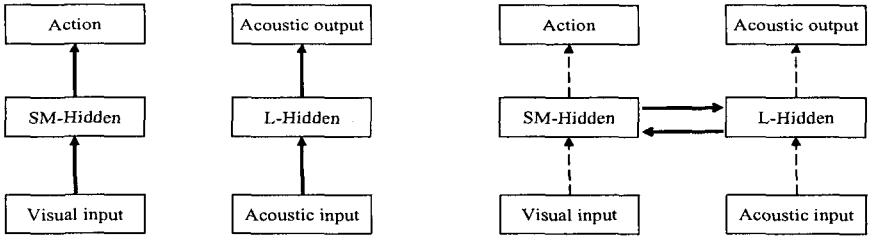


Figure 1. The neural network. During the first stage of learning (a) the two modules learn their respective tasks independently of one another. During the second stage of learning (b) new connections linking the two modules appear or become functional and their synaptic weights are progressively adjusted so that the overall network learns to map objects into the linguistic sounds that designate them (speaking) and linguistic sounds into actions (understanding).

In the second stage, in which language learning in the proper sense begins, the two modules become functionally connected by the development of two new sets of synaptic connections, one linking the hidden layer of the sensory-motor module to the hidden layer of the linguistic module and the other linking the hidden layer of the linguistic module to that of the sensory-motor module (figure 1b). The second stage of learning, which is meant to model language learning in the proper sense, consists in adjusting the weights of these two new sets of connections (through a delta-rule learning algorithm) in such a way that (a) the internal representation evoked in the sensory-motor hidden units by a

[a] There are 120 different objects for each category, created by changing the category's prototypical perceptual pattern.

[b] For symmetrical reasons, there are also 120 real instantiations of the same words, each differing slightly from the word's prototype.

perceived object will tend to produce the internal representation evoked in the linguistic hidden units by the linguistic sound normally co-occuring with the object in the learner's experience, and vice versa, (b) the internal representation of a heard sound in the linguistic hidden units will tend to evoke its 'meaning', that is, the internal representation of the object normally experienced together with the sound. This results in the ability to speak by mapping non-linguistic inputs into the appropriate phono-articulatory movements (e.g., saying "apple" in response to the sight of an apple) and in the ability to understand language by mapping heard linguistic sounds into the appropriate non-linguistic actions (e.g., approaching the speaker in response to the linguistic sound "come here!").

3. How language affects categorization

In order to study how language affects categorization, we analyzed the internal representations of perceived objects in five different conditions, which we call (a) no-learning, (b) no-language, (c) social language, (d) private speech, and (e) inner speech. In our model, the internal representation of a perceived object is the activation pattern that is evoked in the sensory-motor hidden units by the perceived object. Activation patterns in a set of units can be represented as points in an abstract space with the same number of dimensions as the number of units in the corresponding layer of units and with each dimension of the space representing the level of activation of the corresponding unit, say, from 0 to 1. A particular internal activation pattern will be represented by a specific point located in the appropriate position with respect of each dimension, reflecting the activation level of the corresponding unit in the pattern. All the different members of a specific category of objects, say, all different apples, will be represented as a cloud of points. At the beginning of learning in the first stage, since all the connection weights are random, the cloud of points representing the internal activation patterns evoked by the different apples will be very large and it will largely overlap with the cloud of points representing other categories of objects. This is what is actually observed (see figure 2a). At the end of learning, the internal representations of objects belonging to the same category will form much better clouds, that is, clouds which are much smaller, do not overlap with other clouds, and have a considerable distance between their respective centers (figure 2b). Notice that the clouds in both the no-learning and no-language conditions are calculated by giving to the network only the perceptual properties of the object (figure 3a).

In order to test whether social language affects the internal representations of objects we calculated our clouds (categories) by making the network perceive

objects accompanied by instances of the linguistic sounds that normally accompany the objects in the learner's experience (figure 3b). The results, shown in figure 2c, demonstrate that language improves categorization in that the clouds of different categories of objects are now both smaller and more distant from one another than the clouds of objects unaccompanied by language.

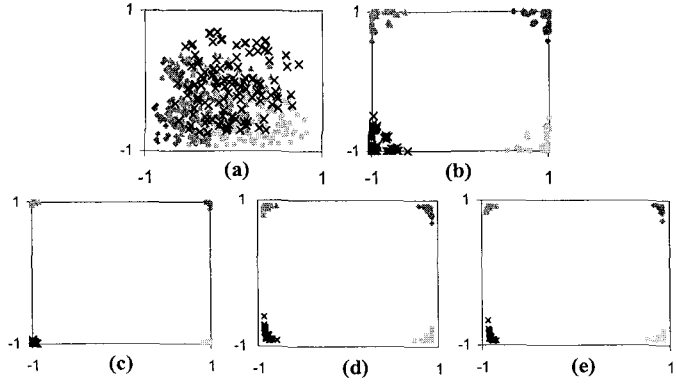


Figure 2. Internal representations of objects in the conditions of no-learning (a), no-language (b), social language (c), private speech (d), and inner speech (e). See text for details.

In order to study whether language can improve categorization not only if it is used socially but also if it is used individually, we simulated two ways in which humans can talk to themselves: externally, by speaking aloud but to themselves (private speech), and internally, by using language inaudibly (inner speech). In the private speech condition (figure 3c) the individual encounters an object and prior to responding to the object the individual produces the linguistic sound that designates the object. The individual hears the sound that he or she has produced and responds to an internal representation of the object which is influenced by the self-produced heard sound.

In the inner speech condition (figure 3d) when the individual sees the object he or she does not produce any externally audible sound. However, the sight of the object evokes in the individual not only the internal representation of the object but also, through the connections leading from the sensory-motor module to the linguistic module, the internal representation of the sound that designates the object. Through the connections leading from the linguistic module to the sensory-motor module, this internal representation of the sound that designates the object can influence the internal representation of the object in the sensory-motor module. The results of the both the private language and the inner language conditions (figure 2d and 2e, respectively) confirm the prediction that categorization is improved not only by social language, but also by talking to

oneself. In fact, both self-produced and internally-thought labels improve internal representations of perceived objects more or less to the same extent as external speech.

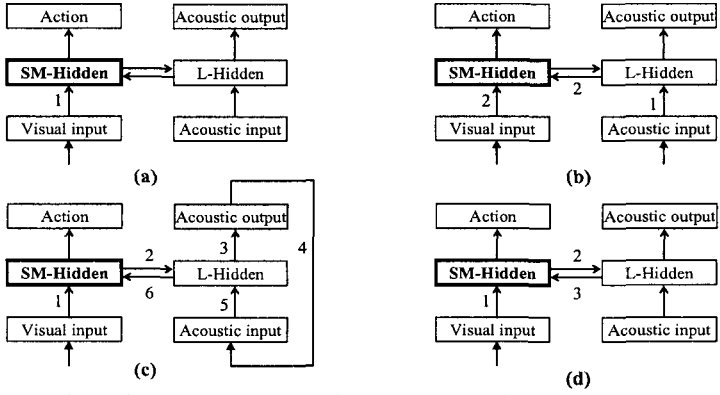


Figure 3. Experimental conditions for testing internal representations. The process of activation is divided up into a sequence of discrete time steps. Numbers indicate connections that are involved in each time step. See text for details.

4. Conclusions

Using a very simple neural network model of language acquisition it has been possible to demonstrate that using language for oneself can improve an individual's categorization of the world. We conclude that human language may have emerged not only due to social pressures, i.e., because language made it possible to have more sophisticated forms of social communication and coordination, but also due the advantages that language conferred on the cognitive functioning of the single individual when it is used to talk to oneself, either aloud or internally. Since these advantages can be demonstrated even with the very simple "language" of our simulations, we conclude that it is not necessary for language to be as sophisticated and complete as present-day language to provide individual cognitive advantages but that these advantages were probably already present in the very early stages of the evolutionary emergence of human language.

Acknowledgements

The research presented in this paper has been supported by the ECAGENTS project founded by the Future and Emerging Technologies program (IST-FET) of the European Community under EU R&D contract IST-2003-1940.

References

Bickerton, D. (1990). *Language and Species*. Chicago: University of Chicago Press.

Cangelosi, A., Harnad, S. (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, 4(1), 117-142

Chomsky, N. (2002). *On Nature and Language*, Cambridge, MA: Cambridge University Press.

Clark, A. (2005). Beyond the Flesh: Some Lessons from a Mole Cricket. *Artificial Life*, 11 (1-2), 233-244

Dennett D.C. (1991). *Consciousness Explained*. New York: Little Brown & Co.

Diaz R., & Berk L. (Eds.) (1992). *Private speech: From social interaction to self regulation*. New Jersey: Erlbaum.

Gentner, D. (2003). Why we're so smart. In Gentner & Goldin-Meadow (2003) (pp.195-235).

Gentner D., & Goldin-Meadow S. (Eds.) (2003). *Language in mind*. Cambridge, MA: MIT Press.

Gruber O., & Goschke T. (2004). Executive control emerging from dynamic interactions between brain systems mediating language, working memory and attentional processes. *Acta Psychologica*, 115, 105-121.

Hurford, J. (2002). The roles of expression and representation in language evolution. In A. Wray (Ed.), *The transition to language* (pp. 311-334). Oxford: Oxford University Press.

Knight, C., Studdert-Kennedy, M., & Hurford, J. (Eds.) (2000). *The evolutionary emergence of language: Social function and the origins of linguistic form*. Cambridge, MA: Cambridge University Press.

Lupyan, G. (2005). Carving nature at its joints and carving joints into nature: How labels augment category representations. In A. Cangelosi, G. Bugmann and R. Borisyuk (Eds.) *Modelling Language, Cognition and Action* (pp. 87-96). Singapore: World Scientific.

Mirolli, M., & Parisi, D. (2005a). Language as an Aid to Categorization: A Neural Network Model of Early Language Acquisition. In A. Cangelosi, G. Bugmann and R. Borisyuk (Eds.) *Modelling Language, Cognition and Action* (pp. 97-106). Singapore: World Scientific.

Mirolli, M., & Parisi, D. (2005b). How can we explain the emergence of a language which benefits the hearer but not the speaker? *Connection Science, 17(3-4)*, 325-341.

Nazzi T., & Gopnik A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition, 80*, 303-312.

Pinker, S., & Jackendoff, R. (2005) The faculty of language: What's special about it? *Cognition, 95(2)*, 201-236.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

LEARNING MODELS FOR LANGUAGE ACQUISITION

SHASHI MITTAL, HARISH KARNICK
Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India 208016.
mshashi@iitk.ac.in, hk@cse.iitk.ac.in

In this paper, we present a model of language acquisition which can be used to explain how children learn a grammar by interacting with their surroundings. We build upon the model proposed by Komarova et al in the context of evolution of grammars. We test our model for two situations : One, in which an individual is trying to learn a grammar in an environment where everybody uses the same grammar, and the other in which different groups in the population use different grammars.

1. Introduction

Komarova et al (Komarova, Niyogi, & Nowak, 2001; Komarova & Nowak, 2002) have proposed a computational model for the evolutionary dynamics of grammar acquisition in a population of individuals. Along similar lines we propose a model for grammar acquisition for a new language learner. The fact that a human learner acquires a language and in particular its grammar, coupled with impossibility results from computational learning theory (Gold, 1967; Angluin & Kharitonov, 1995) imply that there exists an inherent learning bias that makes learning the grammar of a language from limited input computationally feasible. One possibility is the presence of a finite set of candidate grammars one of which gets selected as the most appropriate based on linguistic input (Nowak, Komarova, & Niyogi, 2002). Here, we study language acquisition mainly in the context of learning the correct grammar from the available linguistic data.

2. The grammar system

Komarova et al's model is briefly as follows:

- **UG** = Finite search space of candidate grammars, G_1, G_2, $\ldots G_n$, also called the *Universal Grammar*.
- s_{ij} = Probability that a speaker who uses G_i utters a sentence compatible with G_j. It represents the similarity between different grammars.
- $F(G_i, G_j) = (s_{ij} + s_{ji})/2$, the payoff for mutual understanding.

- x_i = Relative abundance of individuals who use grammar G_i.
- $f_i = \sum_{j=1}^{n} x_j F(G_i, G_j)$, This is the payoff for an individual using grammar G_i.
- $\{q_{ij}\}$ is the stochastic learning matrix, q_{ij} denotes the probability that a child born to an individual using G_i will develop G_j.

The population dynamics of grammar acquisition (Komarova et al., 2001; Komarova & Nowak, 2002) is:

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} f_i q_{ij} x_i - \phi x_j, j = 1, 2, \ldots, n. \tag{1}$$

where $\phi = \sum_{i=1}^{n} f_i x_i$ denotes the average fitness, or *grammatical coherence* of the population. In general, ϕ is a number between 0 and 1.

To model language acquisition dynamics for a single individual, we introduce the following concepts:

- A = number of individuals in the population
- p_{ij} = probability that the ith individual uses G_j at a particular instant in time. Clearly, $\sum_{j=1}^{n} p_{ij} = 1$, and also $x_j = \frac{1}{A}\sum_{k=1}^{A} p_{kj}$.
- $\psi_i = \sum_{j=1}^{n} p_{ij} f_j$, this represents the payoff for the ith individual.

3. Grammar acquisition dynamics of an individual

Following Komarova et al, the learning model for the evolutionary dynamics of grammar acquisition is: Let $\{q_{ij}\}$ be the stochastic learning matrix, where q_{ij} denotes the probability that an individual using grammar G_i will switch to using grammar G_j in the next turn. Note that this interpretation of the stochastic learning matrix is different from the one described in the previous section. Using this stochasticity matrix, the learning dynamics of an individual is given by:

$$\frac{dp_{ij}}{dt} = \sum_{k=1}^{n} f_k q_{kj} p_{ik} - \psi_i p_{ij}, \tag{2}$$
$$i = 1, \ldots, A, \quad j = 1, \ldots, n$$

In all, we have $A \times n$ differential equations, where p_{ij} corresponds to the probability that individual i uses grammar G_j from the **UG**.

4. A simple learning model

The simplest learning model is to assume that all q_{ij} are constants. To simplify our analysis, we assume the q matrix is symmetric, and is given by

$$q_{ii} = q, \ , i = 1, \ldots, n \tag{3}$$

$$q_{ij} = \frac{1-q}{n-1}, i \neq j \tag{4}$$

Further, we assume a *fully symmetrical system*, that is

$$s_{ij} = s, i \neq j, 0 < s < 1 \tag{5}$$

4.1. *Language acquisition dynamics when all population members use the same grammar*

This problem can be formulated as follows : We assume that there are n grammars in the universal grammar set, and there are A individuals in the population. Without loss of generality, we assume that individuals from 1 to $A-1$ have chosen grammar G_1, and we are interested in studying the dynamics of the Ath individual. Assuming that the Ath individual uses all the grammars except G_1 with uniform probability, the above equation reduces to the following form

$$\frac{dp_{A1}}{dt} = \alpha p_{A1}^3 + \beta p_{A1}^2 + \gamma p_{A1} + \delta \tag{6}$$

where, $\alpha = -(\frac{1-s}{A})$, $\beta = (1-s)(\frac{q+1-A}{A})$, $\gamma = q(1-\frac{1-s}{A}) - \frac{s(1-q)}{n-1} - s$ and $\delta = \frac{s(1-q)}{n-1}$.

The initial condition for the equation is

$$p_{A1} = 1/n, \ t = 0 \tag{7}$$

that is, initially each grammar has equal probability of being used. We are interested in studying the behavior of this initial value problem, in particular we want to see whether p_{A1} always attains an equilibrium, if so what is the equilibrium value and how do the parameters s, q, A and n influence the acquisition process. Mathematica was used to study the behaviour of the differential equations.

The probability p_{A1} (henceforth referred to as p in this section for convenience) always converges, though the value to which it converges depends upon the values of the parameters. The value of p reaches 1.0 only if $q = 1.0$ (i.e. learning fidelity is perfect). The effect of changing the parameters q, s, n and A can be summarized as follows :

- If the value of q is increased, keeping other variables fixed, the value of p converges to a higher value as shown in Fig 1. At $q = 1.0$, the final value of p is 1.0, irrespective of the value of s.
- If only s is changed, the value to which p converges decreases, and so does the rate of convergence, as shown in Fig. 2. ($q = 0.7, n = 10$ and $A = 10$).
- With increasing n, the convergence is attained at a slower rate, although it always converges to the same value.
- Changing the value of A does not show any significant impact on grammar acquisition dynamics.

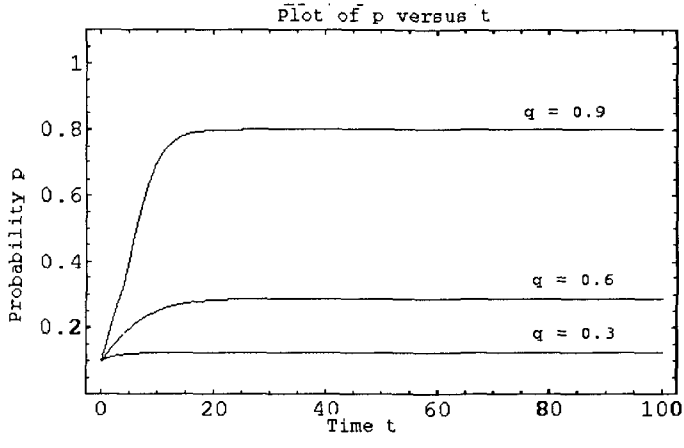


Figure 1. Plot of p versus t when q is varied.

4.2. *Learning mechanisms*

A learning mechanism defines the dependence of $\{q_{ij}\}$ on N, the number of learning events. The results for two learning algorithms, both of which have been extensively studied in the literature, are presented here.

- **Memoryless learning:** The learner starts with a randomly chosen hypothesis (say G_i) and stays with this hypothesis as long as the sentence heard is compatible with this hypothesis. If a sentence is not compatible, the learner randomly chooses another grammar from the **UG**. The process stops after N sentences. For a fully symmetrical system, the dependence of q on N is

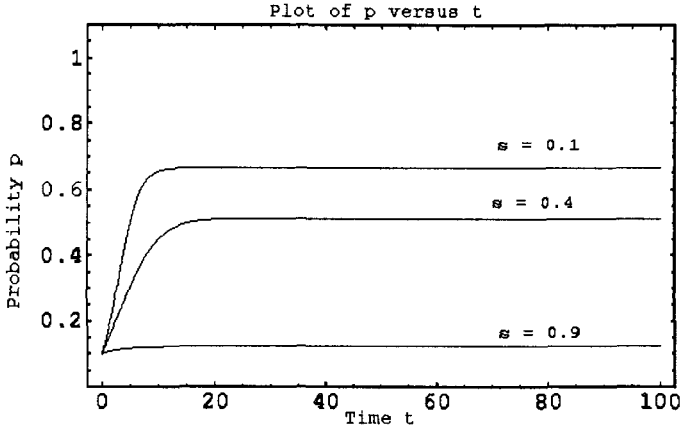


Figure 2. Plot of p versus t when s is varied

given by

$$q = q_{ii} = 1 - (1 - \frac{1-s}{n-1})^N \frac{n-1}{n} \tag{8}$$

- **Batch learning:** The Batch learner is first exposed to and memorizes all N sentences and then chooses a grammar from the **UG** that is most compatible with the input. For a fully symmetrical system,

$$q = q_{ii} = \frac{(1 - (1 - s^N)^n)}{s^N n} \tag{9}$$

It can be seen that the value of q will be higher for the batch learner compared to the memoryless learner, for the same values of s, n and A. Human learning is likely to be intermediate between these two algorithms, and therefore human performance is expected to lie somewhere between these two. Fig. 3 shows the plot of p_{A1} when $s = 0.4$, $N = 40$, $n = 25$ and $A = 10$. The memoryless algorithm converges to $p = 0.40$, whereas the batch learner algorithm converges to $p = 0.85$. The corresponding values of q for the two algorithms are 0.65 and 0.92 respectively.

4.3. *Language acquisition dynamics when different population members use different grammars*

We formulate this problem as follows : The **UG** has n grammars in . Members 1,2,..., A use grammar G_1, members $A + 1$, $A + 2$, ..., $2A$ use G_2, and so on.

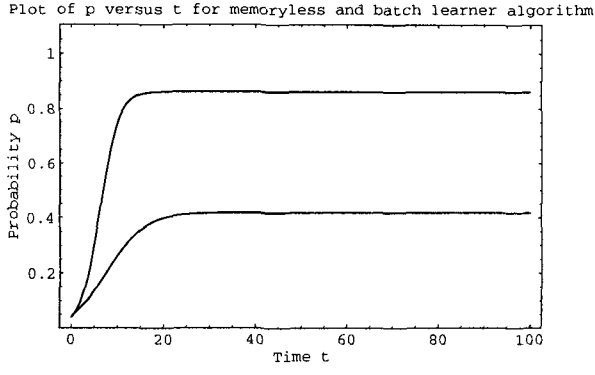


Figure 3. Plots for memoryless and batch learner algorithm

The $(nA + 1)$th member is the learner and is interacting uniformly with all the groups. The dynamics of grammar acquisition for this member is given (in a fully symmetrical situation) by:

$$\frac{dp_j}{dt} = \left(\frac{1-q}{n-1}\right) \sum_{k=1, k\neq j}^{n} f_k p_k + q f_j p_j - \left(\sum_{k=1}^{n} f_k p_k\right) p_j$$

we use p_i for the probabilities of the learning individual.

$$x_j = \frac{\sum_{k=1}^{nA+1} p_{kj}}{nA+1} = \frac{A + p_j}{nA+1}$$

$$f_i = \frac{[(n-1)s + 1]A + s(1-p_i) + p_i}{nA+1}$$

$$\psi = \sum_{k=1}^{n} f_k p_k$$

For this situation, irrespective of the initial values of the probabilities, and the values of s, q, n and A, it is observed that the probability values $p_1, \ldots, p_n$ all converge to the same value. Figure 4 shows the plot for one such case, when $s = 0.2$, $n = 3$, $A = 5$ and $q = 0.8$.

5. A simulated annealing learning model

In the learning models described above, the value of q (the learning coefficient) had been kept constant. In the simulated annealing learning model, the value of q

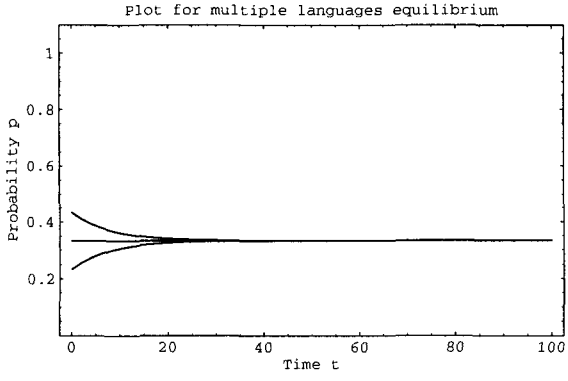


Figure 4. Plots for multiple languages case

changes with time and is given by:

$$q = e^{(\psi - 1)/kt} \tag{10}$$

where $\psi = \sum_{k=1}^{n} f_k p_{ik}$, and k is a constant (fixed at 1.0).
Such a choice of q satisfies the following two important properties (note that q is q_{ii}).

1. If the individual's grammatical coherence is high (i.e. ψ is close to 1, then q is close to 1, i.e. the individual has a lower tendency to switch to another grammar.

2. As time progresses, q tends to 1, i.e. if learning has taken place initially, then there is less likelihood the individual will change to another grammar. However, when t is small q is close to 0 and the learner is likely to switch grammars during early learning.

For the case when all the population members use the same grammar and the simulated annealing learning model is used, the probability of using that particular grammar *always converges to 1*, irrespective of the values of s or n. For the multilingual environment case, the probabilities tend to converge to the same values initially, but subsequently only one of the grammars attains the probability 1 and for other grammars the probability of usage tends to zero. This is shown in Fig. 5.

6. Conclusion

In this paper, we have presented two possible models of learning (the simple learning algorithm and the simulated annealing learning algorithm), and analyzed the behavior of a learner for monolingual and multilingual environments. The simulated annealing model has the interesting consequence that the learner learns a

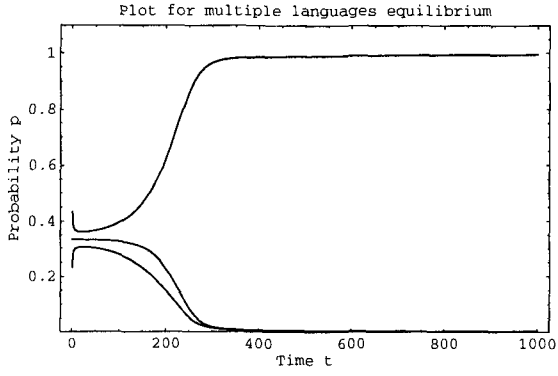


Figure 5. Plot for the multiple language case when simulated annealing model is used.

single language perfectly. The work can be extended by studying the dynamics using more realistic assumptions for the variation of q with time and for the nature of interaction between the learner and the group of mature individuals.

References

Angluin, D., & Kharitonov, M. (1995). When won't membership queries help? *J. Comput. Syst. Sci., 50*(2), 336-355.

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 10*(5), 447-474.

Jain, S., Osherson, D., Royer, J. S., & Sharma, A. (1999). *Systems that learn* (2nd ed.). Cambridge, MA, USA: The MIT Press.

Komarova, N. L., Niyogi, P., & Nowak, M. A. (2001). The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology, 209*, 43-59.

Komarova, N. L., & Nowak, M. A. (2002). Population dynamics of grammar acquisition. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language*. New York, NY, USA: Springer-Verlag New York.

Mittal, S. (2005). *Investigating learning models for language acquisition* (Tech. Rep.). Indian Institute of Technology, Kanpur. Available at http://www.cse.iitk.ac.in/reports/view.jsp?colname=446.

Niyogi, P. (2004). *The computational nature of language learning and evolution.* In press. Available at http://people.cs.uchicago.edu/ niyogi/Book.html.

Nowak, M., Komarova, N., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature, 417*(6), 611-617.

SIMULATING THE EVOLUTIONARY EMERGENCE OF LANGUAGE: A RESEARCH AGENDA

DOMENICO PARISI
Institute of Cognitive Sciences and Technologies
National Research Council,
via S. Martino della Battaglia 44, 00185 Rome
domenico.parisi@istc.cnr.it

1. Why simulations can be useful

If one is interested in studying the evolutionary emergence of human language, one is confronted with two formidable but well recognized problems. First, compared with animal communication systems, human language is a much more complex system for communicating with others, and therefore its evolutionary emergence must necessarily be correspondingly more complex. Second, we don't have much direct empirical evidence concerning how and when human language has emerged in the course of the evolution of the species, and therefore we are restricted to hypotheses and theories that must remain to a large extent speculative. With respect to these two problems there is not much that we can do. However, there is a third, less well recognized problem, and with respect to this problem there is something that we can, and should, do. Hypotheses and theories about the evolutionary emergence of language abound but it is notoriously very difficult to reach a consensus on which one to accept or reject. This is not only the inevitable consequence of a restricted empirical basis. Most hypotheses and theories about language's evolutionary emergence are only verbally formulated and, therefore, they tend to be insufficiently defined, precise, and articulated, and it is hard to derive from them uncontroversial empirical predictions. Hence, not only we do not have much direct empirical evidence on language's evolutionary emergence but we do not address the little, mostly indirect, evidence that we do have with well defined predictions uncontroversially derived from well defined hypotheses and theories.

This is where computer simulations can be of help. Computer simulations are hypotheses and theories which are expressed not in words or mathematical symbols, as is traditional in science, but as computer programs. When the program runs in the computer the results of the simulation are the empirical predictions that are derived from the hypothesis or theory which is incorporated

in the program. Why can this be of help? If a theory is expressed in the form of a computer program, i.e., as a simulation, the theory must necessarily be well defined, precise, and articulated because, otherwise, the program cannot be written or it will not run in the computer. Furthermore, since a simulation's results are the empirical predictions derived from the theory incorporated in the program, a theory expressed as a simulation necessarily (mechanically) and uncontroversially generates many detailed empirical predictions that can be confronted with the empirical data. Furthermore, simulations can be tools for thinking in that they can suggest new hypotheses and new types of empirical evidence.

Researchers in language evolution who do not use simulations but use the more traditional and well established methods of linguistics, psychology, anthropology, ethology, archaeology, and palaeontology, tend to be suspicious of computer simulations. Simulations simplify with respect to reality and to actual empirical phenomena, and these researchers are too well aware of the extreme complexity and diversity of empirical reality to become interested in such simplified models. But all scientific theories simplify with respect to reality, and they make us better understand reality only because they simplify, allowing us to capture the mechanisms and processes that underlie the variety and complexity of empirical phenomena and explain them. Since simulations are theories, they should simplify. The real problem is not that they simplify but that they should make the appropriate simplifications, including the aspects of reality which are critical in order to explain the particular phenomena of interest and leaving other aspects out. Therefore, computer simulations should not be criticized a priori and in general but they should be examined and evaluated case by case.

2. What types of simulations?

If one wants to reproduce in an artificial system - a computer simulation or a community of physical robots - the evolutionary emergence of language, one has to make a number of choices. A simulation of the emergence of language must necessarily assume the existence of a set of agents that represent the human population within which language has evolved. One first set of choices concern the nature of the agents. I believe that, to maintain a continuity between the biological and the cognitive sciences, these agents should possess three properties: (1) their behaviour should be controlled by a neural network, i.e., a control system that simulates the physical structure and the physical way of functioning of the nervous system (Rumelhart and McClelland, 1986); (2) the agents should be embodied, i.e., the neural network should be part of a simulated

physical body, with a given size, shape, and given sensory and motor organs (Nolfi and Floreano, 2000; Pfeifer and Scheier, 2001); (3) they should live in, and interact with, a simulated physical environment that includes conspecifics, other animals, objects, and possibly technological artefacts (Parisi, Cecconi, and Nolfi, 1990).

A second set of choices concern the nature of the process, or processes, of acquisition through which language emerges. If neural networks are used as the control system for the agents, it becomes a necessity to simulate a process of acquisition of whatever ability those agents will eventually possess because neural networks cannot be directly programmed by the researcher. The researcher defines the scenario within which some ability is acquired but then the simulation must show if, how, and in what conditions, the ability is actually acquired. Therefore, one has to decide which algorithm to use to "train" the networks, i.e., to go from networks that do not possess language to networks that do possess language. The problem is that underlying language and its evolutionary emergence there are many distinct acquisition processes and one must be able to simulate all of them, and how they interact with each other.

The first process is biological evolution. Human language is learned during the first years of life but it is learned only because there are inherited predispositions that make the learning possible, as shown by the fact that non-human animals do not acquire a language even if they are exposed to human language. Hence, one thing that one must be able to simulate is the process of biological evolutionary emergence of these inherited predispositions for language learning, which can be either specific for learning language or can be more general predispositions which are species-specific for Homo sapiens but not specific for language. To simulate the evolutionary emergence of biologically inherited abilities or predispositions a procedure called genetic algorithm is used (Holland, 1975). A population of agents with individually different genotypes live in an environment and reproduce by generating offspring that inherit their genotypes. Genotypes encode specifications for the agents' neural networks. Reproduction is selective, with some individuals having more offspring than other individuals, and the offspring's genotypes are somewhat different from the genotypes of their parents because of random genetic mutations and sexual recombination. The two processes of selective reproduction and of constant addition of new variability due to mutations and sexual recombination cause evolutionary changes in inherited genotypes and the emergence of initially absent abilities that make survival and reproduction more likely in the given environment.

The second process of acquisition which is involved in language is learning. A human being is not born with language but he or she learns the particular language which is spoken in his/her environment, although this is only possible because of the biologically inherited predispositions to learn any human language that I have already referred to. Language learning actually is language development, where by development I mean a process of acquisition which needs, and is modulated by, environmental input but goes through stages that are specified in the inherited genotype. This implies that the genotype of the agents must encode a developmental program, not just a set of already existing abilities or a set of predispositions to learn. A developmental program is a genetically inherited schedule for acquiring various behaviours in a sequence of temporal stages. What must evolve, therefore, is not only the content of these stages but also the sequence of stages itself.

Finally, human language involves a third type of acquisition process: cultural or, more precisely, linguistic evolution (language change). Language is learned from others, i.e., culturally. Hence, one has to simulate agents that not only learn in general terms but learn by imitating others. Learning from others creates a second form of evolution, cultural evolution. Behaviours are transmitted (imitated) from one generation to the next, and they are transmitted selectively, with some behaviours more imitated than other behaviours. Furthermore, behaviours can change either because of random errors in transmission, analogous to random mutations, because imitators recombine in new ways different aspects of the behaviour of different models, analogous to sexual recombination, and because of inventions and internal re-organization of the behavioural repertoire. As in biological evolution, this results in cultural evolution, with the emergence of new forms of culturally transmitted behaviour, in our case, new forms of language (language change).

Furthermore, groups of interacting agents tend to learn from each other and therefore to develop similar behaviours, including linguistic behaviours. Shared linguistic behaviours constitute historical languages. If a group of agents, for some reason, splits into separate sub-groups with little interactions between the sub-groups, distinct dialects or even different languages can emerge from a single initially shared language because culturally transmitted behaviours change all the time and the changes that occur in one sub-group tend to diverge from those in other sub-groups. This is similar to biological phenomena of genetic divergence and the emergence of new species.

Some simulations based on the above assumptions and that address some of the phenomena of language evolution that I have briefly described have already

been realized. However, most of the work in this area is still to be done, so what I am talking about is mainly a research agenda. (For a collection of articles on simulating the emergence of language, see Briscoe, 2002; Cangelosi and Parisi, 2002.)

3. Genetically inherited predispositions for language learning

Consider the species-specific biologically inherited pre-dispositions for language learning. As I have said, these predispositions can be either general or specific for language learning. Among the predispositions that are general and not specific for language learning there might be a species-specific tendency to learn by imitating others. Learning by imitating others may presuppose an ability to learn to predict the perceived effects of one's own actions (movements). This ability to predict the effects of one's actions appears to underlie learning by imitating any kind of behaviour, not only linguistic behaviour. Another algorithm, the backpropagation procedure (Rumelhart and McClelland, 1986), can be used to simulate both learning to predict the effects of one's own actions and learning to imitate the actions of others (Jordan and Rumelhart, 1992). In the first case, the learner compares the predicted effects of its own actions with their actual effects and, using the results of this comparison, adjusts the connection weights of its neural network in such a way that, in a series of learning experiences, any initial discrepancy between predicted and actual effects gradually disappears. In the second case the learner compares the predicted effects of its own actions with the perceived effects of the actions of another individual and adjusts its connection weights in such a way that it becomes eventually able to produce the same effects that are produced by the actions of others and, therefore, presumably, the same behaviours.

These two learning processes may underlie the successive stages of prelinguistic and then linguistic behaviour in the child from birth to 1 year and on. Stage 1: the production of all kinds of sounds in the very first months of life, when the child is learning to predict the acoustic consequences of his or her phono-articulatory movements. Stage 2: the emergence of babbling at 4-6 months, when the child is learning to imitate his or her own sounds (self-imitation). Stage 3: the tendency of the sounds produced by the child in the second semester of life to resemble the sounds of the particular language spoken in the child's environment, when the child is learning to imitate the sounds produced by others. Stage 4: the first emergence of true language at around 1 year of age, when the child is learning to produce the same sounds produced by another individual in response to the same objects and actions. The separate

stages of this developmental process have already been simulated (Floreano and Parisi, 1992) but the challenge is to evolve a developmental genotype that encodes the entire developmental sequence.

Stage 4 explains the typically referential nature of human language in contrast to the mostly non-referential nature of animal communication. Linguistic signals acquire their referential meaning because particular sounds are systematically paired, in the learner's experience, with particular objects and actions. Signals that are paired with objects, independently of the specific action which is done with respect to the object, become nouns, while signals paired with actions, independently of the specific object on which the action is done, become verbs. One can also simulate the successive emergence of other parts of speech (Parisi, Cangelosi and Falcetta) and the emergence of complex signals made up of simpler signals (Cangelosi, 1999).

Learning to predict the effects of one's actions and learning by imitating others can be based on species-specific biologically inherited predispositions of Homo sapiens that underlie language learning but are not limited to language. This is consistent with the idea that these same predispositions may underlie other typically human traits such as constructing and using all sorts of technological artefacts by predicting the effects of one's actions on the artefact and the effects of the artefact on the environment. Language learning, however, can also be based on biologically inherited predispositions that are specific for language such as a particular sophistication of the sensory-motor sub-network, or module, that maps heard sounds into phono-articulatory movements or the ability to parse and construct linguistic signals that are combinations of smaller signals (syntax). However, in evaluating these claims of linguistic specificity one must consider that deaf children are able to learn non-acoustic sign languages (which may have preceded the evolutionary emergence of acoustic languages) and that linguistic syntax might emerge from already existing and more general abilities to analyze and to generate complex actions as combinations of simpler actions. Simulations can be of help here by allowing us to test alternative hypotheses. For example, would a population of agents that has developed a language in the acoustic/phono-articulatory mode be able to easily switch to a language in the visual/motor mode? Or, can we demonstrate a general ability to parse and construct complex actions as combinations of simpler actions even in simulated organisms with no language?

4. Adaptive functions of language

Other interesting research issues concerning language evolution that can be addressed with simulations of the type I have described are the adaptive role of linguistic behaviour and the function of language as a tool for communicating with oneself (thinking), and not only with others, which appears to be another property that distinguishes human language from animal communication systems.

The present framework assumes that language has evolved because it was adaptive but one can create and study different evolutionary scenarios in which one contrasts various adaptive roles for language. For example, language is a social behaviour and it involves at least two agents, the speaker and the hearer. Therefore, one can ask: Is language adaptive for the speaker or for the hearer? Imagine an agent which is looking for food but cannot recognize if an encountered but distant mushroom is edible or poisonous. The only option for the agent is to get near to the mushroom and eat the mushroom if it is edible and go away if it is poisonous. In these circumstances the agent would benefit from the linguistic behaviour of another agent which is nearer to the mushroom and tells the first agent if the mushroom is edible or poisonous, saving the first agent's time and energy if the mushroom is poisonous. However the behaviour of the speaker is advantageous for the hearer but is not advantageous for the speaker or can even be disadvantageous if the two individuals compete for survival. So why should speakers evolve in this scenario? And if there are no speakers, there is no language. In fact, if one makes a simulation the results of the simulation show that in this scenario language will not evolve (Mirolli and Parisi, 2005). The behaviour of the speaker is altruistic in that it increases the survival and reproductive chances of the hearer but decreases its own. Therefore, the genes underlying speaking behaviour tend to disappear from the population and language does not emerge. Language emerges only if the speaker and the hearer share the same genes, as predicted by kin selection theory. By benefiting the hearer, the altruistic behaviour of the speaker increases the survival and reproductive chances of an individual which, by sharing the same genes of the speaker, also is a good speaker. Therefore, the genes underlying speaking behaviour remain and diffuse in the population.

The results of this simulation may suggest the hypothesis that language has first emerged in small groups of kin related individuals. However, this implies that different kin groups will speak different languages whereas language is particularly useful when it can be used more widely. How did a language which is spoken and understood by larger groups of non-kin individuals emerge? One

possibility is that language has emerged in a single, small group of kin related individuals, and then it has been culturally inherited in a progressively larger group of more distantly related individuals that were descendants of the original group. Another possibility, of course, is that language was used in situations in which it was useful to both speakers and hearers, such as the speaker asking for something from the hearer and the hearer being interested in knowing what is asked.

Still another possibility is that language was used from its very beginning not only for communicating with others but also for communicating with oneself. In fact, language will emerge if we modify the simulation scenario that I have described and require that the hearer, when it hears a signal from the speaker specifying that the mushroom is edible, has to repeat the signal to itself in order to keep it in memory while approaching the edible mushroom. This is using language for communicating with oneself, not with others. The results of the simulation show that if language is used for communicating with oneself, language emerges even if the speaker has not the same genes of the hearer (Mirolli and Parisi, 2005). The reason is that the hearer must also be a good speaker in order to speak appropriately with itself and be able to increase its reproductive chances. This simulation may suggest the hypothesis that using language to communicate with oneself, i.e., to think, may have been an adaptive pressure for the emergence of language from its earliest evolutionary stages, instead of supposing that the use of language for thinking is a recent discovery which requires an already well developed language.

5. Language change and differentiation

Finally, to demonstrate the scope of simulations I want to briefly mention another, very different, simulation that addresses language change and the historical process of linguistic differentiation which creates patterns of similarities and differences among genetically (historically) related languages. In this simulation what is simulated is the process of diffusion of farming in Europe which originated in Anatolia nine thousand years ago and reached the whole of Europe in 3-4 millennia. In the simulation the entire European territory is divided into relatively small cells with properties that specify the suitability of each cell for farming, with cells more favourable for farming and cells less favourable (sea, mountains, deserts). Farmers (or, more probably, the technology of farming, at least at increasing distances from the point of origin in Anatolia) follow particular paths in their diffusion in Europe which depend on the appropriateness of the particular territories for farming, with some paths dividing

into diverging paths at specific points. It turns out that the resulting tree of paths followed by the simulated farmers in their diffusion in Europe has some similarities with the tree of genetic relatedness of the languages spoken in Europe (Parisi, Antinucci, Cecconi and Natale, in press). This gives some support to theories on the origin of Indo-European languages in Anatolia nine thousand years ago with respect to other theories that hypothesize a more recent origin in the region to the north of the Black Sea and Caspian Sea.

Acknowledgements

Thanks to Marco Mirolli for his useful comments.

References

Briscoe, E. J. (2002). *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge: Cambridge University Press.

Cangelosi A. (1999). Modeling the evolution of communication: From stimulus associations to grounded symbolic associations. In D. Floreano, J. Nicoud, and F. Mondada (eds.), *Advances in Artificial Life*, New York, Springer, 1999, pp. 654-663.

Cangelosi, A., & Parisi, D. (eds.) (2002). *Simulating the Evolution of Language*. New York: Springer.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press, (MIT Press, 1992).

Jordan, M.I., & Rumelhart, D.E. (1992). Forward models: supervised learning with a distal teacher. *Cognitive Science*, 16, 307-354.

Mirolli, M., & Parisi, D. (2005). How can we explain the emergence of a language that benefits the hearer but not the speaker? *Connection Science*, 17, 325-341.

Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics*. Cambridge, MA: MIT Press

Parisi, D., Antinucci, F., Cecconi, F., & Natale, F. (in press). Simulating the expansion of farming and the differentiation of European languages. In B. Laks, and D. Simeoni (eds.) *Origins and Evolution of Language*. New York: Oxford University Press

Parisi, D., Cangelosi, A., & Falcetta, I. (2002). Verbs, nouns, and simulated language games. *Italian Journal of Linguistics*, 14, 99-114.

Parisi, D., Cecconi, F., & Nolfi, S. (1990). Econets: neural networks that learn in an environment. *Network*, 1, 149-168.

Parisi, D., & Floreano, D. (1992). Prediction and imitation of linguistic sounds by neural networks. In A. Paoloni (ed.), *Proceedings of the 1st Workshop on Neural Networks and Speech Processing*. Rome, Fondazione Bordoni, 50-61.

Pfeifer, R., & Scheier, C. (2001). *Understanding intelligence*. Cambridge, MA: MIT Press.

Rumelhart. D.E. and McClelland, J.L. *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*. Cambridge, Mass., MIT Press, 1986.

EVOLVING THE NARROW LANGUAGE FACULTY: WAS RECURSION THE PIVOTAL STEP?

ANNA R. PARKER

Language Evolution and Computation Research Unit, University of Edinburgh, 40 George Square, Edinburgh, EH8 9LL, Scotland

A recent proposal (Hauser, Chomsky & Fitch, 2002) suggests that the crucial defining property of human language is recursion. In this paper, following a critical analysis of what is meant by the term, I examine three reasons why the recursion-only hypothesis[a] cannot be correct: (i) recursion is neither unique to language in humans, nor unique to our species, (ii) human language consists of many properties which are unique to it, and independent of recursion, and (iii) recursion may not even be necessary to human communication. Consequently, if recursion is not the key defining property of human language, it should not be granted special status in an evolutionary account of the system.

1. Introduction

Hauser, Chomsky & Fitch (2002) (henceforth HCF) propose that the human language faculty (FL) consists of two types of property. Those which are found elsewhere in cognition (either human or non-human) form the broad language faculty (FLB), with those which are unique to language forming the narrow language faculty (FLN). This seems an uncontroversial delineation. What is more contentious, is where the dividing line is drawn. By placing only recursion in FLN, HCF suggest it is the single defining property of our linguistic abilities.

A number of interesting questions arise from this proposal. Firstly, what is meant by the term 'recursion'? HCF (and the ensuing rejoinders too) are strikingly vague. We must thus turn to the literature within and outwith our field to develop a clear definition. Secondly, is it true that there is nothing else in human language that is unique to it? Other unique properties would immediately invalidate HCF's argument. Thirdly, is recursion truly unique to human language? For HCF's recursion-only hypothesis to be upheld, this must be the case. Finally, is it the case that all languages exhibit recursion? A recursion-less human language would indicate that recursion cannot be *the* defining property of the system. Echoing Pinker & Jackendoff (2005) (henceforth PJ), this paper adds to their criticisms thorough analysis of recursion, examination of its uniqueness, and pinpointing of the crux of the recursion-less language argument.

[a] The recursion-only hypothesis is just that – a hypothesis. The authors do not "...define FLN as recursion by theoretical fiat..." (Fitch, Hauser & Chomsky, 2005:183) (henceforth FHC), and indeed in places they seem to retreat to a weaker position. However, as the authors also note, "[t]he contents of FLN are to be empirically determined" (ibid: 182). That is precisely the aim of this paper – to use empirical data to assess the hypothesis that " ...FLN only includes recursion..." (HCF: 1569).

2. Defining Recursion

A survey of the definitions of recursion available in the linguistics literature reveals a vagueness not conducive to our assessment. The computer science literature offers a little more formalisation, but in both cases there is little consensus on where to place the burden of explanation; certain definitions highlight the embedded nature of recursive structures, others use recursive phrase structure rules as their basis, others simply equate recursion with repetition. The most significant difficulty with definitions of recursion is their failure to make three important distinctions: recursion is not the same as iteration, recursion is not the same as phrase structure, and there are differing types of recursion.

One merit of the computer science definitions is that they draw our attention to an important feature of recursion – its memory requirements. In processing recursion we need to be able to keep track of where to return to once the embedded portion of the structure is complete. For this, we need a last-in-first-out type of storage device such as a pushdown stack.

2.1. *Three Crucial Distinctions*

Recursion versus iteration The first of three distinctions crucial in understanding recursion is the difference between recursion and the oft-confused iteration. This boils down to a distinction between embedding and repetition. While iteration simply involves repeating an action or object an arbitrary number of times, recursion involves embedding the action or object within another instance of itself.

When baking a cake, we might encounter a recipe instruction such as *"stir the mix until it becomes smooth"*. Following the instruction involves repeating some action over and over again until we reach the terminating condition. Importantly, each stirring action does not rely on the previous or the next. This is iteration. Once the cake has been baked, serving an equal-sized piece to each of sixteen guests involves repeating a cutting action over and over. We first cut the whole cake in half, then cut each half in half, then cut each quarter in half, and then cut each eighth in half. Here the process differs from the iterative example in that there is a dependency between actions; the output of each cutting action becomes the input to the next. Further, we cannot omit any intermediate action and end up with the same result; it is not possible to go from halves to eighths leaving out the step that gives us quarters. This is recursion.

Tail versus nested recursion The second distinction is between tail and nested recursion. The former is illustrated in possessive constructions - (1), and relative clause constructions - (2), the latter in centre embeddings - (3).

(1) John's brother's teacher's book is on the table.
(2) The man that wrote the book that Pat read in the cafe that Mary owns.

(3) The mouse the cat the dog chased bit ran.

While tail recursion involves embedding at the edge of a phrase, nested recursion involves embedding in the centre, leaving material on both sides of the embedded component. The latter type of embedding produces long-distance dependencies. It is these dependencies that, in turn, necessitate a device for keeping track. In processing (3), we must store the subject noun phrases we encounter in memory (in the order we find them), retrieving them only (in the opposite order) when we reach the verbs with which they are associated.
Tail recursion might appear to be just a case of iteration, given that it looks like the simple repetition of identical phrases. However, consider (4):

(4) John's mother loves him.

This cannot be analysed as a simple proposition with another NP tacked on the front. Instead, it must be analysed as a sentence with a complex subject NP, containing within it another NP. This is exactly what Pinker and Bloom (1990) were referring to when they noted that recursion allows us to specify reference to an object to an arbitrarily fine level of precision. The iterative analysis of (4) is not true to the complex meaning it reflects. In other words, in natural language semantics forces us to analyse iteration and tail recursion differently.
Recursion versus phrase structure The final important distinction is between recursion and phrase structure, concepts which are often erroneously equated in the linguistics literature. Phrase structure is the hierarchical ordering of phrases within a sentence. Importantly, a structure may be hierarchical without being recursive. While hierarchy involves phrases embedded within other phrases, recursion involves *identical* phrases embedded inside each other. Phrase structure is thus required in language for recursion, because we need the capacity to embed before we can embed identical elements, but phrase structure does not guarantee recursion.
We are now in a position to define recursion (and iteration) as follows:
Iteration : the simple unembedded repetition of an action or object an arbitrary number of times.
Recursion : the embedding at the edge or in the centre of an action or object one of the same type. Further, nested recursion leads to long-distance dependencies and the need to keep track, or add to memory.

3. The Uniqueness of Recursion

Armed with a better understanding of recursion, we can turn to the next question: is recursion unique? HCF define FLN as that which is unique to language and unique to humans. If recursion fits this characterisation, there are three places we should not find it: human non-linguistic cognition, non-human non-communicative cognition, and non-human communicative cognition.

3.1. *Human Non-Linguistic Cognition*

Within the non-linguistic cognition of our species, a number of domains suggest themselves. Number is a reasonable possibility, but this should be ruled out as language and number may be evolutionarily linked (PJ, Chomsky, 1988, Hurford, 1987). In the visual domain, processes responsible for decomposition of complex objects and scenes may work in a recursive fashion, analogously to the earlier cake-cutting example. In social cognition, our theory of mind allows us to embed minds within minds - I can think that John thinks that Bill thinks that Mary thinks X. This is only possible with a complex conceptual structure capable of generating recursive propositions.

Music, like language, is organised hierarchically. However, ascertaining if a piece consisting of repeated phrases should be analysed iteratively or recursively will be very difficult. In language, semantics provides a pointer to structure, but in music there is no such pointer. Nevertheless, music offers more definitive evidence of recursion. Hofstadter (1980) suggests that on encountering a key change, the listener must store the tonic key in memory. Once the tonic key is resolved, the stack item can be popped off. In other words, there is a nesting of one musical key within another. Bach's "Little Harmonic Labyrinth", so called because its key modulations are so frequent and so complex that the listener is left confused as to where they are in relation to the tonic key, suggests a parallel with difficulties in processing nested recursion in language.

3.2. *Non-Human Non-Communicative Cognition*

In non-human non-communicative cognition, number can be ruled out as animals lack comprehension of the successor function, the basis of numerical recursion. Navigation studies within the travelling salesman paradigm point to a good place to look for recursion. Animals' complex action sequences, such as the food preparation techniques of mountain gorillas (Byrne & Russon, 1998), or the artificial fruit solving techniques of chimpanzees (Whiten, 2002), offer evidence of hierarchical reasoning, and may also provide an arena for future experimental testing for recursion.

Although attributing a full theory of mind to other species is controversial, it is less disputable that they have some degree of social cognition. Experiments (e.g. Tomasello *et al*, 2003) indicate that chimpanzees cannot embed minds within minds. However, the work of Bergman *et al* (2003) suggests that even with rudimentary aspects of a theory of mind, other species may be capable of recursive conceptual manipulation. Baboons classify themselves and their conspecifics both in a linear hierarchy of dominance, and in matrilineal kin groups. In other words, they are capable of forming conceptual structures such as [X is mother of Y [who is mother of Z [who is mother of me]]] or [X is more dominant than Y [who is more dominant than Z [who is more dominant than

me]]] – tail-recursively embedded associations, which (unlike the iterative counterparts) cannot be re-ordered while maintaining the correct relations.

3.3. *Non-Human Communicative Cognition*

Unfortunately, for non-human communication systems, the question of recursion turns out to be much more challenging. Animal communication systems can be divided into two types: (i) those with limited semantics, but a flat, non-hierarchical organisation, e.g. the dance of the honeybee (von Frisch, 1966), or the alarm calls of the Campbell's monkey (Zuberbühler, 2002), and (ii) those with a complex hierarchical organisation, but no semantics, e.g. bird song (Okanoya, 2002). While recursion will not be found in the first type, as hierarchy is required for recursion, the second type may have embeddings which could plausibly be recursive. The problem is that faced only with a string, and no pointer to its structure, we cannot distinguish tail recursion from simple iteration. Nested recursion, on the other hand, could be evidenced by a complex enough string alone. Although such strings are not currently attested in these systems, we can use this knowledge to narrow the scope of future research - to design experiments to test specifically for this type of recursion (nested) in this type of system (hierarchically organised).

In sum, recursion is not uniquely human or uniquely linguistic, and thus should not be characterised as a property of FLN. More interesting, however, is the fact that no hint of nested recursion is to be found in non-human domains, suggesting that the difference between human and non-human cognition may boil down to a difference in memory capabilities. Despite claims of methodological flaws (Perruchet & Rey (in press)), recent experimental work (Fitch & Hauser (2004)) may be interpreted as supporting this hypothesis. Tamarins, shown to be able to only learn strings of the form a^nb^n, might differ from humans, who can also learn those of the form $(ab)^n$, in being able to deal only with recursion of the tail variety[b]. This would suggest that what was crucial in the evolution of human language was not recursion but the enhanced stack-type memory necessary to deal specifically with nested recursion.

4. The Contents of FLN

HCF's recursion-only hypothesis means that recursion should be the only aspect of language that is unique to language and unique to humans; as PJ put it, the only feature of language that makes it 'special'. The next question is whether there are other such properties of language which are independent of recursion.

From a wide-ranging literature in linguistics, we can discern a number of uniquely linguistic features which cannot be explained in terms of recursion. A non-exhaustive list would include the following (see also PJ for alternatives): (i)

[b] A pointer to the structure involved would need to be incorporated to test the hypothesis.

structure dependence, (ii) the lexicon, (iii) movement, (iv) duality of patterning, (v) word order, and (vi) syntactic devices. All are to be found only in humans, and more specifically, only in human language. Moreover, none fall out of recursion either directly or indirectly.

Future research may, of course, uncover evidence of such properties in non-linguistic domains. This would mean re-assigning them to FLB, FLN then being the empty set. Yet the current state of play suggests expansion of FLN. Conceptually, the FLB/FLN distinction makes sense; empirically, HCF's division is in the wrong place.

5. Language without Recursion – the Case of Pirahã

The claim of HCF implies that a lack of recursion would reduce human language to something more like an animal communication system: "...animal communication systems lack the rich expressive and open-ended power of human language (based on humans' capacity for recursion)" (HCF: 1570). The question is: do languages without recursion exist? If they do, are they as expressive as languages which do make use of recursion? And importantly, would a language without recursion still look like a human language, or would we wish to class it as closer to non-human communication?

Everett (1986, 2005) has argued that the Amazonian language Pirahã does not make use of recursion. Pirahã uses alternate means to express what would be expressed in English-type languages using recursive subordinate embedding.

(5) ti baósa -ápisí 7ogabagaí. Chico hi goó bag -áob.
I cloth -arm want. name 3 what sell –completive
'I want the hammock. Chico what sold'

(6) hi gái- sai xahóápáti ti xi aagá-hóág-a
3 say NOMLZR name 1 hunger have-INGR-REM
(i)'Xahóápáti said, "I am hungry"'
or (ii) 'Xahóápáti said (that) I am hungry' (Everett 1986, 2005)

(5) shows juxtaposition used to express a clausal modification of the noun, while (6) shows that indirect speech is expressed in the same way as direct speech, leaving it up to the pragmatics to determine the referent of the pronoun.

Pirahã permits only one possessor - (7). Again, juxtaposition is used to express recursive possession - (8).

(7)a. *kó7oí hoagí kai gáihií 7íga
name son daughter that true
'That is Kó7oí's son's daughter'

b. kó7oí kai gáihií 7íga
name daughter that true
'That is Kó7oí's daughter'

(8) 7ísaabi kai gáihií 7íga. Kó7oí hoagí 7aisigí -ai
name daughter that true. name son the same be
'That is 7ísaabi's daughter. Kó7oí's son being the same' (Everett, 2005)

If the criterion for syntactic recursion is that there must be embedded inside a phrase one of the same type, the Pirahã data cannot be analysed as recursive.

This data tells us that human language without recursion is indeed possible. It also tells us that Pirahã speakers are perfectly capable of expressing the same underlying conceptual structures as English speakers (although arguably in a somewhat less efficient or less compressed way): "...Pirahã most certainly has the communicative resources to express clauses that in other languages are embedded..." (Everett, 2005: 631). The crucial point missed in the later installments in the recursion-only debate is that Pirahã is a full human language, not a system akin to the communication systems of other species. It is a language that exhibits uniquely human, uniquely linguistic properties, and that can only be acquired by those in possession of a human LAD.

So, here we appear to be faced with a human language lacking the one property HCF set out as the defining characteristic of human language. FHC's invocation of Jackendoff's (2002) toolkit hypothesis: that "...our language faculty provides us with a toolkit for building languages, but not all the languages use all the tools" (FHC: 204), just will not wash. For HCF, recursion is the one tool which defines human language. But, if a language can get on just as well without recursion, surely it must be only one of a number of tools in the set which makes language unique. And, if recursion is not the crucial defining property of human language, then its place in an evolutionary account of the system becomes far less important.

6. Conclusion

The initial question posed - is recursion the pivotal step in the evolution of FLN? – must be answered with a resounding no. Three arguments support this answer. Firstly, recursion exists in domains outside language. In other words, it is not unique to human language, and so should not be placed in FLN. Secondly, many properties of human language, which are entirely independent of recursion, are absent from non-linguistic domains. That is, FLN consists of much more. Finally, data from a full human language without recursion suggests that it is not crucial to the communication system of our species. Therefore, I submit that the recursion-only hypothesis of HCF is flawed.

References

Bergman, T., Beehner, J., Cheney, D. & Seyfarth, R. (2003). Hierarchical classification by rank and kinship in baboons. *Science, 302*, 1234-6.

Berwick, R. (1998) Language evolution and the minimalist program. In, J. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), *Approaches to the Evolution of Language*, 320-40. Cambridge: Cambridge University Press.

Byrne, R. & Russon, A. (1998) Learning by imitation: a hierarchical approach. *Behavioral and Brain Sciences, 21*, 667-721.

Chomsky, N. (1988) *Language and Problems of Knowledge: The Managua Lectures.* Cambridge, MA: M.I.T. Press.

Everett, D. (1986) Pirahã. In D. Derbyshire & G. Pullum (Eds.), *Handbook of Amazonian Languages, volume 1,* 200-326. Berlin: Mouton de Gruyter.

Everett, D. (2005) Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language. *Current Anthropology, 46(4)*, 621-46.

Fitch, W. T., Hauser, M. & Chomsky, N. (2005) The evolution of the language faculty: clarifications and implications. *Cognition, 97(2),* 179-210.

Fitch, W. T. & Hauser, M. (2004) Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*, 377-80.

Hauser, M., Chomsky, N. & Fitch, W. T. (2002) The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*, 1569-79.

Hofstadter, D. (1980) *Gödel, Escher, Bach: An Eternal Golden Braid.* London: Penguin.

Hurford, J. (1987) *Language and Number: The Emergence of a Cognitive System.* Oxford: Basil Blackwell.

Jackendoff, R. (2002) *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford: Oxford University Press.

Okanoya, K. (2002) Sexual display as a syntactic vehicle: the evolution of syntax in birdsong and human language through sexual selection. In A. Wray (Ed.), *The Transition to Language*, 46-64. Oxford: Oxford University Press

Perruchet, P. & Rey, A. (in press) Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review.*

Pinker, S. & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences, 13(4)*, 707-84.

Pinker, S. & Jackendoff, R. (2005) The faculty of language: what's special about it? *Cognition, 95(2)*, 201-36.

Tomasello, M., Call, J. & Hare, B. (2003) Chimpanzees understand psychological states – the question is which ones and to what extent? *Trends in Cognitive Sciences, 7*, 153-6.

Von Frisch, K. (1966) *The Dancing Bees: An Account of the Life and Senses of the Honey Bee.* London: Methuen.

Whiten, A. (2002) Imitation of sequential and hierarchical structure in action: experimental studies with children and chimpanzees. In K. Dautenhahn & C. Nehaniv (Eds.), *Imitation in Animals and Artifacts,* 38-46. London: M.I.T. Press.

Zuberbühler, K. (2002) A syntactic rule in forest monkey communication. *Animal Behaviour, 63*, 293-9.

FROM MOUTH TO HAND

DENNIS PHILPS

Department of English (IRPALL/ELANG), University of Toulouse-Le Mirail, 5, allées Antonio-Machado, 31058 Toulouse cedex 9, France

Within a semiogenetic theory of the emergence and evolution of the language sign, I claim that a structural-notional analysis of submorphemic data provided by certain reconstructed PIE roots and their reflexes, projected as far back as theories of the evolution of speech will permit by a principle of articulatory invariance, points to the existence of an unconscious neurophysiologically grounded strategy for 'naming' parts of the body. Specifically, it is claimed that the occlusive sounds produced by open-close movements of the mouth, which have been shown experimentally to be synchronized with open-close movements of the hand(s), may have functioned as 'core invariants'. Morphogenetically transformed into conventionalized language signs, these could have served to 'name' not only the mouth movements and articulators involved, but also the hand movements with which they appear to be coordinated, as well as the hand itself.

1. Linguistics and the evolution of language

As linguistic theories become more refined, and as the scientific study of language evolution advances, so interpenetration of knowledge has increased, encouraging some linguists to attempt to bridge the gap between the two fields. Yet the very nature of Saussurian linguistics, based as it is on the principle of the arbitrariness of the sign and on the conventional status of the latter, means that, to quote Nichols, "[T]here is no hope of recovering information about language origins by tracing linguistic descent." (1998: 128). In the field of neurolinguistics however, Buoiano seems to want to bridge this gap when he suggests that "we need a device that can define the sign as non-arbitrary within the frame of a neurolinguistic theory in order to explain why neurocognition and language have phylogenetically developed using (also) arbitrary 'signs', since this would appear as an irreducible contradiction in itself." (2001).

Here, I take my cue from Gentilucci *et al.* (2001), who suggest, as a result of their experimental work, that hand gestures may have been transformed into articulatory gestures by means of multiple motor commands to hand and mouth. The authors also hypothesize that open-close hand and mouth movements are strictly synchronized by means of brain-mediated, somatotopically mapped circuits, since grasping an object with the hand appears to influence mouth opening, and vice-versa. They go on to speculate, following Armstrong *et al.* (1995), that speech has evolved from a communication system based on hand gestures, a stance echoed by Corballis (2003), who argues that human language emerged from manual gestures rather than from primate calls. The semiogenetic theory of the conditions of emergence and evolution of the language sign

(henceforth SGT) sketched out in Philps (2000) suggests that if open-close hand gestures were indeed transformed into open-close articulatory gestures, then the latter could have served to *refer back to* these hand gestures deictically, and to *stand for* them symbolically by means of an unconscious, neurophysiologically grounded, cognitive body-naming strategy. The processes involved in this putative strategy appear to include self-reference (Philps 2000: 217), vocomimesis (Donald 2001: 291) and conceptual mapping (Lakoff 2003: 246).

2. The SGT and the concept of 'sublexical marker'

The SGT, constrained empirically by a corpus compiled from Proto-Indo-European (PIE) and Indo-European languages, postulates that the language sign was originally configured vocomimetically during a period in the evolution of *H. sapiens* when the oral apparatus, originally used for purposes of nutrition, respiration and visuofacial communication, began to be employed additionally for articulatory purposes. One major assumption of this theory is that the initial conditions of a system largely determine its subsequent conditions, though not exclusively so. Moreover, whereas the linguistic sign is arbitrary by definition and by conception, the language sign is envisaged as having *become* arbitrary.

The theory developed initially from an analysis of those initial consonant clusters of English with recurrent form and 'meaning' called 'phonæsthemes' by Firth (1930: 50), e.g. *bl-*, *gr-*, *sl-*, and *sn-*, although these "frequently recurring sound-meaning pairings" (Bergen 2004: 290) were identified by grammarians as long ago as the 17th century (Wallis 1653). In view of the lack of any rigorous definition of phonæsthemes and of criteria for classifying words containing them, I applied a principle of submorphemic invariance to the heuristically set up semiological classes in which they are found, i.e. '*gr-* words', '*sn-* words', etc. This allows one to identify subsets of words attesting a given phonæstheme whose members display both semiological and notional invariance, e.g. nasality in the subset of '*sn-* words' that includes *sneeze*, *sniff* and *snore*, and prehension in the subset of '*gr-* words' that includes *grasp*, *grip* and *grope*. I call the word-initial cluster thus conceptualized a 'sublexical marker' (Philps 2003), defined as a submorphemic unit displaying semiological and notional invariance within the subset(s) of words of which it conditions the meaning(s). These markers are noted typographically between angled brackets (<sn->, <gr->, etc.).

A 'notion' is not to be equated with Saussure's *signifié*, since it is envisaged as a cognitive entity which may be defined as a bounded set of complex mental representations resulting from the mind's attempts to categorize experience, notably the formal, functional, and inter-relational properties of the latter. Seen in this light, notions (noted between slashes, e.g. /nasality/), may be analysed according to principled cognitive criteria into hierarchically organized, topologically constrained notional domains to which a metric may be applied.

Now there is structural evidence in PIE, notably root-final **-r-/*-l-* alternation that does not correlate with a change in 'meaning', as in **gal-* 'to call, shout'/**gar-* 'to call, cry' and **ghel-* 'to call'/**gher-* 'to call out', that **g-/*gh-*, which occupy the C_1 slot in the canonical PIE root structure C_1VC_2-, function as 'core invariants' (<*g->/<*gh->), and **-r-/*-l-*, consequentially, as variables (C_2). A 'core invariant' may be defined synchronically as the minimal invariant structural-notional unit in a given subset belonging to a pre-established class of words (e.g. **g-* in PIE '**g-* roots', or *gr-* in English '*gr-* words'). A diachronic definition must, however, account for the fact that this unit can be zero (e.g. in the Middle/Modern English 'phonosemantic doublet' *gnip* (obs.) / *ønip* 'to bite'). Moreover, one of the above roots (**gher-* 'to call out') furnishes English with the '*gr-* word' *greet* (< Germanic **grōtjan* < PIE **ghrēdh-* 'to call out'), while the '*gr-* word' *grope* may derive, via **ghreib-*, from (apparently unattested) **gher-* 'to grasp' (Mallory & Adams 1997: 564). Hence, in spite of the fact that *r-* forms part of the semiologically invariant segment *gr-* in English '*gr-* words', notably in that subset having meanings which refer to /prehension/ (*grab*, *grasp*, *grip*, etc.), it nevertheless appears to occupy the variable slot (C_2) in the class of PIE '**g-/*gh-* roots' from which some '*gr-* words' are derived.

There is also empirical evidence that a notional relation exists between the subset of '*gr-* words' including *grip*, *grope*, etc., and certain members of the semiological class of '*gVr(-)* words', e.g. *gird* (v.) 'to surround, encircle; to bind (a horse) with a saddle-girth'. This relation, which may be expressed by the function {<g(V)r(-)>, /prehension/}, can be traced back to PIE, since the '*gr-* word' *grope* and the '*gVr(-)* word' *gird* may derive, via a base II and a base I extension respectively, from the same PIE root, namely **gher-* 'to grasp' (oldest form **ĝher-*, Rix *et al.* 2001: 177). Now there is in PIE a homonymic root **gher-* 'to call out', one reflex of which is, as mentioned above, the English '*gr-* word' *greet* (v.). The relation in question may be expressed by the function {<g(V)r(-)>, /orality/}, in which <g(V)r(-)> is one possible actualization of the complex sublexical marker <g(V)C(-)>, as attested by, e.g., *gan* (n., slang) 'the mouth', *gape* (v.) 'to open the mouth wide, to shout', *gob* (n., slang) 'the mouth', etc. (*The Oxford English Dictionary*, 2nd edition, 1989, henceforth *OED*). If one examines the different actualizations of <g(V)C(-)> in PIE, a fairly consistent pattern of homonymic relations emerges, since this marker appears to condition the meanings of a number of apparently or clearly distinct roots referring to various subdomains of the superordinate domains /orality/ and /manuality/:

a) /manuality/: /scratching/ [*scratch* (v.) 'To wound superficially by dragging the claws or finger-nails over the skin']: **gerbh-* 'to scratch', **gher-* 'to scrape, scratch' and **ghrebh-* 'to dig, bury, scratch', /rubbing/ [*rub* (v.) 'To subject (a surface or substance to the action of something (as the hand, a cloth, etc.) moving over it']: **ĝhreh$_1$i-* 'to rub' and **ghreh$_1$u-* 'to rub, grind', an extension of **gher-* 'to scrape, scratch', and /grasping/ [*grasp* (v.) 'To make clutches with

the hand']: *ĝher- 'to grasp, enclose', *ghabh- 'to give, take, seize' (> OInd. gábhastin- 'hand', cf. OInd. hásta- 'hand' < *ĝhós-to-s < *ĝhés-r- 'hand') and *ghreib- 'to grip'. This marker occupies the C_1 slot in the original PIE root from which words for the 'hand' are derived, namely *ĝhes- (Markey 1984);

b) /orality/: /calling/ [*call* (v.) 'to shout, utter loudly, cry out, summon']: *gal- 'to call, shout', *gar- 'to call, cry', *gerh$_2$- 'to cry hoarsely', *ghel- 'to call', *gher- 'to call out', *ĝheu(h)- 'to call, invoke', etc., /yawning/: *ĝheh$_2$i- 'to yawn, gape', /swallowing/: *g^{w}elh$_1$- 'to swallow', *g^{w}erh$_3$- 'to swallow', and /biting/: *gh(e)n- 'to gnaw', *ĝ(y)euh$_x$- 'to chew, eat'. This marker occupies the C_1 slot in many roots whose derivatives denote mouth-related features in various IE languages, e.g. *ĝembh- 'tooth, nail', *ĝep(h)-/*ĝebh- 'jaw, mouth', and the compound *ghel-unā 'jaw' (Watkins 2000).

This analysis seems to confirm that the consonant occupying the C_1 slot in PIE roots, e.g. *g- in *gal-, *gh- in *ghel-, *ĝh- in *ĝher-, and *g^w- in *g^{w}elh$_1$-, functions as a core invariant, which may take the form of a voiced occlusive tectal ('occlusive' being "an older term for *plosive*", Trask 1996: 246), whether aspirated (*gh-), aspirated and palatalized (*ĝh-), labialized (*g^w-), or not (*g-).

3. From occlusive to occlusion

Analytical methods such as archaeological inference, lexico-cultural assessment and glottochronology tend to converge, in spite of their respective shortcomings, on a time-depth of some 6,000-8,000 years BP for the earliest form of PIE (Mallory & Adams 1997: 586). If one accepts this estimation on the one hand, and the possibility of reconstructing the sound-notion functions {<g->, /orality/} and {<g->, /manuality/} for PIE on the other, then one hypothesis that will require exploration is whether the depth to which this function can be reconstructed is an indication that it has always existed. Although no linguistic methodology permits us to do this, there does exist a theory known as articulatory phonology (Browman & Goldstein 1992), in which the basic units of phonological contrast are seen as *gestures*. Applied to the theory of articulatory invariance outlined here, it may help us, indirectly, to extend the above-mentioned time-depth quite substantially, since, even though this theory is formulated without specific reference to language evolution, it incorporates phonetic and kinetic parameters. Now according to McNeill (2005: 255-256), the emergence of a thought-language-hand link could have begun as long as five million years ago with the emergence of bipedalism in *Australopithicus*, and the selection of self-responsive mirror neurons some two million years ago with the advent of *H. habilis* and later *H. erectus*. McNeill further estimates that the whole process could have been completed around a hundred thousand years ago.

Assuming that the phonetic feature common to PIE *g- in *gal-, *gh- in *ghel-, *ĝh- in *ĝher-, and *g^w- in *g^{w}elh$_1$-, etc., i.e. [occlusive], implies that a

constriction/release has occurred at some point along the vocal tract, I contend that the manner feature which characterizes the occlusive realization of the core invariant <g-> in *gal-, etc., is the static equivalent in the place-manner classification system for consonants used in phonetics, of the dynamic, open-close gesture [occlusion] in the theory of articulatory phonology. Reanalysing occlusives as articulatory gestures of occlusion allows us to trace the invariant manner feature [occlusive] in its gestural guise as [occlusion] beyond PIE, as far back as theories of speech evolution such as MacNeilage's "Frame/Content Theory" will permit, without abandoning the guiding principle of invariance on which the SGT is based. According to MacNeilage, who bases his comments on empirical studies of early consonantal articulation and syllable-formation in infants' babbling, "[syllabic] frames may derive from cycles of mandibular oscillation present in humans from babbling onset, which are responsible for the open-close alternation." (1998: 499). By this view, human speech seen as a motor function emerged when the cyclical, open-close alternations characteristic of ingestive activities underwent a series of sequenced adaptations resulting in them being employed in content-modulated syllabic frames for purposes of visuofacial and phonatory communication.

Other scholars who have adopted a similar stance include Studdert-Kennedy & Goldstein, who suggest that the hominid protosyllable may have arisen from cyclical lowering and raising of the jaw for mastication, adding that "Such a protosyllable can be viewed as a gesture, that is, as constriction and release of one of the vocal organs, set in the context of an overall vocal tract posture and combined with phonatory action." (2003: 240). The authors also note that "a CV word can be produced by a single organ forming a constriction and release without any precise coordination of consonant and vowel gestures. By contrast, a CVC word requires precise inter-gestural coordination – either consonant gestures to vowel or consonant gestures to each other." (2003: 252). If the earliest languages only had *CV* syllable structure (Hurford 2003: 53), then the phonetic realizations of a marker such as <g-> would have been incorporated into a *CV* syllable. And significantly, Southern (1999: 152) suggests that fully linguistic *CVC* frames may have evolved from *CV* frames by consonantal augmentation at some early stage in the evolution of the protolanguage.

4. Conceptual projection

In the SGT, the open-close sounds produced by contact between the back of the tongue and the soft palate (a constriction-release gesture), with vibration of the vocal folds during the compression stage and cyclical jaw lowering-raising, are hypothesized to function as 'core invariants' of self-referential, goal-orientated language signs. The latter would enable the speaker — and the hearer, by what Gallese (2004) terms "intentional attunement" —, to 'name' not only the oral movements, functions and articulators such as the jaws involved in the

production of the sounds, but also, by conceptual projection, other symmetrical parts of the body such as the 'hands' that feature goal-orientated, open-close (or otherwise oscillatory) movements too, notably in the form of extension-flexion or abduction-adduction cycles, possibly accompanied by sonority (clicking, etc.).

Within Lakoff & Johnson's source-to-target mapping theory (2003: 252), this body-naming strategy may be seen as one of top-down intradomain conceptual projection. In the SGT, the mouth is taken to be the 'source domain' and the hands the 'target domain' of the projection on the assumption that the vocal organs and their anatomical environment can function not only self-referentially (Philps 2000: 230-231), but also as a structural template for denoting other parts of the body (Heine 1997: 134). This hypothesis implies that the process leading to the 'naming' of the open-close movements of the vocal organs, their different functions, and the organs themselves, is metonymically based, i.e. an open-close sound for the open-close movements and articulators involved. The process leading to the 'naming' of apparently synchronized open-close hand movements, and the hand itself, is however partly metonymic, i.e. an open-close sound for the open-close movement(s) of the hands (coupled with the movement for the effector in the case of the body part), and partly metaphorical, i.e. top-down projection of common topological properties, functions and relations such as protrusion, angularity, movement and prehension.

One PIE root with an initial, voiced occlusive tectal that furnishes reflexes attesting a process of mouth-to-hand projection, observable linguistically as polysemy, is **ghrendh-* 'to grind', derivatives of which possess both an 'oral' sense, e.g. in Mod. Eng. *grind* (v.): 'Denoting the action of teeth, or apparatus having the same function', and a 'manual' sense, e.g., in *to grind the coffee mill*: 'to imitate with the hand the action of grinding, by way of contempt' (*OED*). Two other PIE roots testify to a cognitive process of mouth-to-knee projection, observable linguistically as homonymy, namely **ĝenu-* 'jaw, chin' and **ĝenu-* 'knee' (> Mod. Eng. *knee*). Also implicated is the hypothetical base **g(e)n-* 'to compress into a ball', since it furnishes a subset of English '*kn-* words' other than *knee* with meanings referring to /articulated body parts/, e.g. *knop* (n., obs.) 'The rounded protuberance formed by the front of the knee or the elbow-joint', *knuckle* (n.) 'the end of a bone at a joint, which forms a more or less rounded protuberance when the joint is bent, as in the knee, elbow, and vertebral joints…', and *knead* (v.) 'to work and press into a mass (as if) with the hands'.

5. Conclusions

If the hypothesis of a strict relation between speech control and hand control put forward by Gentilucci and co-workers is correct, then it is conceivable that voiced occlusive sounds produced by open-close movements of the mouth synchronized with open-close movements of the hand(s) could, once

morphogenetically augmented by syllabification and differential consonantal accretion (e.g. *G-* > *GV-* > *GVC-*, as in PIE **g-* > **ga-* > **gal-*), have served to 'name' not only open-close mouth movements such as 'gnawing' and the articulators involved, but also coordinated hand movements such as 'grasping' and the effectors involved. The conventionalized signs thus formed would have meanings that, being of bodily origin, would be common to the entire speech community concerned. Once integrated into a linguistic system and subjected to its constraints, the 'body words' thus configured may have undergone desemanticization (or 'body bleaching') and grammaticalization. This is attested by English spatial grams such as *aback, abreast, afoot, a hand* (phr., obs.), *ahead, aknee* (obs.), etc., an indication that the invariant, topological relations which characterize the body, transposed into grammar via the lexicon, may provide a structural template for certain types of syntactic relations.

To sum up, the proposed body-naming strategy appears to be grounded in the brain's apparent capacity to dynamically and empathically simulate the cyclical, articular, goal-orientated, open-close movements of the hands by means of synchronized cyclical, articulatory, goal-orientated, open-close movements of the jaws. This hypothesis is accredited by recent research on the reciprocal influence between hand and mouth movements (e.g. Gentilucci *et al.* 2001), mirror neurons (e.g. Rizzolatti & Craighero 2004) and embodied simulation (e.g. Feldman & Narayanan 2004, Gallese & Lakoff 2005). Further exploration of the relevant language data, and a deeper understanding of the embodied processes of conceptual projection and simulation, may well set us on the road to attaining the neurolinguistic goal contained in the suggestion by Buoiano quoted earlier.

References

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). *Gesture and the nature of language*. Cambridge, UK: Cambridge University Press.

Bergen, B. K. (2004). The psychological reality of phonæsthemes. *Language, 80, 2*, 290-311.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica, 49*, 155-180.

Buoiano, G. C. (2001). http://fccl.ksu.ru/winter.2001/discuss.htm.

Corballis, M. (2003). From hand to mouth: the gestural origins of language. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 201-218). Oxford: Oxford University Press.

Donald, M. (2001). *A mind so rare. The evolution of human consciousness*. New York & London: W.W. Norton.

Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language, 89*, 385-392.

Firth, J. R. (1930). *Speech*. London: Ernest Benn.

Gallese, V. (2004). Intentional attunement. The Mirror Neuron system and its role in interpersonal relations. *Interdisciplines* (European Science Foundation), http://www.interdisciplines.org/mirror/papers/1.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22 (3/4)*, 455-479.

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S., 2001. Grasp with hand and mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology, 86*, 1685-1699.

Heine, B. (1997). *Cognitive foundations of grammar*. New York: Oxford University Press.

Hurford, J. R. (2003). *The language mosaic and its evolution*. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 38-57). Oxford: Oxford University Press.

Lakoff, G., & Johnson, M. (2003 [1980]). *Metaphors we live by*. Chicago & London: University of Chicago Press.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*, 499-546.

Mallory, J. P., & Adams, D. Q. (1997). *Encyclopedia of Indo-European culture*. Chicago & London: Fitzroy Dearborn.

Markey, Th. L. (1984). The grammaticalization and institutionalization of Indo-European *hand*. *Journal of Indo-European Studies, 12*, 261-292.

McNeill, D. (2005). *Gesture and thought*. Chicago & London: University of Chicago Press.

Nichols, J. (1998). The origin and dispersal of languages: linguistic evidence. In N. G. Jablonski & L. C. Aiello (Eds.), *The origin and diversification of language* (pp. 127-170). San Francisco: California Academy of Sciences.

Philps, D. (2000). Le sens retrouvé ? De la nomination de certaines parties du corps: le témoignage des marqueurs sub-lexicaux de l'anglais en <CN->. *Anglophonia / Sigma, 8*, 207-232.

Philps, D. (2003). *S-* incrémentiel et régénération submorphémique en anglais. *Bulletin de la Société de Linguistique de Paris, XCVIII, 1*, 163-196.

Rix, H. (Ed.), Kümmel, M., Zehnder Th., Lipp, R., & Schirmer B. (2001). *Lexikon der indogermanischen Verben*. Wiesbaden: Reichert.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Southern, M. R. V. (1999). *Sub-grammatical survival: Indo-European s-mobile and its regeneration in Germanic*. Washington: Journal of Indo-European Studies 34 / Institute for the Study of Man.

Studdert-Kennedy, M., & Goldstein L. (2003). Launching language: the gestural origin of discrete infinity. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 235-254). Oxford: Oxford University Press.

Trask, R.L. (1996). *A dictionary of phonetics and phonology*. London: Routledge.

Wallis, J. (1653). *Grammatica linguæ anglicanæ*. R.C. Alston (Ed.), Reprint 142 (1969). Menston: The Scolar Press.

Watkins, C. (2000). *The American Heritage dictionary of Indo-European roots*, 2nd ed. Boston: Houghton Mifflin.

DIFFUSION OF GENES AND LANGUAGES IN HUMAN EVOLUTION

ALBERTO PIAZZA

Dipartimento di Genetica, Biologia e Biochimica,
Università di Torino,
via Santena 19, 10126 Torino, Italy
alberto.piazza@unito.it

LUIGI CAVALLI SFORZA

Department of Genetics,
Stanford University,
Stanford, CA 94305, USA
cavalli@stanford.edu

In a study by Cavalli-Sforza et al. (1988), the spread of anatomically modern man was reconstructed on the basis of genetic and linguistic pieces of evidence: the main conclusion was that these two approaches reflect a common underlying history, the history of our past still frozen in the genes of modern populations. The expression genetic history' was introduced (Piazza et al. 1988) to point out that if today we find many genes showing the same geographical patterns in terms of their frequencies, this may be due to the common history of our species. A deeper exploration of the whole problem can be found in Cavalli-Sforza et al. (1994). In the following, some specific cases of structural analogies between linguistic and genetic geographical patterns will be explored that supply further and more updated information. It is important to emphasize at the outset that evidence for coevolution of genes and languages in human populations does not suggest by itself that some genes of our species determine the way we speak; this coevolution may simply be due to a common mode of transmission and mutation of genetic and linguistic units of information and common constraints of demographic factors.

1. The Genetic Analysis of a Linguistic Isolate: The Basques

The case of the Basques, a European population living in the area of the Pyrenees on the border of Spain and France who still speak a non-Indo-European language, is paradigmatic. What are the genetic relations between the Basques

and their surrounding modern populations, all of whom are Indo-European speakers?

Almost half a century ago it was suggested (Bosch-Gimpera 1943) that the Basques are the descendants of the populations who lived in Western Europe during the late Paleolithic period. Their withdrawal to the area of the Pyrenees, probably caused by different waves of invasion, left the Basques untouched by the Eastern European invasions of the Iron Age. In their study of the geographic distribution of Rh blood groups, Chalmers et al. (1948) pointed out that the Rh negative allele, which is found almost exclusively in Europe, has its highest frequency among the Basques.

Chalmers et al. hypothesized that modern Basques may consist of a Palaeolithic population with an extremely high Rh negative frequency, who later mixed with people from the Mediterranean area. In more recent times genetic analyses have produced the following conclusions:

(a) Mitochondrial and Y-chromosome DNA polymorphisms support the idea that the Basques are genetically different from the other modern European populations (Richards et al., 2000, 2002; Semino et al. 2000).

(b) Mitochondrial and Y-chromosome DNA polymorphisms support the idea that the Basques are the descendants of a Palaeolithic population (Richards et al., 2000, 2002; Semino et al. 2000). The main haplogroups contributing to the European mitochondrial geography are H, pre-V, and U5. Haplogroup H is the most frequent haplogroup in both Europe and the Near East but occurs at frequencies of only 25% 30% in the Near East and the Caucasus, whereas the frequency is generally 50% in European populations and reaches a maximum of 60% in the Basque country. The age ranges of the mitochondrial founders of these lines are mostly palaeolithict: specifically the age ranges of the mitochondrial haplogroup V which is found at the highest frequency among the Basques and the Saami are pre-neolithic. In agreement with the suggestion proposed to explain the distribution of mtDNA haplogroup V (Torroni et al. 1998), the distributions of Y chromosome groups R* and R1a have been interpreted by Semino et al. (2000) to be the result of postglacial expansions from refugia within Europe. European mtDNA estimates the Neolithic component in the Basques to be the lowest for any region in Europe. Although the criteria used to identify Near Eastern founder types are somewhat heuristic and involve many assumptions, the relative number of types in different European populations should still be informative, and the Basque component, estimated at 7%, clearly lies outside the distribution for the rest

of Europe, estimated to range between 9% and 21% (Richards et al. , 2000).

(c) The linguistic hypothesis originally put forward by Trombetti (1926) that Basques share a common ancestry with the modern Caucasian speaking people living in the northern Caucasus (see Ruhlen 1991) to form the Dene-Caucasian linguistic macrofamily according to Greenberg is in agreement with some genetic evidence: Wilson et al. (2001) report that the paternal ancestors of modern Basques could have shared a common genetic origin with Celtic speaking populations. In fact, the Y chromosome complements of Basque- and Celtic-speaking populations are strikingly similar. The similarity and homogeneity of the Basque, Welsh and Irish samples suggest one of two explanations: (i) pre-agricultural European Y chromosomes were homogeneous or (ii) there was a specific connection between the Basques, the pre-Anglo-Saxon British, and the Irish. With regard to the latter hypothesis, it is interesting that a northward expansion from a glacial refugium in Iberia has been postulated from the diffusion of Magdalenian industries (Otte et al., 1990) and patterns of Y-chromosome (Semino et al., 2000) and mtDNA variation (Torroni et al., 1998). More detailed investigation of the genetic diversity present in and around Europe may allow these hypotheses to be distinguished.

2. Coevolution of Genes and Languages: The Origin of Indo-European

Barbujani and Sokal (1990) found a correlation between linguistic and genetic boundaries in Europe. In the majority of cases (22 out of 33) there were also physical barriers that may have caused both genetic and linguistic boundaries. In nine cases there were only linguistic and genetic boundaries but not physical ones: three of them (northern Finland vs. Sweden, Finland vs. the Kola peninsula, Hungary vs. Austria) separate Uralic from Indo-European languages. It remains to be determined whether in these cases linguistic boundaries have generated or enhanced genetic boundaries, or if both are the consequence of political, cultural, and social boundaries that have played a role similar to that of physical barriers.

The problem of the origin of the Indo-European linguistic family and of the people speaking its languages has roused much more interest over the last years than in earlier times partly owing to the book by Renfrew (1987), who suggested that farmers, beginning to spread from Anatolia around 9,000 years ago, spoke Indo-European languages. His hypothesis was based on the suggestion originally

put forward by Ammerman and Cavalli-Sforza (1984) that the spread of Neolithic farming from the Fertile Crescent was due to the spread of the farmers themselves and not only of the farming technology, and on the consideration that migrating people retain their language, if at all possible. Renfrew's hypothesis was criticized by most Indo-European linguists (for a review, see Mallory 1989, Lehmann 1993: 283-8) and did not fare well when contrasted with earlier hypotheses, now identified with the name of another archaeologist, Marjia Gimbutas (1985), that Indo-Europeans migrated to Europe from the Pontic steppe area of south Russia from Dniepr to the Volga (which she called 'Kurgan' from the Russian name of mounds covering the graves), beginning with the early Bronze Age, that is, around 5,500 years ago.

Genetic data cannot give strong evidence on dates of migration, especially since the 'Kurgan' area, one of the largest pre-historic complexes in Europe, probably remained very active in generating population expansions for a long time after the Bronze Age. In that area we find at c. 6,000 years ago the Sredni-Stog culture, later (5,500-4,500 years ago) the Yamnaya cultures (formerly called pit-grave cultures) which stretched from the Southern Bug River over the Ural River and which dates from 5,600 to 4,200 years ago. From about 5,000 years ago we begin to find evidence for the presence in this culture of two and four-wheeled wagons (Anthony 1995).

Genetic data on European populations using blood typing (Piazza et al. 1995) and Y-chromosome DNA markers (Semino et al. 2000) have strongly supported a centre of radiation in the Ukraine. It has been suggested (Cavalli-Sforza et al. 1994, Piazza et al. 1995) that the hypotheses of Renfrew and Gimbutas should not be treated as mutually exclusive; they may be compatible, as Schrader anticipated as long ago as 1890: 'the Indo-Europeans practiced agriculture at a site between the Dniepr and the Danube where the agricultural language of the European branch was developed' (quoted from Lehmann 1993, p. 279). The settling of the steppe by Neolithic farmers must have occurred after the beginning of their migration from Anatolia, and if the expansions began at 9,500 years ago from Anatolia and at 6,000 years ago from the Yamnaya culture region, then a 3,500-year period elapsed during their migration to the Volga-Don region from Anatolia, probably through the Balkans. There a completely new, mostly pastoral culture developed under the stimulus of an environment unfavourable to standard agriculture, but offering new attractive possibilities. Our hypothesis is, therefore, that Indo-European languages derived from a secondary expansion from the Yamnaya culture region after the Neolithic

farmers, possibly coming from Anatolia and settled there, developing pastoral nomadism.

A new treatment of the problem has been given in a still unpublished analysis (Piazza et al., but see Cavalli-Sforza, 2000 where main results are anticipated) of a set of lexical data (200 words) in 63 Indo-European languages published by Dyen et al. (1992). From a linguistic distance matrix whose elements are the fraction of words with the same lexical root for any pair of languages and its transformation to make the matrix elements proportional to time of differentiation, we were able to reconstruct a linguistic tree. The root of the tree separates Albanians from the others, with a reproducibility rate (the error in reconstructing the tree) of 71 percent. The next oldest branch is Armenian. The simplest interpretation is that the language of the first migrant Anatolian farmers survives today in two direct descendants, Albanian and Armenian, which diverged from the oldest pre-Indo-European languages in different directions but remained relatively close to the point of origin.

If we give to the first split the time depth of the beginning of the expansion of the pre-Indo-European Anatolian farmers, about 9,000 years ago, we can then calculate that the origin of the European branch dates to about 6,000 years ago. The four major branches (pre-Celtic, pre-Balto-Slavic, pre-Italic, pre- Germanic) may correspond to some extent to different migratory waves, but archaeological dating is too scanty to provide unambiguous associations. It is reasonable to suggest that a first migration corresponds to the first branch, the pre-Celts (6,000 years ago, according to the tree), who settled first and went further west. Their only linguistic remnants are still alive today at the extreme of their original range. They profited from being among the first to develop an Iron Age culture, and were able to develop a wide community that spoke their language. Before Roman rule they spread to half of Europe, extending from Spain to France, most of the British Isles, northern Italy, and central Europe.

Very recently Gray and Atkinson (2003) have analyzed the same data set. They generated a tree of 87 languages which can be compared with our tree. We eliminated a small number of modern languages of the Dyen et al. set (1992), and Gray and Atkinson added interesting information on three extinct languages, Hittite and Tocharian A and Tocharian B, which we did not include. Their inclusion may have the advantage of providing some support for the root, but the noticeable shortening of the Hittite branch in their tree introduces some doubt on its usefulness. We believe it is worth discussing the differences between the two trees in some detail, because they are relevant to the problem of Indo-European origins and also to the general problem of evolutionary tree analysis.

Both approaches used for inferring the tree use information on the variation of evolutionary rates of different words. This is essential but very rarely done, because it affects strongly the shape of the curve formed by the rate of cognate retention rates (C) versus separation times t, causing a serious underestimation of longer times when compared with the standard glottochronological approach (Swadesh, 1952), that assumes a proportionality of log C and t. They use a method that assumes a normal distribution of log C of the retention rate of individual words and estimate it directly from the data, while in our analysis we estimate it using another source of information: the number of the different roots used to express the same meaning of each word.There remain a few differences between the trees, and it is worth considering them in detail. Gray and Atkinson seem to agree with us that there could be two origins of Indo-European languages, the first in coincidence with the origin of agriculture as suggested by Renfrew (1987), to be located in the Middle East or Anatolia, and a later one in the Ukraine, as suggested by Gimbutas (1985). The oldest languages, Armenian, Albanian and Greek, are among the oldest in both trees, but there is some disagreement in the relevant dichotomies. These are, however, those that have the highest errors in both trees, as shown by the percentage of agreement among repetitions of the analysis.

The other discrepancy is the dichotomy of Celtic, which in our tree is the oldest of the European subfamilies, while in theirs the oldest is Balto-Slavic. Our bootstrap value is higher than in their tree, indicating our method has smaller error in this part of the tree. There is information from other disciplines that supports our tree for both discrepancies.

If history can support some separation dates, though very weakly, geography may again be of help. Albanian is weakly related to Indic Iranian, while in our tree it is nearest to the root, closest to Armenian and Greek, in agreement with geography. Given the long distance between Albania and south Asia, and the local tree uncertainty it may be better to make the first dichotomy of the tree as a branch leading to a trichotomy of Albania, Greece and Armenia, corresponding with what remains of the first spread of farmers from Anatolia, and another branch leading to all the rest, reflecting later farmers expansions starting from the Ukraine, that gave rise to an early split into the Indic-Iranian branch going east and south, and the European branch, with the splitting sequence in time Celtic/Italic-Germanic/ Balto-Slavic. Making the Celtic branch the eldest is in agreement with other information : 1) Celtic languages are believed to have been spoken in Austria, Switzerland and northern Italy by the La Tene culture at least in the early part of the third millennium BC ; 2) in Julius Caesar's time

Celtic languages were spoken in France and Great Britain, while Germanic languages were spoken east of the Rhine; the later spread northwards and westwards of Germanic languages and southwards and westwards of Italic languages confined Celtic languages to the most peripheral parts of the British Isles, with Brittany speaking Celtic because of a secondary migration from the British Isles at the time of the Anglo-Saxon invasion, in the V-VI century AD. A remarkable help from weavings: the La Tène culture used Scottish style tartans , which were found over 3000 years ago also in the clothes of the mommies of west China. It is not entirely clear but these people may have spoken Tocharian in later times.

From a methodological points of view it is clear the retention rates of the Indo-European core vocabulary of 200 meanings considered in the analysis not only are heterogeneous but also fit to a bimodal gamma distribution and this adds further uncertainty to the dates associated to the major branchings in the tree.

From a general point of view it is of some interest to explore how the linguistic classification correlates with genetic data. Poloni et al. (1997) showed, for the Y chromosome, an important level of population genetics structure among human populations, mainly due to genetic differences among distinct linguistic groups of populations. A multivariate analysis based on genetic distances between populations shows that human population structure inferred from the Y chromosome corresponds broadly to language families ($r = .567$, $P < .001$), in agreement with autosomal and mitochondrial data. Times of divergence of linguistic families, estimated from their internal level of genetic differentiation, are fairly concordant with current archaeological and linguistic hypotheses. Variability of the p49a,f/TaqI Y polymorphic marker is also significantly correlated with the geographic location of the populations ($r = .613$, $P < .001$), reflecting the fact that distinct linguistic groups generally also occupy distinct geographic areas. Comparison of Y-chromosome and mtDNA polymorphisms in a restricted set of populations shows a globally high level of congruence, but it also allows identification of unequal maternal and paternal contributions to the gene pool of several populations.

3. Towards a Global Perspective

More than 5,000 languages are spoken today in the world, and it does not take a linguist to recognize that some languages are more closely related than others due to history. The official origin of historical linguistics can be dated to 1786, when the English judge Sir William Jones advanced the idea that Sanskrit, a

classical language in India, Greek, Latin, and possibly Celtic and Gothic (the ancestor of Germanic languages) shared a common origin. These old languages were the first members of a family of languages that would become known as the 'Indo-European' family (or 'phylum'). As Indo-European is the earliest and best studied linguistic family, coevolution of genes and languages has been documented. Since the eighteenth century, however, many other linguistic families or superfamilies have been recognized. The most complete classification on a world basis was proposed by Ruhlen (1994) on the basis of Greenberg's published and unpublished writings: he lists 12 linguistic families (Khoisan, Niger-Kordofanian, Nilo-Saharian, Afro-Asiatic, Dravidian, Kartvelian, Euroasiatic, Dene-Caucasian, Austric, Indo-Pacific, Australian, Amerind).

The reconstruction of the relationships above the family level is hotly debated among historical linguists who have yet to agree on the existence of a single tree linking all the existing language families, that is on the possible differentiation of modern languages from a single ancestor language. Even unification at a lower level such as that of the (pre-Columbian) American languages proposed by Greenberg (1987), who grouped them into just three macro-families (Eskimo-Aleut, Na-Dene, and Amerindian), has been strongly opposed by the majority of American linguists. Interestingly, Greenberg's proposal seems to agree with the analysis of genetic markers in extant Native Americans (Cavalli-Sforza et al. 1994) and these three families seem to identify three major migrations suggested by archaeological data. Amerindian speakers appear to have come first (between 30,000 and 15,000 years ago according to genetic data), followed by Na-Dene speakers and finally Eskimo-Aleut (both in a period between 15,000 and 10,000 years ago). It must be said, however, that at a finer level of classification contemporary Amerindian speakers show high genetic variability, and this is not easy to reconcile with linguistic taxonomy.

Even without an agreed genealogy of the linguistic families covering all tongues spoken today, it is relevant to note the impressive one-to-one correspondence of the genetic phylogeny of the world populations with the classification into the 12 large linguistic families listed above (Cavalli-Sforza et al. 1988). This correspondence is expected because there are important similarities between the evolution of genes and languages. In either case: (a) a change which first appears in a single individual can subsequently spread throughout the entire population (for genes they are called mutations; they are rare, are passed from one generation to the next and can, over many generations, eventually replace the ancestral type; linguistic innovations are much more frequent and can also pass between unrelated individuals); and (b) the dynamics

of change is affected by the same demographic pressures, isolation, and migration. Two isolated populations differentiate both genetically and linguistically because isolation, which could result from geographic, ecological, or social barriers, reduces the likelihood both of marriages and cultural exchanges and, as a common result, reciprocally isolated populations will evolve independently and gradually become different. Both genes and languages will drift apart regularly over time, the former slowly, the latter much more quickly.

In principle, therefore, the linguistic tree and the genetic tree of human populations should agree since they reflect the same history of population splitting and subsequent independent evolution. The different rate of change, however, is a major source of divergence: one language can be replaced by another in a relatively short time. In Europe, for example, Hungarian is spoken in a land surrounded by Indo- European speakers but it belongs to the Finno-Ugric subdivision of Uralic. At the end of the ninth century AD, the nomadic Magyars left their land in Russia and invaded Hungary. The number of conquerors was probably less than 30 percent of the conquered population so that their genetic contribution was limited, but they imposed their language on the local Romance-speaking population. Today all Hungarians speak a Uralic language, but barely 10 percent of their genes can be attributed to the Uralic conquerors.

Generally it is intuitive that the total substitution of one language for another occurs more easily under the pressure of a strong political power of the newcomers, as witnessed in the Americas. The case of Basques, on the other hand, shows that separate languages spoken in nearby countries can remain relatively unaffected for thousands of years, even when their genes experience a partial substitution. It is remarkable that, despite the above sources of confusion, the correlation between genes and languages has been maintained through the centuries until today and is still statistically significant. The ties between biology and linguistics were already evident since the times of Darwin, who in chapter XIV of his The Origin of Species wrote:

> "If we possessed a perfect pedigree of the mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world; and if all extinct languages, and all intermediate and slowly changing dialects, were to be included, such an arrangement would be the only possible one. Yet it might be that some ancient language had altered very little and had given rise to few new languages, whilst others had altered much owing to the spreading, isolation, and state of civilization of the several co-descended races, and had thus given rise to many new dialects and languages. The various degrees of difference between the languages of the same stock, would have to be

> expressed by groups subordinate to groups; but the proper or even the only possible arrangement would still be genealogical; and this would be strictly natural, as it would connect together all languages, extinct and recent, by the closest affinities, and would give the filiation and origin of each tongue."

The increasing resolving power of modern genetic data makes it possible to follow Darwin and to use the genetic phylogeny of our species to infer the earliest branches of a hypothetical linguistic tree. The most comprehensive genetic phylogeny reconstructed in Cavalli-Sforza et al. (1988) was used by Ruhlen (1994) to draw the tree of origin of human languages (some reference dates from genetic and archaeological evidence have been added). The oldest linguistic families must be African: Khoisan is probably the oldest and Afro-Asiatic the most recent, while Niger-Kordofanian and Nilo-Saharian believed by some linguists to descend from an ancestor tongue, and the Congo- Saharan, were probably spoken at an intermediate time. A more exhaustive discussion of this hypothetical tree can be found in Cavalli-Sforza (2000). As the genetic data improves with the inclusion of more representatives from those geographical areas of the world where the sampling is still scanty, the tree will be more complex but it is likely that its main features will remain unchanged.

In conclusion, our present genome keeps the record of its past evolution with an impressive richness of detail that is also reflected by our languages. Genes and languages contribute to the understanding of human history by highlighting human diversity; both are instrumental in giving some of the silent voices of our past a chance to be heard.

References

Ammerman, A.J., Cavalli-Sforza, L.L (1984). *Neolithic Transition and the Genetics of Populations in Europe*. Princeton University press, Princeton, NJ

Anthony, D.W. (1995). Horse, wagon & chariot: Indo-European languages and archaeology. Antiquity 69: 554-65

Barbujani, G., Sokal, R.R. (1990). Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences* 87: 1816-9

Bosch-Gimpera, A. (1943). El problema de los origines vascos. *Eusko-Jakintza* 3: 39

Capelli, C., Redhead, N., Abernethy, J.K., Gratrix, F., Wilson, J.F., Moen, T., Hervig, T., Richards, M., Stumpf, M.P., Underhill, P.A., Bradshaw, P., Shaha, A., Thomas, M.G., Bradman, N., Goldstein, D.B. (2003). A Y chromosome census of the British Isles. *Current Biology*, **13,** 979 - 984.

Cavalli-Sforza, L.L. (2000). *Genes, Peoples, and Languages*. North Point Press, New York

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ

Cavalli-Sforza, L.L., Piazza, A., Menozzi, P., Mountain, J. (1988). Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences* 85: 6002-6

Chalmers, J.N.M., Ikin, E.W., Mourant, A.E. (1948). Basque blood groups. *Nature* 162: 27

Dyen, I., J.B Kruskal & P.Black. (1992). An Indoeuropean classification: a lexicostatistical experiment, *Transactions of the American Philosophical Society* 82: Part 5. Philadelphia, American Philosophical Society.

Gimbutas, M. (1985) Primary and secondary homeland of the Indo-Europeans. *Journal of Indo-European Studies* 13: 185-202

Gray, R.D., Atkinson, Q.D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426, 435-439.

Greenberg, J.H. (1987). *Language in the Americas*. Stanford University Press, Stanford, CA

Lehmann, W.P. (1993). *Theoretical Bases of Indo-European Linguistics*. Routledge, London

Mallory, J.P. (1989). *In Search of the Indo-Europeans: Language, archaeology and myth*. Thames and Hudson, London

Otte, M. Soffer, O., Gamble, C., (eds.) (1990). In *The World at 18,000 BP*, Unwin Hyman, London.

Piazza, A., Cappello, N., Olivetti, E., Rendine, S. (1988). A genetic history of Italy. *Annals of Human Genetics* 52: 203-13

Piazza, A., Rendine, S., Minch, E., Menozzi, P., Mountain, J., Cavalli-Sforza, L.L. (1995). Genetics and the origin of the European languages. *Proceedings of the National Academy of Sciences* 92: 5836-40

Piazza, A., Minch, E., Cavalli-Sforza, L.L. (in preparation). The Indo-Europeans: Linguistic tree and genetic relationships. Manuscript

Poloni, E.S., Semino, O., Passarino, G., Santachiara-Benerecetti, A.S., Dupanloup, I., Langaney, A., Excoffier, L. (1997). Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am J Hum Genet.* 61, 1015-35.

Renfrew, C. (1987). *Archaeology and Language: The Puzzle of Indo-European Origins*. Cambridge University Press, New York

Richards, M., Macaulay, V., Hickey, E., Vega, E., Syke,s B., Guida, V., Rengo, C., Sellitto, D., Cruciani,. F, Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Calò, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Oppenheim, A., Novelletto, A., Nùrby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A., Bandelt, H.-J. (2000). Tracing European founder lineages in the near eastern mtDNA pool. *The American Journal of Human Genetics* 67, 1251-76.

Richards, M., Macaulay, V., Torroni, A., Bandelt, H.-G. (2002). In Search of Geographical Patterns in European Mitochondrial DNA. *Am. J. Hum. Genet.* 71,1168-1174.

Rosser, Z.H., Zerjal, T., Hurles, M.E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., et al (2000). Y-Chromosomal Diversity in Europe Is Clinal and Influenced

Primarily by Geography, Rather than by Language. *The American Journal of Human Genetics* 67, 1526-1543.

Ruhlen, M. (1994). *On the Origin of Languages: Studies in Linguistic Taxonomy.* Stanford University Press, Stanford, CA

Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A.S., Cavalli-Sforza, L.L., Underhill, P.A. (2000). The genetic legacy of paleolithic Homo sapiens sapiens in extant Europeans: A Y chromosome perspective. *Science* 290: 1155-1159

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96:452-463.

Swadesh, M. (1955). Towards greater accuracy in lexicostatistical dating. *International Journal of American Linguistics* 21:121-137.

Torroni, A., Bandelt, H.-J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savantaus, M.-L., Bonné-Tamir, B., Scozzari, R. (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137-1152.

Trombetti, A. (1926). *Le origini della lingua Basca.* Bologna, Italy

Wilson, J.F., Weiss, D.A., Richards, M., Thomas, M.G., Bradman, N., Goldstein, D.B. (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci USA* 98:5078–5083.

DIFFERENCES AND SIMILARTIES BETWEEN THE NATURAL GESTURAL COMMUNICATION OF THE GREAT APES AND HUMAN CHILDREN

SIMONE PIKA
School of Psychology, University of St. Andrews, St. Andrews, Fife, KY16 9JP, Scotland

KATJA LIEBAL
Department of Psychology, University of Portsmouth, King Henry Building, King Henry 1st Street, Portsmouth PO1 2DY, United Kingdom

The majority of studies on animal communication provide evidence that gestural signaling plays an important role in the communication of nonhuman primates and resembles that of pre-linguistic and just-linguistic human infants in some important ways. However, ape gestures also differ from the gestures of human infants in some important ways as well, and these differences might provide crucial clues for answering the question of how human language –at least in its cognitive and social-cognitive aspects- evolved from the gestural communication of our ape-like ancestors. The present manuscript summarizes and compares recent studies on the gestural signaling of the great apes (*Gorilla gorilla, Pan paniscus, Pan troglodytes, Pongo pygmaeus*) to enable a comparison with gestures in children. We focused on the three following aspects: 1) nature of gestures, 2) intentional use of gestures, 3) and learning of gestures. Our results show, that apes have multifaceted gestural repertoires and use their gestures intentionally. Although some group-specific gestures seem to be acquired via a social learning process, the majority of gestures are learned via individual learning. Importantly, all of the intentional produced gestures share two important characteristics that make them crucially different from human deictic and symbolic gestures: 1) they are almost invariably used in dyadic contexts and 2) they are used exclusively for imperative purposes. Implications for these differences are discussed.

1. Introduction

One of the enduring questions is how spoken language, which is thought to be unique to humans, originated and evolved. One important way to address this question is to compare speech to the systems of vocal communication evolved in other animals, especially in non-human primates (hereafter primates) (e.g., Marler, 1977; Seyfarth, 1987; Snowdon, 1988; Zuberbühler, 2003).

The majority of studies investigated vocal communication and revealed that call morphology and call usage seem to have only limited flexibility (Liebermann, 1998; Corballis, 2002). However, recent data provided evidence that vervet monkeys use different alarm calls in association with different predators (leading to different escape responses in receivers) and therefore raised the possibility that some nonhuman species may, like humans, use vocalizations to make reference to outside entities (Cheney & Seyfarth, 1990). But it has

turned out since then that alarm calls of this type have arisen numerous times in evolution in species that also must organize different escape responses for different predators, including most prominently prairie dogs and domestic chickens (Owings & Morton, 1998). And importantly, there is currently no evidence that any species of ape has such referent specific alarm calls or any other vocalizations that appear to be referential (Cheney & Wrangham, 1987; however see Crockford & Boesch, 2003 for context specific calls). This implies that it is highly unlikely that alarm calls of monkeys could be the direct precursor of human language - unless at some point apes used similar calls and have now lost them.

Interestingly, gestural or ideographic communication systems have to some extent been mastered by human-reared great apes (e.g. Gardner, et al, 1989; Savage-Rumbaugh, et al, 1993). Though by no means 'language', these projects have shown intentional, referential use of numerous gestures and ideograms (Gardner, et al., 1989; Savage-Rumbaugh, 1986), accurate usage under double-blind conditions, and understanding of human speech. These findings support the hypothesis that the evolutionary roots of language might have evolved in the visual-gestural modality (e.g., Condillac, 1971; Hewes, 1976; Armstrong et al., 1995; Dunbar, 1996; Arbib, 2002). In addition, recent studies provide evidence that gestural signaling plays an important role in the natural communication of primates and resembles that of prelinguistic and just-linguistic human infants (Plooij, 1978, Tomasello et al., 1985). However, ape gestures also differ from the gestures of human infants in some important ways, and these differences might provide crucial clues for answering the question of how human language – at least in its cognitive and social-cognitive aspects- evolved from the gestural communication of our ape-like ancestors. The question thus arises: what is the nature of the gestural communication of nonhuman primates, and how do they relate to human gestures and language?

The present manuscript is based on observations of the communicative signaling of the four great apes species (*Gorilla gorilla, Pan paniscus, Pan troglodytes, Pongo pygmaeus*). To enable a qualitative comparison with gestures in children, we focused on the three following aspects: First we investigated the nature of gestures by examining whether they are used dyadic, triadic, imperative (used to get another individual to help in attaining a goal, cf. Bates, 1976) and/or declarative (used to draw another's attention to an object or entity merely for the sake of sharing attention, cf. Bates, 1976).

Second, we investigated if apes use their gestures intentionally, focusing on the key characteristics for intentional communication in children (Piaget, 1952; Bates, 1976; Bruner, 1981), –a) means-ends dissociation and b) special sensitivity to the social context.

A) Means-ends dissociation can be characterized by the flexible relation of signaling behavior and goal. An individual uses for instance a single

gesture for several goals (touch for nursing and riding) or different gestures for the same goal (*slap ground* and *bodybeat* for play).

B) Sensitivity to the social context: The sender performs a gesture toward a recipient for the purpose of communication. Evidence for specifically communicative intent includes the signaler's alternation of gaze between goal and recipient (Bates, 1979; observed in wild chimpanzees, Plooij, 1978), persistence to the goal, or adjustment to audience effects (Tomasello et al., 1997).

Our third goal concerned the learning of gestures by focusing on individual and group variability to distinguish between underlying social and individual learning processes. Following Tomasello and colleagues (Tomasello et al., 1994) similarities in the gestural repertoires within a group and group specific gestures would provide evidence for the existence of a social learning process, whereas individual differences that overshadow group differences (i.e., a lack of systematic group differences, idiosyncratic gestures) imply that an individual learning process is involved.

2. Methods

Two chimpanzee, two bonobo, two gorilla and two orangutan groups were observed in different European zoos. The communicative behavior of 46 subadult focal animals was videotaped for an average of 12.5hrs/individual (sampling rule: behavior sampling/focal animal sampling; recording rule: continuous recording). We analyzed an average of 1530 gestures per species.

3. Results

Gestural repertoire

Based on auditory, tactile and visual components we formed three signal categories: auditory gestures generate sound while performed, tactile gestures include physical contact with the recipient, and visual gestures generate a mainly visual component with no physical contact.

Bonobos: The bonobos used 20 different distinct gestures: one auditory (5%), eight tactile (40%) and eleven visual gestures (55%). On average each individual used 11 gestures.

Chimpanzees: The chimpanzees used 28 different distinct gestures: three auditory (11%), nine tactile (32%), and 16 visual gestures. On average each individual used 9.5 gestures.

Gorillas: Overall the gorillas performed 33 different distinct gestures: six auditory (18%), 11 tactile (33%) and 16 visual gestures (49%). On average each individual used 20 gestures.

Orangutans: The orangutans used 26 different distinct gestures (see figure 1): 12 tactile and 14 visual gestures. On average each individual used 16 gestures.

The majority of these gestures were dyadic and imperative. Exceptions to this pattern were the gestures *move, peer* (bonobos), *palm-up* (chimpanzees), *move, object shake, peer, straw wave* (gorillas), *hold hand in front of the mouth, offer arm with food pieces, offer food, present object, shake object* (orangutans). These gestures although imperative, were clearly triadic since they involved an outside entity (food, object), the sender and the receiver.

Intentional use of gestures

Means-ends dissociation: The bonobos used on average in every context approximately two (± 0.6) different gestures, the chimpanzees 3.2 (± 0.4), the gorillas 3.2 (± 1), and the orangutans 5.3 (±1.2) gestures. Concerning the use of gestures in different contexts, the bonobos utilized on average 2.7 (± 1.48) gestures in more than one context, the chimpanzees 1.3 (± 0.2), the gorillas 3.8 (± 2.6), and the orangutans 1.5 (± 0.9).

Sensitivity to the social context -adjustment to audience effects-: We found a significant difference between the use of tactile and visual gestures among all species based on a variation in the degree of visual attention of the recipient (Wilcoxon-test: $P < 0.05$, for further details see Liebal et al., 2004, Liebal et al., in review, Pika et al. 2003, Pika et al., 2005, Tomasello et al., 1994). There was no significant difference between the uses of auditory versus visual gestures and auditory versus tactile gestures. On average, the bonobos performed 79% (± 10) of their visual gestures to an attending recipient, the chimpanzees 87% (± 2), the gorillas 89% (± 12), and the orangutans 98.8% (± 2). However, tactile gestures were performed to an attending recipient in 50% (bonobos and chimpanzees, ± 10), 66% (gorillas, ± 13), and 67% (orangutans, ± 10.3).

Learning of gestures

Following Tomasello and colleagues (Tomasello et al., 1994) high levels of concordance of gestural repertoires within a group and group-specific gestures would provide evidence for the existence of a social learning process, whereas individual differences that overshadow group differences (i.e., a lack of systematic group differences, idiosyncratic gestures) imply that mainly an individual learning process is involved. To assess the degree of concordance in the performance of gestures between and within the two groups we used Cohen's Kappa statistics (see, Tomasello et al., 1997). The between and within-group Kappas of the bonobos (within-group Kappa: 0.5; between group Kappa: 0.45) and chimpanzees (within-group Kappa: 0.34; between group Kappa: 0.24) showed very low degrees of concordance (Altmann, 1991), the between and

within-group Kappas of the orangutans (within-group Kappa: 0.7; between group Kappa: 0.68) 'moderate' levels of agreement, and the between and within-group Kappas of the gorillas showed an 'excellent' strength of agreement (within-group Kappa: 0.8; between group Kappa: 0.72) (Altmann, 1991). All species showed similar degrees of concordances between and within-groups.

The bonobos and gorillas used three idiosyncratic gestures, the chimpanzees 13, and the orangutans two.

The bonobos and gorillas performed two group-specific gestures and the orangutans one. All group-specific gestures cannot be easily explained due to different physical conditions or different social settings.

4. Discussion

This manuscript aimed to provide a qualitative overview of the gestural communication of the great apes to enable a qualitative comparison with gestures in children. We focused on the three following aspects, 1) nature of gestures, 2) intentional use of gestures, and 3) major learning mechanism involved in the acquisition of gestures.

Overall, our results showed that apes have multifaceted gestural repertoires. The majority of these gestures were dyadic and imperative. However, some gestures to obtain food or to play with an object were used triadically.

Concerning the intentional use of gestures, all apes used their gestures flexibly, by utilizing one signal for several contexts and several signals for a single context. In addition, all four species adjusted the use of gestures to the attentional state of the recipient, preferentially performing visual gestures to an attending recipient. Therefore, we can conclude that apes communicate by using intentional acts identified through their flexible relation of signaling behavior and goal and the signaler's sensitivity to the social context.

Focusing on the learning of gestures, our data showed that the gorillas showed the highest level of concordances of gestural repertoires between and within-groups, the chimpanzees and bonobos the lowest. Furthermore, concordances in gestural repertoires between and within-groups did not differ significantly. In addition, all great ape species developed idiosyncratic gestures. Overall these findings support, based on our defined indicators for individual learning, the hypothesis that ontogenetic ritualization is the main learning process involved.

However, we found group-specific gestures in a group of bonobos, gorillas and orangutans. These findings imply that at least some gestures are acquired via a social learning process.

All of the intentional gestures used by apes therefore share two important characteristics that make them crucially different from human deictic and symbolic gestures: 1) They are mainly used in dyadic contexts and attract the attention of others to the self and, not triadically, to some outside entity. Human

infants in contrast gesture from their very first attempts in addition to dyadic gestures triadically, that is for persons to external entities (Carpenter et al., 1998). 2) Ape gestures seem to be exclusively used for imperative purposes to request actions from others. Human infants in contrast use gestures imperatively but also declaratively to direct the attention of others to an outside object or event, simply for the sake of sharing interest in it or commenting on it. Although the majority of differences are quantitative and not qualitative, the crucial findings is that apes don't use gestures to communicate about outside entities or comment on it. This propensity seems to be unique for human communication and might have been derived from the cognitive ability that enables humans to understand other persons as intentional agents with whom they may share experience (Tomasello, 1999).

References

Altmann, D. (1991). *Practical statistics for medical research.* CRC: Chapman and Hall.

Arbib, M. A. (2002). The mirror system, imitation, and the evolution of language. In K. Dautenhahn & C. L. Nehaniv (Eds.), *Imitation in animals and artifacts. Complex adaptive systems* (pp. 229-280). Cambridge, Masachusetts, USA: MIT Press.

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). *Gesture and the nature of language.* Cambridge: Cambridge University Press.

Bates, E. (1976). *Language and context the acquisition of pragmatics.* New York: Academic Press.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York: Academic Press.

Bruner, J. (1981). Intention in the structure of action and interaction. In L. Lipsitt (Ed.), *Advances in Infancy Research* (Vol. 1, pp. 41-56). New Jersey: Ablex, Norwood.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 255*, 176-179.

Cheney, D., & Wrangham, R. (1987). Predation. In B. Smuts & D. L. Cheney & R. M. Seyfarth & R. Wrangham & T. Struhsacker (Eds.), *Primate societies.* Chicago: University of Chicago Press.

Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world.* Chicago and London: University of Chicago Press.

Condillac, E. B. d. (1971). *An essay on the origin of human knowledge; being a supplement to Mr. Locke's Essay on the human understanding. A facism. reproduction of the translation of Thomas Nugent.* Gainesville: Scholars' facsimiles and reprints.

Corballis, M. C. (2002). *From hand to mouth, the origins of language.* Princeton, New Jersey: Princeton University Press.

Crockford, C., & Boesch, C. (2003). Context-specific calls in wild chimpanzees, *Pan troglodytes verus*: Analysis of barks. *Animal Behaviour, 66*(1), 115-125.

Dunbar, R. (1996). *Grooming, Gossip and the Evolution of Language.* London: Faber and Faber Ltd.

Gardner, R. A., Gardner, B., & Van Cantford, T. E. (1989). *Teaching sign language to chimpanzees*. Albany: State University of New York Press.

Hewes, G. W. (1976). The current status of the gestural theory of language origin. In S. Harnad & H. D. Steklis & J. Lancaster (Eds.), *Origins and evolution of language and speech* (pp. 482-504). New York: New York Academy of Sciences.

Liebal, K., Pika, S., & Tomasello, M. (2004). Social communication in siamangs (Symphalangus Syndactulus): Use of gestures and facial expression. *Primates*, 45(2).

Liebal, K., Pika, S., & Tomasello, M. (in review). Gestural communication of orangutans (Pongo pygmaeus). *Gesture*.

Liebermann, P. (1998). *Eve spoke: Human language and human evolution* (Vol. 11). New York: Norton, W. W. & Co.

Marler, P. (1977). The evolution of communication. In T. A. Sebeok (Ed.), *How animals communicate* (Vol. 2, pp. 45-70). Bloomington: Indiana University Press.

Owings, D. H., & Morton, D. S. (1998). *Animal Vocal Communication: A New Approach*. Cambridge: Cambridge University Press.

Piaget, J. (1952). *The origins of intelligence in children*. New York: Norton.

Liebal, K., Pika, S., & Tomasello, M. (2004). Social communication in siamangs (Symphalangus Syndactulus): Use of gestures and facial expression. Primates, 45(2).

Liebal, K., Pika, S., & Tomasello, M. (in review). Gestural communication of orangutans (Pongo pygmaeus). Gesture.

Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (Gorilla gorilla): Gestural repertoire, learning and use. *American Journal of Primatology, 60(3)*, 95-111.

Pika, S., Liebal, K., & Tomasello, M. (2005). Gestural communication in subadult bonobos (Pan paniscus): Gestural repertoire and use. *American Journal of Primatology, 65(1)*, 39-51.

Plooij, F. X. (1978). Some basic traits of language in wild chimpanzees? In A. Lock (Ed.), *Action, gesture and symbol* (pp. 111-131). London: Academic Press.

Savage-Rumbaugh, E., Murphy, J., Sevcic, R. A., Brakke, K. E., Williams, S. L., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development, 58*, 1-256.

Savage-Rumbaugh, E. S., McDonald, K., Sevcic, R. A., Hopkins, W. D., & Rupert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General, 115*, 211-235.

Seyfarth, R. M. (1987). Vocal communication and its relation to language. In B. Smuts & D. L. Cheney & R. Seyfarth & R. Wrangham & T. Struhsaker (Eds.), *Primate societies* (pp. nv). Chicago: University of Chicago Press.

Snowdon, C. (1988). A comparative approach to vocal communication. In D. Legwe (Ed.), *Comparative perspectives in modern psychology*. Lincoln: University of Nebraska Press.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard: Harvard University Press.

Tomasello, M., Call, J., Nagell, K., Olguin, R., & Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees: A trans-generational study. *Primates, 35*(2), 137-154.

Tomasello, M., George, B. L., Kruger, A. C., Farrar, M. J., & Evans, A. (1985). The development of gestural communication in young chimpanzees. *Journal of Human Evolution, 14*, 175-186.

Tomasello, M., Call, J., Warren, J., Frost, T., Carpenter, M., & Nagell, K.

(1997). The ontogeny of chimpanzee gestural signals. In S. Wilcox, King, B. & Steels, L. (Ed.), *Evolution of Communication* (pp. 224-259). Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Zuberbühler, K. (2003). Referential signalling in non-human primates: Cognitive precursors and limitations for the evolution of language. *Advances in the study of behavior, 33*, 265-307.

THE EVOLUTION OF LANGUAGE AS A PRECURSOR TO THE EVOLUTION OF MORALITY

JOSEPH POULSHOCK

Language Evolution and Computation Research Unit, University of Edinburgh, 40 George Square, Edinburgh, UK EH8 9LL; Tokyo Christian University, 3-301-5 Uchino, Inzai-City, Chiba Japan, 270-1349

This paper argues that the evolution of human language is a prerequisite to the evolution of human morality. Human moral systems are not possible without fully complex language. Though protolanguage can extend moral systems, the design features of human language greatly extend human moral ability. Specifically, this paper focuses on how recursion, linguistic creativity, naming ability, displacement, and compositionality extend moral systems. The argument descriptively defines altruism as self-sacrificial behavior for others and morality as how a group classifies right and wrong behavior. No comment is made on how altruism squares with the replicatory selfishness of genes, or on the controversy of group selection. However, along with Dawkins (Dawkins, 1976), the author concurs that humans can use linguistically based concepts to help constrain genetic selfishness and promote degrees of altruism and morality. Though drawing on previous research, the ideas presented here are novel to the extent that they demonstrate how the design features of language support and extend human altruism and morality.

1. Recursive Linguistic Creativity Enhances Morality

Recursion refers to the "computational mechanisms [that provide] the capacity to generate an infinite range of expressions from a finite set of elements" (Hauser et al., 2002: 1571). According to these authors, recursion may be the only characteristic that distinguishes human language from non-human communication systems; thus, it clearly is at least one important distinguishing feature of human language. Other species may use recursion in other domains, such as, navigation and social relations. Nevertheless, for humans, recursion may enable us to express our moral ideas about an infinite number of situations, objects, and relations. If this is the case, creative recursive language makes human morality creatively recursive.

For instance, we can moralize about the usage of cellular phones in public places, about humane or inhuman treatment of animals, about issues pertaining to sexuality and personhood, about the responsibility of wealthy nations to poorer nations, about dress codes, about the amount of money wasted globally each year on necktie purchases, which could be used to instead for charity, about the inappropriateness or appropriateness of different kinds of humor, about use

and abuse of natural resources, and we can meta-moralize about morality itself, including why we think it immoral for people to moralize about our actions.

Besides being able to moralize about an infinite number of things, we can also moralize recursively about one thing. For example, we can use the following "if/then" and "not only/but also" recursive construction. If you do not return the money you found on the street, then the police may find out about it, and if the police find out about it, you could be charged with stealing (since this is a crime in this country), and if you are charged with a crime, then you will go to jail, and if you go to jail, then you will not be able to take care of your family, and if you cannot take care of your family, then you will not only be a criminal, but you will also be an irresponsible nincompoop of no count for putting your family into poverty, and if all these things could happen for not returning the money, then it would be better to simply return it, but if you do return it, then...

In addition, if this were not enough, we can meta-moralize about whether real morality exists or not. Nevertheless, the point here pertains not to whether recursion leads us to moral realism or anti-realism, but rather that (1) linguistic recursion helps us moralize about an infinite number of things, and (2) it also helps us moralize infinitely about any one thing. If this is the case, then linguistic recursion perpetually enables, extends, and enhances the range and number real and even imaginary scenarios we can moralize about.

Besides the fact that the creative and recursive nature of language makes human morality recursive, a recursive moral code also stands as a uniquely distinguishing feature of human morality compared to the proto-morality or altruism of non-human species. That is, degrees of recursive ability between species will differentiate the degrees of moral ability between species. Hauser, Chomsky, and Fitch (2002) say examples of animal recursive ability (navigation, number, and social calculus) stand as potential precursors to recursive language, and they suggest that domain-specific aspects of recursion became domain-general in humans. Along these lines, humans can combine recursive abilities; thus, recursive social calculus relates to how recursive language helps humans possess a recursive theory of mind. With language one can think: "I think that Henry thinks that Kenny borrowed Jim's book and should return it lest he fall out of favor with Jim and the rest of us." This stands as a linguistic form of social calculus that demonstrates moral differences between species. For example, how might apes express a recursive theory of mind? If Chimp A thinks that Chimp B and Chimp C are in conflict with each other, and if Chimp A attempts to help B and C reconcile, this behavior might stem from a recursive theory of mind. However, the important point here concerns how recursive language extends other recursive abilities in the moral realm--for example in how we think about and attempt reconciliation.

Regarding chimpanzee reconciliation (de Waal, 1982; Arnold and Whiten, 2001), recursion, and theory of mind, we cannot easily substantiate the claim that apes can read mental states (Povinelli and Vonk, 2004; Premack, 2004), which they could employ when reconciling. Moreover, though there may be some cases where chimpanzees can know what conspecifics know and do not know (Hare et al., 2001), whether they have the ability to attribute states of mind remains a controversial, complex, and debated point (Arnold and Whiten, 2001). This is because, "there is no easy way of making an a priori transition from behavioral similarity to psychological similarity" (Povinelli et al., 2000: 27). Interestingly, Povinelli, Bering, and Giambrone propose...

> ... that the majority of the most tantalizing social behaviors shared by humans and other primates (deception, grudging, reconciliation) evolved and were in full operation long before humans invented the means for representing the causes of these behaviors in terms of second-order intentional states (Povinelli et al., 2000: 25).

If this hypothesis obtains, then higher order representational abilities such as recursive language would add a whole new array of behavioral repertoire to the organism on top of these already existing behaviors.

More importantly for this discussion, language stands as a primary means to access the mental states of others, for though I may be able to deceive another about my intentions, I can also make my real intentions known. Moreover, I can tell you what I think you think, and you can tell me whether I am correct or not, or I can tell you what I think you think Frank is thinking, and you can tell me whether you think I am right or not. Therefore, if language does not make possible recursive theory of mind, at least language greatly extends it. Thus, no matter what ultimately causes apes to reconcile, linguistic recursion and recursive theory of mind greatly extend this behavior in human beings.

For example, you may be a noisy neighbor, and you may not know that your noise bothers your neighbor, but your bothered neighbor could solve this problem directly by talking to you, or she could recursively communicate with you through another neighbor. She may tell another neighbor of the problem, and ask him to approach you with a request to be quieter. When she does this, you can apologize to her through the mediator without even seeing or speaking to her. Something similar to this happened when the US President George Bush apologized for Iraqi prisoner abuse through King Abdullah of Jordan. To describe his conversation with the King, Mr. Bush said that he told the King he was sorry for the humiliation suffered by the Iraqi prisoners and the humiliation suffered by their families. This may not represent a valid admission of guilt, but it does demonstrate a socially and linguistically recursive apology.

The above exemplifies how humans use recursive language with a theory of mind to extend the range and variety of human moral behavior. Language

gives us recursive access to other minds and our moral relations to them, enabling us to recursively socialize and moralize. Regarding recursion, Aitchison (1999: 79) says, "we can never make a complete list of all the possible sentences in any language," and this suggests an infinite number of things, events, and people we can moralize about. Thus, recursion stands as a defining feature of human language and social calculus with its linguistic access to other minds, which strongly affects human sociality and morality.

2. Creativity, Naming, and Morality

In addition, a building block of recursion, "the naming insight" also extends and expands the human ability to moralize. Speaking of the origin of human language, Aitchison (1999: 19) asserts that besides being able to produce a range of sounds, humans "must have attained the 'naming insight,' the realization that sounds sequences can be symbols which 'stand for' people and objects." Non-human species such as some primates have the cognitive abilities to name things (Savage-Rumbaugh et al., 2001), and other animals, such as dogs, have the ability to recognize names for things (Kaminski et al., 2004). Nevertheless, except for the type of alarm calls we see in vervet monkeys, these naming abilities appear to emerge only after intensive training under the tutelage of language-enabled humans.

Hence, though whales and dolphins have signature calls that indicate their presence to the group, and vervet monkeys have a number of calls for predators, we generally see extensive name-production and name-recognition in non-human species because we use human language first to teach "names" to these species. Moreover, for animals that do possess minimal naming insight, they do not appear to use it to attribute moral values to named items, but the possibility raises some questions. Do animals that possess a naming insight on their own without human instruction, such as vervets, attribute moral-like qualities to the objects they name? Do language-trained animals name objects with a moral sense of good, bad, right, and wrong? This may be unlikely, but for our discussion here, human use of the naming insight stands as a distinguishing feature of human morality. Not only can we name objects, people, events, and concepts, but we can also coin new names for anything, and most importantly, we can attribute the values of good, bad, right, and wrong to the things we name.

Hence, because we can name stuff, we can moralize about what we name in a very simple and protolinguistic fashion. For example, "monogamy is good; "polygamy is bad;" or for those opposing the legalization of marijuana: "weed is bad;" or for those in favor of trickle down economics "greed is good." Such moralizings are relatively simple because they do not require recursion, syntax, and argument structure; that is, changing syntax does not change the meaning: "bad is weed" and "good is greed." Moreover, argument structure "who does

what to whom" does not function in these phrases. Thus, we can moralize protolinguistically, with simple labels and without argument structure.

Regarding how naming ability and language enhance morality, the skeptical reader might wonder how we might use language for selfish and immoral purposes. Thus, before moving on, a small caveat is needed. For example, a large literature exists on the human ability to deceive with language (Renshaw, 1993; Stiff and Miller, 1993; Wortham and Locher, 1999; Galasinski, 2000; Meltzer, 2003; Newman et al., 2003). Hence, though language has the power to extend moral behavior, it also holds the opposite power to deceive others, negate morality, and advance malevolence. Thus, language may give us the ability to create an alternative morality, such as in George Orwell's novel 1984, in which "Newspeak" is used to teach, "War is peace. Freedom is slavery. Ignorance is strength" (Orwell, 1950: 7). The topic of how language can facilitate anti-altruism and immorality transcends the focus of this paper. However, though we must acknowledge the negative power of language to deceive and serve selfishness, this does not negate the positive power of language to enable, extend, and maintain human altruism and morality.

3. Displacement Enhancing Morality

In addition to how naming ability helps us assign moral values to what we name, language also helps us make abstractions, and this highlights the unique feature of human language called displacement. Crystal (1992: 26) defines displacement as the ability "whereby language can be used to refer to contexts removed from the immediate situation of the speaker (as in the cases of tenses which refer to past or future time)." Animal calls, on the other hand, only refer to "specific situations, such as danger and hunger, and have nothing comparable to displaced speech" (26). Hence, displacement enables humans to refer to things removed in space, time, and even reality from the speaker, referencing the hypothetical or unreal. Though some species exhibit limited displacement ability, as in bee dancing, this still refers to the specific physical location of displaced nectar. Thus, displacement exhibits unique features in human language that transcend concrete situations.

How could linguistic displacement uniquely enhance and extend human morality? For one thing, as previously mentioned, it enables us to moralize about the past and the future, and though some animals might feel regret about past events, such as an elephant or gorilla mourning the loss of kin, this is still quite different from moralizing about past events. Is it possible that two bonobo chimps could be made to regret their secretive copulation through a verbal rebuke even if the dominant male who might physically oppose such behavior never found out about it? Would it be possible through verbal or any other means to make a male elephant mourn the death of conspecifics he has not actually physically seen? However, even a human child in the first grade of

elementary school can reflect on a parent's scolding: "it was not good that you lied to your teacher, telling her your dog ate your homework, instead of the truth that you simply forgot to do it."

Besides past-event-moralizing, with language we can turn our attention to the future and instruct a child in the following way. "Tomorrow you will apologize to your teacher, and tonight (future-displacement) you will write your ancestors (abstract-displacement) an apology, reflecting on how you can remember your homework and reasons why (hypothetical-displacement) you should not lie again (future-displacement)." Besides moralizing about the past and future, displacement enables us to moralize about the hypothetical and unreal. For example, "if your boss pressured you to lie about your company's financial accounting, would you follow your boss or blow the whistle on him?" Moreover, in an ethics course, participants can discuss ways to carefully deal with ethical issues before they ever encounter them. Additionally, we can think about fictional or futuristic moral dilemmas. If you suddenly found yourself with the ability to foresee the future with 80% accuracy, and the government asked you to predict terrorist activity and arrest "pre-crime" terrorists before they can act, what would you do about it? In short these examples show that displacement, as a defining feature of human language also distinguishes human moralizing from proto-moralities because it enables us to think morally about that which is removed from us in space, time, and even reality.

Moreover, it is interesting to note how displacement relates to recursion. First, displacement does not require recursive embedding, for we can refer to the future, the past, places, and non-realities in proto-linguistic ways (with 1-word utterances): tomorrow, yesterday, Venus, Mars, Hercules, and Zeus. Incidentally, though we can name these concepts in 1-word utterances, we may need recursive ability to understand at least some of them. For example, even if we see statues or images of the god Zeus (upholder of justice and morals), we still cannot understand what the name means without a recursive explanation. Nevertheless, though displacement does not require recursion, with recursion, displacement becomes unlimited--enabling us to moralize without end about anything removed from us in space, time, and reality.

4. Compositionality, Recursion, and Morality

Besides displacement and stimulus freedom, how do the design features of recursion and compositionality affect human morality? Smith says:

> Recursiveness allows the creation of an infinite number of utterances. Compositionality makes the interpretation of previously unencountered utterances possible--in a recursive compositional system, if you know the meaning of the basic elements and the effects associated with combining

elements, you can deduce the meaning of any utterance in the system (2003: 4)

Hence, while recursion allows humans to create an infinite number of novel moral utterances, compositionality refers to our ability to comprehend them. Regarding compositionality the nuance here does not concern our ability to endlessly moralize about everything or any one thing, but rather our ability to comprehend all this moralizing. Humans can compositionally comprehend recursive moralizing through hearing speech, reading texts, and viewing sign language. In sum, human beings can linguistically produce an infinite and novel moral output (recursion) as well as comprehend an infinite and novel moral input (compositionality).

Regarding actual behavior, infinite and novel moralizing does not necessarily create altruism in people; that concerns a rather different question. However, as recursion and compositionality enable us to incessantly send and receive moral messages, this ability may dramatically affect our general moral nature as humans, whether we behave altruistically nor not. Hence, language not only remarkably defines human uniqueness, but these linguistic abilities also significantly determine our moral nature through what they enable us to moralize about. We can recursively and compositionally moralize about not just everything or any one thing, but everything and any one thing embedded in and in combination with everything else. Thus, in principle nothing is necessarily morally neutral, and no meaning can escape the reach of moralizing language.

5. Conclusion

For lack of space, the discussion has ignored many topics, such as cultural transmission, stimulus freedom, UG, and categorical ability enhanced by language. Neither has it touched the topic of genetic selfishness or the problem of group selection. However, the argument implies that language-based moral concepts may give humans a lever that sometimes can help us overcome genetic constraints on altruism. Moreover, the argument briefly outlines how the evolution of human morality requires a pre-existing linguistic system. Moral systems could evolve along with linguistic systems, but when we look at our moral abilities, this paper makes clear that human morality requires language. Moreover, it also raises many other important questions. For example, did early human groups experience a conflict between their social needs and genetic interests? If so, could this conflict of interest have pressured them into developing their moral systems? If these moral systems require language, could these pressures and conflicts have forced an evolution in the complexity of human language? These are interesting questions worthy of further inquiry, and that further inquiry should take place as much as this paper has demonstrated the strong relationship between human language and morality.

References

Aitchison, J. (1999). Linguistics: An introduction. London: Hodder & Stoughton.

Arnold, K., & Whiten, A. (2001). Post-conflict behaviour of wild chimpanzees (pan troglodytes schweinfurthii) in the budongo forest, Uganda. Behaviour, 138, 649-90.

Crystal, D. (1992). Introducing linguistics. London: Penguin.

Dawkins, R. (1976). The Selfish Gene. Oxford, UK: Oxford University Press.

de Waal, F. (1982). Chimpanzee politics: Power and sex among apes. London: Counterpoint.

Galasinski, D. (2000). The Language of Deception: A Discourse Analytical Study. Sage Publications.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? Anim Behav, 61(1), 139-51.

Hauser, M., Chomsky, N., & Fitch, T. (2002). The faculty of language: What is it, Who has it, and How did it evolve? Science, 29822, 1569-79.

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: evidence for "fast mapping". Science, 304(5677), 1682-83.

Meltzer, B. (2003). Lying: Deception in human affairs. International Journal of Sociology and Social Policy, vol. 23, no. 6-7, 61-79.

Newman, M., Pennebaker, J., Berry, D., & Richards, J. (2003). Lying Words: Predicting Deception from Linguistic Styles. Personality and Social Psychology Bulletin, vol. 29, no. 5, 665-75.

Orwell, G. (1950). 1984. Harmondsworth, Middlesex, UK: Penguin Books.

Povinelli, D., Bering, J., & Giambrone, S. (2000). Toward a science of other minds: Escaping the argument by analogy. Cognitive Science, Vol. 24 (3), 509-41.

Povinelli, D., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. Mind and Language, Vol. 19 No. 1, 1-28.

Premack, D. (2004). Is language the key to human intelligence? Science, 303, 318-20.

Renshaw, D. C. (1993). Lies and medicine: reflections on the etiology, pathology, and diagnosis of chronic lying. Clin Ther, 15(2), 465-73; discussion 432.

Savage-Rumbaugh, S., Shanker, S. G., & Taylor, T. J. (2001). Apes, Language, and the Human Mind. Oxford University Press.

Smith, K. (2003). The Transmission of Language: models of biological and cultural evolution. (Doctoral dissertation, University of Edinburgh, 2003).

Stiff, J. B., & Miller, G. R. (1993). Deceptive Communication. Sage Publications.

Wortham, S., & Locher, M. (1999). Embedded metapragmatics and lying politicians. Language & Communication, 19(2), 109-25.

MODELLING THE TRANSITION TO LEARNED COMMUNICATION: AN INITIAL INVESTIGATION INTO THE ECOLOGICAL CONDITIONS FAVOURING CULTURAL TRANSMISSION

GRAHAM R. S. RITCHIE and SIMON KIRBY

Language Evolution and Computation Research Unit,
University of Edinburgh, Edinburgh, UK
graham.ritchie@ed.ac.uk / simon@ling.ed.ac.uk

Vocal learning is a key component of the human language faculty, and is a behaviour we share with only a few other species in nature. Perhaps the most studied example of this phenomenon is bird song which displays a number of striking parallels with human language, particularly in its development. In this paper we present a simple computational model of bird song development and then use this in a model of evolution to investigate some of the ecological conditions under which vocal behaviour can become more or less reliant on cultural transmission.

1. Introduction

One of the most unusual characteristics of language, when compared to many of the other communication systems found in nature, is the extent to which it relies on vocal signals transmitted culturally rather than genetically. This is of considerable interest as other modelling work has demonstrated the role that cultural transmission, via 'iterated learning', may play in explaining many prominent features of human languages, e.g. the emergence of compositional syntax (e.g. Brighton, 2002), regular and irregular word forms (Kirby, 2001), and dialects (Livingstone, 2002). The evolution of learning can therefore be seen as a key transition in the evolution of human language.

Vocal learning is a comparatively rare evolutionary development, it appears to have only evolved in three groups of mammals: humans, bats and cetaceans, and three groups of birds: songbirds, hummingbirds, and parrots (Jarvis, 2004). Of these, the development of bird song and human language have a number of striking similarities, e.g. both nestlings and human babies have a critical period for learning, both rely on auditory feedback for normal development, and both exhibit a form of early babbling (known as subsong in birds) (Doupe & Kuhl, 1999). This suggests that there may be strong epigenetic constraints on the evolution of a learned vocal system (Jarvis, 2004), and so studying the evolution of learning in bird song may help us to elucidate possible ecological factors which played a role in the transition to learned communication in our own species.

In this paper we use a computational model to investigate the possible role of two very simple ecological conditions which we think may affect the transition

to learning; namely the reliability of cultural transmission, and the stage of life at which communication is required.

2. The auditory template model of song development

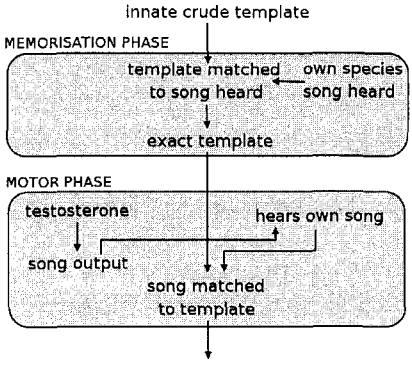


Figure 1. The auditory template model of song development. Figure after (Catchpole & Slater, 1995).

The song learning behaviour of many different species of the oscine passerines has been extensively studied, for an introduction see the reviews in (Catchpole & Slater, 1995) and (Marler & Slabbekoorn, 2004). The exact pattern of song development varies greatly among different species, but in attempting to capture the general features, bird song biologists have developed what is known as the 'auditory template model' of song learning, depicted in figure 1. This model posits two distinct phases to song learning; an early memorisation phase in which songs that are heard as an infant which are recognised as conspecific by an innate 'crude template' are memorised, and a later motor phase when song production is trained to produce songs that match the learned template. This behaviour can be contrasted with the sub-oscine passerines which appear to have a largely innately specified song, and will develop normal song production without hearing conspecific song and without auditory feedback.

3. A simple computational model

We take this model as our inspiration and develop a computational model of the two stages of learning in bird song, described in the proceeding two sections. We then use this model to investigate some conditions under which song perception and production can come to be increasingly influenced by cultural transmission.

3.1. *Phase 1: Observational learning*

To model the memorisation phase of song learning we hypothesise a module which we term the Species Recognition Device (SRD). This is intended to model the auditory biases birds appear to show towards conspecific song.

We model the SRD as a note transition matrix which defines the transition probabilities between every available note (or song element)[a]. We assume that the notes are fixed and identical for every agent in the simulation, and the number of notes used here was 6. We realise that this is unrealistic and that many species learn the form of song elements from their tutors as well as the element sequence. We also realise that element transitions or sequence are not the only cues birds use to identify conspecific song. However, a note transition matrix provides us with a simple and computationally tractable model of these sorts of biases.

Each agent in the model has 'genes' which code for an innate SRD, this is intended to model the 'crude template' as described above. An agent uses its SRD to categorise songs it hears as either conspecific or not by comparing the note transitions in the song with the transition probabilities in the matrix.

Such a matrix can be more or less biased to a particular song-type, if all the probabilities in the matrix are equal then the matrix has no preference to any particular song, while if each row has exactly one high probability transition, the matrix is maximally biased to one particular song. We can measure this bias by calculating the Shannon entropy for each transition distribution, and we can measure the preference of a matrix for a particular song by comparing the transitions found in the song and the probabilities in the matrix. We have used these measures of matrix preference and bias in earlier work (Ritchie & Kirby, 2005), and the reader is referred there for a more detailed definition.

An agent's adult SRD is also subject to being altered by songs heard in early life, we model this by 'exposing' each agent to 100 songs from its environment and getting it to select the ones preferred by its innate SRD (crude template). The note transitions in the songs that are selected at this stage are then reinforced in the agent's SRD to produce the agent's adult SRD, or 'exact template'. The degree to which an agent's SRD is modifiable by songs heard in early life is determined by genes which code for the agents SRD plasticity (SRDP), this will be a value between 0 and 1, with 0 meaning the innate SRD is entirely fixed, and 1 meaning that the agent relies only on songs heard early in life to construct its adult SRD.

3.2. *Phase 2: Reinforcement learning*

The SRD as described in the previous section models an agent's sensory biases (or lack thereof) to a particular song-type. We also require a model of song production. We also model this as a note transition matrix[b], but here the probabilities determine the probabilities of singing one note after another. We call this the Song Production Device (SPD). Just as for the SRD, an agent encodes innate biases for

[a] While we implement the SRD as a note transition matrix here, we hope that this component could be modelled in many different ways, e.g. as a neural net with the initial weights specified genetically.

[b] Again, we hope that the SPD component could be modelled in a number of different ways, not necessarily using the same mechanism as for the SRD.

its SPD in its genes.

To model plasticity in the production mechanism, we allow the SPD to be trained by reinforcement learning using the agent's SRD as a critic, using a very simple learning algorithm. This is intended to model the process by which a bird uses its memorised exact template to guide its vocal development. As for the SRD, the degree to which the adult SPD is allowed to be influenced by learning, the SPD plasticity, (SPDP), is determined genetically. If the plasticity is 0 then the SPD is not influenced at all by the learning procedure described below, higher values mean the SPD becomes increasingly influenced by learning.

The SPD is trained by getting the agent to produce a song and then to 'listen' to this song with its adult SRD, if a note transition in the song is 'accepted', i.e. has a high probability in the SRD matrix, that transition's probability is increased slightly in the SPD. This process is repeated 250 times, after which the agent's SPD is said to have 'crystallised' and will not change again in he agent's lifetime[c].

3.3. *Determining fitness*

We define an agent's fitness as its ability to recognise and be recognised by conspecifics. This seems a reasonable model of one of the main pressures acting on song (Catchpole & Slater, 1995), although there are of course many other pressures acting on song in he wild (e.g. sexual selection for variation, adaptation to the local acoustics etc.), and we hope to model some of these in future work.

To calculate an agent's fitness we perform 250 fitness trials. In each trial we get the agent to produce a song using its crystallised SPD and we then randomly select another member of the population and check that this second agent correctly recognises the song using its adult SRD. We also get the second agent to produce a song and check that the first agent correctly recognises the song. Every correct recognition means that the agent's fitness is incremented by 1.

Defining fitness in this way means that there is a strong selection pressure for the agents to develop and maintain a stereotypical and easily recognised species-specific song. As the SRD is modelled as note transition probability matrix, this corresponds to a matrix with a single high probability transition for each individual note. In short, in this environment it is adaptive to have strongly biased matrices.

3.4. *Overall model design*

The overall model works with an evolving population of 100 agents. As we want to investigate how a genetically specified song can come to be learned we initialise the agents innate SPD and SRD genes to one particular song "abcd", and the plasticity genes to 0. This means that the population will start off receiving maximal

[c]Unfortunately we do not have space to describe the learning algorithm in detail here. Further details are available upon request, and will be described more fully in future work.

fitness values and any mutations that degrade an agent's ability to sing and recognise conspecific song will be selected against. Each agent in each generation then goes through the following 'life stages':

Birth The agent's innate SRD and SPD, along with its SRDP and SPDP, are decoded from its genes.

Development Each agent is exposed to the songs of the previous generation, and picks those which will be used for learning using its innate SRD. The agent then goes through the two stages of learning described above to give them their adult SRD and crystallised SPD.

Adulthood The agents are tested in 250 fitness trials as described above to see how many times it can correctly recognise a bird of its own species and how many times its song is correctly recognised by a bird of its own species. These values are summed to give a bird's fitness score.

Reproduction Parents from the population are selected probabilistically according to their fitness score and their genes are recombined and subject to a low mutation rate to produce new child agents[d].

Death Each bird in the population is sampled 5 times and the resulting songs are stored for the next generation to learn from. All of the current birds in the population are removed and their children become the new population.

We repeat this process over many generations and record various measures over the course of a run.

4. Experiments

In this initial investigation we only model two very simple ecological conditions:

Environmental reliability For the first experiment we vary the reliability of the environment, that is the degree to which the previous generation's songs are faithfully recorded and then passed on to the new generation to learn from. We have two conditions: a reliable environment where we keep 80% of the previous generation's songs, and an unreliable environment where we keep only 20% of the previous generation's songs. The remaining songs are randomly generated songs which use the same notes and are constrained to within the same length as the agents' songs. This intended to model heterospecific song or other extraneous sounds in the birds environment.

Timing of song requirement In the first experiment we only test the bird's fitness after learning has taken place, in this experiment we also check the bird's fitness *before* learning. This intended to model a possible environment in which song is

[d]This is implemented with a standard genetic algorithm (GA), using tournament selection, a crossover rate of 0.7 and a mutation rate of 0.01. Mutation is modelled by simply replacing the gene that is to be mutated with a uniform random number between 0 and 1.

required immediately after birth as well as later in life.

5. Results

We provide results for each of the three different conditions described above in figure 2. The measures shown in each are the population average fitness, SPDP, SRDP, SPD change and SRD change. The SPD and SRD change are simply the absolute difference of the bias value of the innate and adult matrices (as discussed in section 3.1 above). We measure this as well as the plasticity values as the plasticity values can vary *without* a correlated variation in the change values (as demonstrated in figure 2c).

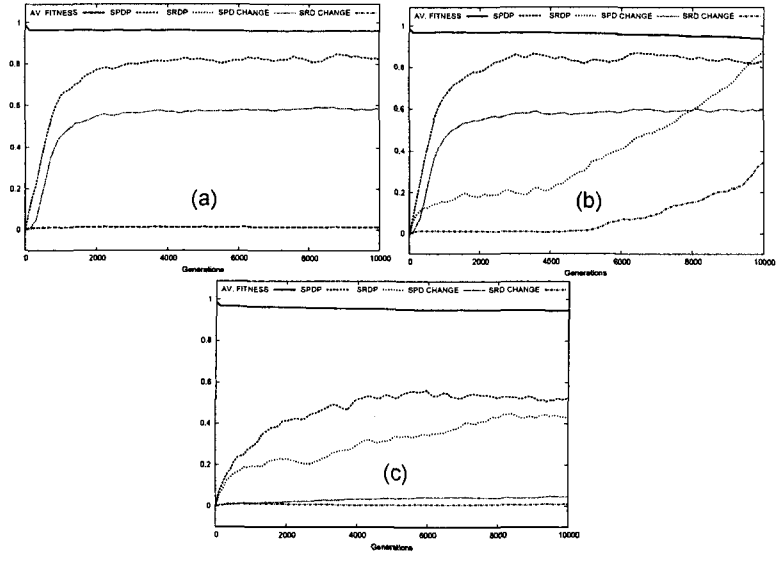


Figure 2. Results for the three different environments. The X-axis in each graph is the number of generations, set to 10000 for all results shown here. The Y-axis in each graph measures the population average fitness, SPDP, SRDP, SPD change and SRD change for each different condition. Graph (a) shows results for an unreliable environment where only 20% of the previous generation's song are faithfully passed on. Graph (b) depicts a reliable environment where 80% of the songs are passed on. Graph (c) shows results for a reliable environment in which the agents' fitness is checked both before and after learning. These results are the averages of 10 separate runs for each condition with a different random number generator seed for each. We have smoothed the graphs to allow us to better see the overall trends.

In all of the conditions we found that fitness stayed fairly fixed throughout all of the runs. However, the degree to which song remained being transmitted genetically depended on the environment, as demonstrated by the different values of SRDP and SRD change at the end of each simulation.

In the unreliable environment the population cannot count on hearing conspecific song as infants. The agents therefore have to keep transmitting their song

genetically, as demonstrated by the much lower SRD change and SRDP at the end of the run in figure 2a. In contrast, in the reliable environment shown in figure 2b, towards the end of the runs the population begins to transmit their song culturally as demonstrated by the coincident rise in the population's SRDP and SRD change.

In both experiments, however, the SPD change and SPDP quickly rise, indicating that the SPD is always being trained using the adult SRDs and the reliability of the environment appears to have no bearing on this. As long as the adults can construct a faithful copy of the species song in their SRDs as a result of either cultural or genetic transmission, it can always be used to train the SPD, and so there is no pressure for the copy of song stored in the SPD to be transmitted genetically and mutation pressure quickly erodes the genetic copy.

Figure 2c show results when the timing of song requirement is changed, where we test an agent's fitness both before and after learning. The SPD and SRD change values stay low throughout the run, demonstrating that SPD and SRD copies of song remain genetically transmitted throughout the run. The average SPDP and SRDP values drift to around 0.5 as there is no selection pressure acting to maintain these at any particular value.

6. Discussion

The results described here predict two simple environmental conditions which could affect the transition to a learned communication system; the reliability of the cultural environment, and stage of life at which communicative behaviour is required. These conditions seem fairly widely applicable and it seems reasonable that these conditions may have played a role in the transition to increased reliance on learning in human communication as well.

We think that this model also provides an interesting case study of the interaction of genetic and cultural transmission and phenotypic plasticity. We see that where the environment is reliable enough, and a learning mechanism is available to the population, the genes need not code for a song explicitly as an agent can rely on obtaining a copy of the 'correct' song via cultural transmission. Cultural transmission can thus, in some conditions, be seen as a *masking* force (Deacon, 2003) on genetic transmission, with a similar end result to that we found in earlier work (Ritchie & Kirby, 2005) for rather different environmental conditions.

Another interesting result is that in all of the experiments described here the agents come to rely solely on their auditory copy of song (in the SRD) to guide later production behaviour. We feel that this again represents a form of genetic parsimony, as it seems rather inefficient for an agent to store two 'copies' of their song genetically, even though these copies are likely to be represented in rather different ways; one being a sensory and the other a motor mechanism. Nevertheless, if there is enough phenotypic plasticity to allow these to interact, and if the genetic 'cost' of this plasticity is lower than the cost of encoding a song genetically, we see that even in the unreliable environment the agents rely on only

their auditory copy, but need it always be this way round? In the case of bird song it seems so, as a bird only needs to produce a song when it is sexually mature while it needs to be able recognise conspecific song earlier. This means that the song recognition system should be more genetically constrained than the song production system, which seems to match the biological data.

While this may be true of bird song it is not so clear for human language as children become capable talkers well before puberty. In future work we would like to relax some of the assumptions built into the current model with regard to the timing of each of the learning phases and allow this to be under genetic control. The two ecological conditions we discuss here are the simplest relevant condition we could think of, and we would also like to model other relevant ecological conditions, such as sexual selection pressure, to see what role these may play in conjunction with the conditions investigated here.

References

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*(1), 25–54.

Catchpole, C. K., & Slater, P. J. B. (1995). *Bird song: Biological themes and variations.* Cambridge University Press.

Deacon, T. (2003). Multilevel selection in a complex adaptive system: the problem of language origins. In B. Weber & D. Depew (Eds.), *Evolution and learning: the baldwin effect reconsidered* (pp. 81–106). MIT Press, Cambridge, MA.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Reviews of Neuroscience*, *22*, 567–631.

Jarvis, E. D. (2004). Brains and birdsong. In P. Marler & H. Slabbekoorn (Eds.), *Nature's music: The science of birdsong* (pp. 226–271). Academic Press Inc. (London) Ltd.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, *5(2)*, 102–110.

Livingstone, D. (2002). The evolution of dialect diversity. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 99–118). London: Springer Verlag.

Marler, P., & Slabbekoorn, H. (2004). *Nature's music: The science of birdsong.* Academic Press Inc. (London) Ltd.

Ritchie, G., & Kirby, S. (2005). Selection, domestication, and the emergence of learned communication systems. In *Proceedings of AISB 2005: Social intelligence and interaction in animals, robots and agents.*

TOWARDS A SPATIAL LANGUAGE FOR MOBILE ROBOTS

RUTH SCHULZ, PAUL STOCKWELL, MARK WAKABAYASHI, JANET WILES
School of Information Technology and Electrical Engineering,
The University of Queensland,
Brisbane, QLD, 4072, Australia

We present a framework and first set of simulations for evolving a language for communicating about space. The framework comprises two components: (1) An established mobile robot platform, RatSLAM, which has a "brain" architecture based on rodent hippocampus with the ability to integrate visual and odometric cues to create internal maps of its environment. (2) A language learning system based on a neural network architecture that has been designed and implemented with the ability to evolve generalizable languages which can be learned by naive learners. A study using visual scenes and internal maps streamed from the simulated world of the robots to evolve languages is presented. This study investigated the structure of the evolved languages showing that with these inputs, expressive languages can effectively categorize the world. Ongoing studies are extending these investigations to evolve languages that use the full power of the robots representations in populations of agents.

1. Introduction

While all human languages can describe spatial representations, people speaking different languages will use different frames of reference: intrinsic (from the point of view of the object), relative (from the point of view of the speaker or some other viewer) or absolute (e.g. North, South, East and West) (Levinson, 1996). These frames of reference can be used to construct or describe spatial relationships in the world. The use of different frames of reference in different languages indicates that language may restructure the spatial representations of the language speaker, rather than the existence of innate and universal spatial concepts (Majid, Bowerman, Kita, Haun, & Levinson, 2004).

Computational modeling of language evolution provides a means of investigating ontology, grounding, learnability, and generalization in languages that evolve in populations of agents (See Steels, 2005 for an outline of the major stages in the evolution of language using computational models). The use of simulation techniques can add to the debate on the origins and evolution of language by determining factors that are important for evolving communication systems. Language games are a possible framework for language models in which agents engage in tasks requiring communication. These games have been used to evolve lexicons (Hutchins & Hazlehurst, 1995), categories (Cangelosi & Harnad, 2001), and grammars (Batali, 2002) in populations of agents.

The symbol grounding problem (Harnad, 1990) is a major issue for computational models of language. Without the grounding of meanings in the world, symbols refer only to other symbols with no association between the symbols and the world. One way to address the symbol grounding problem in computational models of language is to conduct language research with real or simulated robots (Marocco, Cangelosi, & Nolfi, 2003; Roy, 2001; Steels, 1999; Vogt, 2000).

In robot language research, the environments are often simplified and idealized compared to the real world. In the Talking Heads Experiment (Steels, 1999) geometric shapes were used rather than 'real world' objects such as tables and chairs. The languages evolved in the Talking Heads Experiment used a relative frame of reference to talk about the different shapes in the scene using meanings such as 'left' and 'right'.

One way to extend robot language research is to use mobile robots that interact with a real world environment, using navigation systems to build up internal maps of the world. The use of mobile autonomous agents that move in a real environment enables the evolution of spatial languages using both relative and absolute frames of reference. The visual input of the robot would be used in a relative frame of reference, where the scenes can be categorized with respect to what the world looks like from the perspective of the robot. The internal maps would be used in an absolute frame of reference. The languages evolved could provide a methodology to investigate the structure of languages that describe space.

This paper introduces RatChat, a project that uses RatSLAM, an established mobile robot platform, to develop a framework for the robots to evolve a language describing their environment. The RatChat and RatSLAM projects are described in Section 2. A study using this platform to evolve spatial languages is presented in Section 3, followed by a general discussion and conclusion.

2. RatChat

Simultaneous Localisation and Mapping (SLAM) is a methodology for robot map building and navigation. RatSLAM is a model of SLAM, based on the hippocampal complex in rodents, that uses a combination of the properties of grid based, topological, and landmark representations to keep a sense of space while adding robustness and adaptability (Milford, Wyeth, & Prasser, 2004).

The inputs to the RatSLAM system include odometry and vision with the resulting map represented by pose cells. Active pose cells represent the current location and orientation of the robot, and are arranged in (x,y,θ) for ease of

visualization. With RatSLAM, robots use the appearance of an image to aid localization by learning to associate the appearance of a scene and its position estimate (Prasser, Wyeth, & Milford, 2004).

RatChat aims to evolve a shared lexicon between robots grounded in perceptions, local views, and behaviors using a language game framework (see Figure 1). The evolution of languages for locations will be explored, later extending the vocabulary of the robots to include objects. The challenge is for the robots to categorize their internal representations and label these with appropriate generalization and variability. The shared lexicon should allow the robots to agree on words for categories while including sufficient diversity for different categories to have different labels. As the language is expanded to include objects, more emphasis will be on the visual inputs of the robots (see Figure 2).

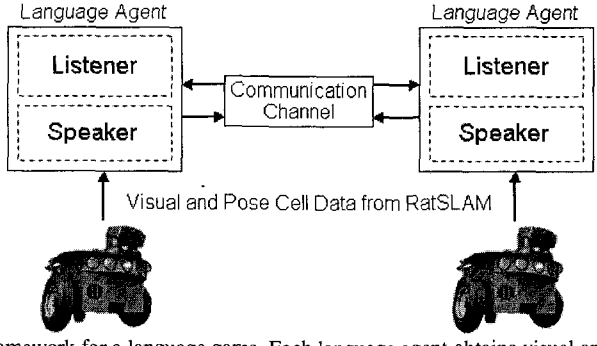


Figure 1 The framework for a language game. Each language agent obtains visual and pose cell data from the RatSLAM system. A communication channel is set up between the agents, allowing the speaker for each agent to produce utterances, and the listener for each agent to receive utterances for comprehension.

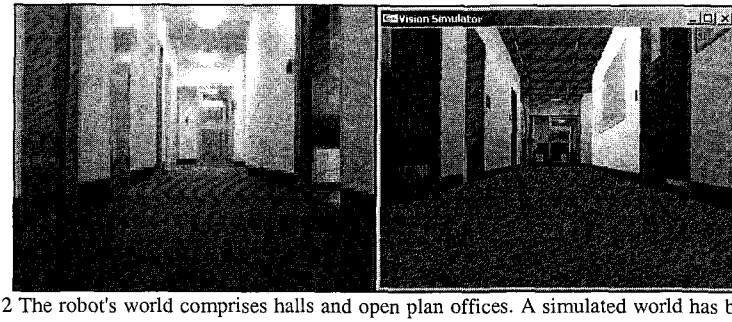


Figure 2 The robot's world comprises halls and open plan offices. A simulated world has been built to mirror the real world. The features of the environment shown in the visual images seen by the robot include the floor, walls, desks, chairs, and filing cabinets. The left image is from the robot's camera and the right image is the same location in the simulated world.

The RatChat language agents consist of a speaker and a listener based on simple recurrent neural networks (Elman, 1990; Tonkes, Blair, & Wiles, 2000). Speaker networks are extended to include the output of the network in the context for the next time step. Preliminary simulations showed that languages are easier to learn when the meaning space patterns are non-orthogonal and that distributed representations in signal space enable expressive languages to be found more easily than if localist representations are used.

3. A Spatial Language

This study investigated the evolution of spatial languages using the visual and pose cell representations of the robot, looking at the expressivity of the languages evolved, and how the languages categorized the world of the robot.

Methods: The visual input for this study was every 100^{th} scene in a series of 10000 visual scenes of 12x8 gray scale arrays obtained from a run of the robot in the simulated world. The pose cell input for the study was every 100^{th} pattern in a series of 10000 pose cell patterns from the same run. The number of cells was reduced from 440640 to 610 by reducing the resolution of the pose cells (4x4x4 pose cells to 1 pose cell), and by discarding cells that are inactive in every pattern. For a third representation, the pose cells were processed using a hybrid system based on Self Organizing Maps (SOMs) (Kohonen, 1995). In the processing system, a SOM was trained on the input series for 1000 epochs. The output of the SOM was a 12x8 set of competitive units organized in a hexagonal pattern. To construct a distributed activation the actual output values of the units were converted to values between 0 and 1.

For the signal representation, utterances consisted of a sequence of three syllables. Each syllable was represented by a ten unit binary vector in which the two most active units were set to one, with all other units set to zero.

One way to measure understanding is to test how well an agent has categorized the world. The representations of the world are presented to the speaker, resulting in words associated with each pattern. Listeners produce a prototype for each unique utterance. If the original input pattern presented to the speaker is closest to the prototype for the utterance used by the speaker, this pattern has been correctly categorized. When many of the patterns are associated with one word, the agents will categorize more patterns correctly, but the language does not divide the meaning space effectively. A more appropriate measure of understanding is the number of patterns correctly categorized divided

by the largest category size, indicating how well the language divides up the meaning space, and how well the agent understands the language.

In this study, ten agents were evolved individually for 100 generations to produce languages based on each set of inputs (vision, pose cells and processed pose cells). A simple (1+1)-evolutionary strategy (Beyer & Schwefel, 2002) was used to evolve the agent's speaker, introducing variability in the language. At each step, the agent's speaker was evolved and the agent's listener was trained on the language from the speaker for 500 epochs using the Back Propagation Through Time algorithm (Rumelhart, Widrow, & Lehr, 1994). The agents were evaluated with a fitness function based on the measure of understanding described above. If the listener trained on the mutant languages were better at categorizing the input patterns than the listener trained on the current champion language, then the mutant became the champion. The languages produced by the agents for each set of inputs were compared for expressiveness, categorization and how the meaning space was divided.

Results: The agents evolved with visual scenes as inputs produced languages with an average of 24.2 words (see Table 1). The average number of scenes correctly categorized by the agents was 53.4 out of 100. One highly expressive language had 67 unique words of which 47 were associated with single scenes. Words often appeared to group several different types of images together, with the resulting prototype visual scene for the word a combination of these scenes. One set of similar scenes were those in which the robot faced a white wall with a strip of black next to the floor. All of the languages other than the most expressive language grouped together some of these scenes (see Figure 3).

Table 1 Properties of the languages evolved with different sets of input

	Number of Unique Words (avg (std))	*Number of Patterns Correctly Categorized (avg (std))*
Vision	24.2 (17.3)	53.4 (13.5)
Pose Cells	23.2 (12.4)	22.6 (10.4)
Processed Pose Cells	10.9 (6.4)	58.7 (10.4)

The agents evolved with pose cells as inputs produced languages with an average of 23.2 words. The average number of scenes correctly categorized by the agents was 22.6 out of 100. The majority of the words were associated with single input patterns or a small number of input patterns, scattered across the space. Some words group together input patterns that are close together in space, but these words are also generally associated with a small number of input patterns from other areas.

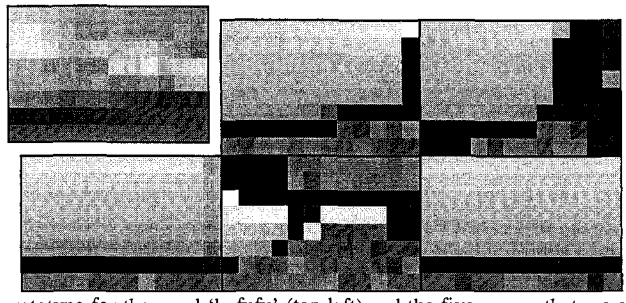


Figure 3 The prototype for the word 'kufufu' (top left) and the five scenes that are associated with this word in a language with 27 unique words. Most of the scenes associated with 'kufufu' show a white wall with a black strip, although the bottom middle scene has different features.

The agents evolved with processed pose cells as inputs produced languages with an average of 10.9 words. The average number of scenes correctly categorized by the agents was 58.7 out of 100. These languages had less words associated with single input patterns and more words associated with many input patterns spread across the entire space. The larger languages had more words associated with groups of input patterns that were close together in space.

Discussion: Expressivity is an important feature of language, where unique words are used for unique meanings. In this simulation, expressivity is indicated by the number of unique words. The vision and pose cell representations resulted in languages with an average of over 20 unique words for the 100 input patterns, while the processed pose cell representation resulted in languages with an average of 10.9 unique words. This reduction in expressivity for the processed pose cell representation indicates that the unique information in the input patterns may be lost when the pose cell representation is processed.

The number of categories correct indicates how well the language categorizes the world. The processed pose cell languages were most successful at clustering input patterns that were close together in space, with distinct clusters associated with single words. The unprocessed pose cell languages were not as successful at categorizing the patterns, which may be due to the size and sparseness of the pose cell representation, and can be addressed by processing the pose cell representation.

Some of the agents using languages evolved with vision were successful in grouping together similar scenes, however many of the words in the vision languages grouped together images that were dissimilar as well as similar, or were associated with single images. In this study, raw vision as an input provided a structure that allowed some languages to evolve to successfully categorize the

world. Processing the scenes prior to the language agent may extract the important information from each scene that is necessary for languages to consistently evolve with expressivity and categorization.

4. General discussion and conclusion

The RatChat project aims to explore the structure of languages that describe space using mobile robots. The simulations presented in this paper represent agents developing their internal representations of the world prior to playing naming games in populations of agents, and have provided insight into the expressivity, categorization, and structure of languages that can evolve from visual and pose cell representations.

There is a tradeoff between expressivity, with unique words for unique meanings, and categorization, with the use of one word for a group of similar meanings. The degree of expressivity and categorization can be altered by processing the inputs, as can be seen with the pose cell representation: the unprocessed languages are more expressive, while the processed languages are better at categorizing the world.

We are currently running simulations to scale up these results with further studies into processing the robot representations prior to the language networks and evolving languages in populations of agents.

Acknowledgements

We thank members of the RatSLAM team Michael Milford, David Prasser, Shervin Emami, and Gordon Wyeth. This research is funded in part by a grant from the Australian Research Council.

References

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. J. Briscoe (Ed.), *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*. Cambridge, UK: Cambridge University Press.

Beyer, H.-G., & Schwefel, H.-P. (2002). Evolution Strategies: A comprehensive introduction. *Natural Computing, 1*, 3-52.

Cangelosi, A., & Harnad, S. (2001). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication, 4*(1), 117-142.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42*, 335-346.

Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert & R. Conte (Eds.), *Artificial Societies: The Computer Simulation of Social Life*. London: UCL Press.

Kohonen, T. (1995). *Self-organizing maps*. Berlin: Springer.

Levinson, S. C. (1996). Language and Space. *Annual Review of Anthropology, 25*, 353-382.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Science, 8*(3), 108-114.

Marocco, D., Cangelosi, A., & Nolfi, S. (2003). The role of social and cognitive factors in the emergence of communication: experiments in evolutionary robotics. *Philosophical Transactions of the Royal Society London - A, 361*, 2397-2421.

Milford, M. J., Wyeth, G. F., & Prasser, D. (2004). RatSLAM: a hippocampal model for simultaneous localization and mapping. In *IEEE International Conference on Robotics and Automation (ICRA 2004)*: IEEE Press.

Prasser, D., Wyeth, G. F., & Milford, M. J. (2004). *Biologically inspired visual landmark processing for simultaneous localization and mapping.* Paper presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai.

Roy, D. (2001). Learning visually grounded words and syntax of natural spoken language. *Evolution of Communication, 4*(1), 33-56.

Rumelhart, D. E., Widrow, B., & Lehr, M. A. (1994). The basic ideas in neural networks. *Communications of the ACM, 37*(3), 87-92.

Steels, L. (1999). *The Talking Heads Experiment* (Vol. I. Words and Meanings). Brussels: Best of Publishing.

Steels, L. (2005). The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Connection Science, 17*(3-4), 213-230.

Tonkes, B., Blair, A., & Wiles, J. (2000). Evolving learnable languages. In S. A. Solla, T. K. Leen & K.-R. Muller (Eds.), *Advances in Neural Information Processing Systems 12*. Boston: MIT Press.

Vogt, P. (2000). Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication, 4*(1), 87-116.

WHY TALK? SPEAKING AS SELFISH BEHAVIOUR

THOM SCOTT-PHILLIPS

Language Evolution and Computation research unit, University of Edinburgh, George Square, Edinburgh, EH8 9LL, UK

Many theories of language evolution assume a selection pressure for the communication of propositional content. However, if the content of such utterances is of value then information sharing is altruistic, in that it provides a benefit to others at possible expense to oneself. Close consideration of cross-disciplinary evidence suggests that speaking is in fact selfish, in that the speaker receives a direct payoff when successful communication takes place. This is congruent with the orthodox view of animal communication, and it is suggested that future research be conducted within this context.

1. Introduction

1.1. *The Neglect of Pragmatics in Theories of Language Evolution*

The generative emphasis on the transfer of propositional information as the defining trait of language has meant that other features – particularly pragmatic ones – have sometimes been neglected in the study of its origins. For example, Hauser and Fitch's attempt to define the uniquely human aspects of language (2003) makes no mention at all of pragmatics. Hauser, Chomsky and Fitch (2002) do similarly, and nowhere in Bickerton's self-styled introduction to the field (in press) does he consider the relevance of pragmatics to evolutionary accounts of the language faculty. For many researchers in the field of language evolution, pragmatics appears not to be a foundational issue.

Yet it is necessarily core. If language were approached anew, from a Darwinian standpoint, then the first questions we might ask would arguably be about linguistic *use*; in other words, pragmatics. As one prominent evolutionary psychologist has put it: "The issue here is a purely empirical one. How *do* we use language?" (Dunbar, 2004, italics in original). Despite its importance, this fundamental question is little addressed, let alone answered.

1.2. *The Illusion of Linguistic Communism*

In asking just such questions about conversational behaviour a paradox emerges. Pinker and Bloom (1990) argue that language evolved in response to pressures of communicative efficiency, the adaptiveness of which is clear: pooled knowledge

will usually result in better outcomes for all. However, it is equally true that in such an environment there is scope for a selfish individual to listen as much as possible, and thereby acquire information, but not to speak, since doing so may dilute the value of the information held. Such an individual would prosper; she can make use of knowledge held by others at no cost to herself.

Yet we do not pursue such a strategy. On the contrary, we are a species that is motivated to speak. In the words of one researcher, we have a "robust and passionate urge of some kind to communicate" (Bates, 1994, p.139). Although some individuals talk more than others, nobody is obstinately silent. In contrast, efforts to teach language to non-human primates often suffer from the primate's lack of motivation to use what they have learnt, unless food or some other stimulus is provided: "monkeys and apes rarely seem to 'donate' information... there is little evidence... that primates use their voices *in order to inform*" (Locke, 2001, p.39, italics in original). Humans could hardly be more different. Our willingness to tell others things we think worthy of comment is taken for granted. Even pre-linguistic human infants seem keen to convey illocutionary content; lacking words, they use intonation instead (Ninio & Snow, 1996). The fact that we willingly and pro-actively converse with each other – and thereby, supposedly, provide listeners with the valuable currency of information – presents a challenge to adaptationist theories of language evolution that assume communicative efficiency is/was the overriding selection pressure. This paradox has been termed "the illusion of linguistic communism" (Bourdieu, 1991, p.43).

1.3. *Talk as Altruistic Behaviour*

Miller has expressed the same problem another way: "The trouble with language is its apparent altruism" (2000, p.346). Although both the usual explanations of altruism – *inclusive fitness* (Hamilton, 1964) and *reciprocal altruism* (Trivers, 1971) – have been proposed as the solution or partial solution to the problem (e.g. Fitch, 2004; Pinker, 2003) they cannot tell the whole story. The first says nothing about our apparent willingness to share information with non-kin. The second depends upon efficient policing (see, e.g. Fehr & Gächter, 2002), yet the one-to-many nature of conversation ensures that the social balance sheet of all but the most introverted individuals will be permanently in the red. Moreover, a range of cross-disciplinary evidence exists that, taken together, suggests not only that the speaker benefits from conversation, but that, in fact, they receive direct benefit from speaking. If this is true, then speaking contains a direct pay-off, over-and-above any desire to communicate. Thus, a solution to the paradox is offered:

sharing information would no longer be altruistic; it would, instead, be a selfish act that happens to benefit the listener at the same time.

In fact, brief consideration of our everyday experience of language is suggestive: "People compete to say things. They strive to be heard... those who fail to yield the floor... are considered selfish, not altruistic. Turn-taking rules... regulate not who gets to listen, but who gets to talk" (Miller, 2000, p.350). These observations are hard to explain within an altruistic framework. On the contrary, they appear decidedly selfish. If that were not the case then we would not compete to be heard (at least not to the same degree), yielding the floor would be selfish, and turn-taking rules would regulate whose turn it is to receive valued information. Of course, all this leaves open the question of what the benefit to the speaker might be. Dessalles (1998) suggests that it is status; Miller (2000) and Burling (2005) cite sexual selection. Other propositions can be imagined. Here, however, that question is deferred; instead the focus is simply on the evidence that speaking is a selfish act. That evidence comes from three distinct fields: evolutionary psychology, anatomy and computational modelling.

2. Speaking as Selfish Behaviour

2.1. *Evolutionary Psychology*

The central tenet of evolutionary psychology is that our brains are evolved organs that are susceptible, as all organs are, to the pressures of natural selection. Consequently, our innate psychological tendencies leave us suitably-equipped to deal with the challenges of complex social interaction as they were encountered in the environment in which we have evolved. One well-attested example of such wisdom is the existence of strategies for detecting social cheats: problems contextualised in terms of a social contract are far easier to solve than those expressed in any other terms (Cosmides, 1989). For example, when asked which facts are relevant to the preservation of the rule "If you take a pension then you must have worked here ten years" subjects will, if asked to put themselves in the position of the employ*er*, pick out the correct answers. However, when asked to consider the matter as though an employ*ee*, sentences like "worked here twelve years" and "did not get a pension" – phrases that do not inform the question being asked – are deemed relevant (Gigerenzer & Hug, 1992). The headline conclusion from a series of such experiments is that we have a mind that "includes cognitive processes specialized for reasoning about social exchange" (Cosmides, 1989, p.187, but see Gray, 2003 for a different view). We should therefore be able to

draw conclusions about the nature of behaviour from the presence of such mechanisms. That is, by reverse engineering from the situations in which we suspect and detect deception, we can deduce the form of our social contract.

From this perspective, two observations are telling. The first is that introversion – listening but doing little speaking – is not a conversational offence. Quiet individuals are able to collect information from others without reciprocation, yet the assumption that the listener is the main beneficiary would predict the opposite. Thus, we should expect to find psychological mechanisms geared to detecting and ostracising individuals that remain silent during conversation. In contrast, one particular form of speaking – lying – *is* frowned upon. If we may characterise lying as talking on false premises, then it can be understood in selfish terms: as attempting to gain whatever payoff is on offer in conversation without concern for truth. As such, the psychology of conversational behaviour suggests that speaking is a selfish act.

2.2. *Anatomy*

Brief consideration of anatomical data suggests that selection has acted more on our ability to speak than it has on our ability to listen and thus, supposedly, to acquire information. Put simply, our ears are little evolved from primates whereas our vocal tracts have evolved significantly since the last common ancestor (Lieberman, 1984). Indeed, they are more developed than is necessary in order to produce unambiguous utterances. In fact, the vocal tract is massively redundant if we assume its purpose is the production of evermore unambiguous utterances. Even in a language with relatively few distinct phonemes the potential number of, say, four-syllables words that a human can produce is far greater than the number of words in the average lexicon. For example, Hawaiian, on some measures, has a particularly small phonological set of just eight consonants and four vowels. Yet even here, a consistent CV syllable structure produces 8x4=32 possible two-phoneme words, 32^2=1,024 possible four-phoneme words, 32^3=32,768 possible six-phoneme words and 32^4=1,048,576 possible eight-phoneme words. At the other extreme, a language with, say, 20 vowels or diphthongs and 24 consonants (as the southern British English accent has) and CV syllable structure would have 20x24=480 syllables and 480^2=230,400 four-phoneme combinations. Estimates of the size of an individual's lexicon are typically in the 50,000 to 75,000 range (e.g. Oldfield, 1966; Pulvermüller, 1999), and many words are much longer than four phonemes anyway. The full range of linguistic content could still be produced with a vastly simplified vocal tract. Though it has been suggested that the larynx

may have descended in *Homo sapiens sapiens* for reasons other than speech (Fitch, 2000), this does not in itself explain further evolutionary developments.

In contrast, no similar development of redundancy is observed in our ears: background noise remains just that, whereas a pressure to consume information would be expected to produce a catch-all listening device. However, we have not evolved ear trumpets as part of our anatomy (Miller, 2000, p.350-351). The situation is summarised thus: "human languages are adapted to general mammalian perceptual capabilities... [whereas] human speech has clearly evolved with the *production* of language as its primary adaptive context" (Tomasello & Bates, 2001, p.3, italics added).

2.3. *Computational Modelling*

Finally, a computational model (Hurford, 2003) gives us further evidence that natural selection acted on our ability to communicate rather than interpret. Here, agents engage in communicative tasks with one speaker and one hearer. Agents' abilities were evolved using a genetic algorithm, and the basis for selection was set to either *communicative* or *interpretative* success. In the former case, the languages that emerged were those in which synonymy was rare and homonymy tolerated, just as is observed in virtually all recorded languages. In contrast, when *interpretative* success was used as the basis for selection then the converse situation – unknown in natural language – arose: homonymy was rare and synonymy tolerated. As Hurford concludes, and as we have now seen in a variety of different ways: "humans evolved to be well adapted as senders of messages; accurate reception of messages was less important... *we may be primarily speakers, and secondarily listeners*" (p.450, italics added). This is because, it is suggested, the greater payoff in most conversational interaction is available to the speaker rather than the hearer.

3. Concluding Remarks – Marrying Animal Communication with Pragmatic Behaviour

Implicit in the orthodox evolutionary view of animal communication is that it is, typically, a selfish act. Signallers emit signals in order to manipulate the behavioural machinery of receivers, and receivers evolve behavioural mechanisms – characterised as mind-reading – that allow them to make the best use of any observed behaviour of the signaller (Krebs & Dawkins, 1984). Thus, a signal becomes so only when the receiver makes use of it as such; to the receiver, there is

no meaningful difference between signals intentionally produced by the signaller and any other observation they may make of the signaller's behaviour.

It is probably no coincidence that this view of animal communication maps well onto the pragmatic notion of inference. Where listeners infer meaning, they are, in the terminology of animal communication, reading a mind: they use the utterance to gain an insight into the speaker's intended meaning (Origgi & Sperber, 2000). It seems reasonable to propose, similarly, that when giving a signal – that is, making an utterance – speakers are trying to manipulate the behaviour of others. Certainly, given the clues reviewed above, more detailed examination of language as selfish manipulation is merited.

Although, as already mentioned, some researchers have proposed individual payoffs to speaking, it is surely more likely that the payoffs will take a wide variety of forms. Increased status within the group (Dessalles, 1998) is likely to be a payoff in some scenarios, and greater sexual opportunity (Miller, 2000; Burling, 2005) in others. But in other circumstances neither of these will apply. Rather than see such examples as exceptional, it seems more appropriate to conceive of all signalling (linguistic or otherwise) in the terms of animal communication systems: as attempts to manipulate the behaviour of others. For example, in issuing the utterance "Make me a cup of tea" I am attempting to manipulate their body of the listener so as to perform an act on my behalf. Whether or not the imperative is obeyed is a function of their ability to infer my state of mind – that is, to mind-read (a straight-forward task in this example, since I have made my state of mind explicit, though this would not necessarily be the case in a more complex example) – and of whether they consider it in their interest to comply.

Exploration of how well this perspective of human language is congruent with traditional accounts of pragmatic behaviour is surprisingly little-addressed by language evolution researchers. This is especially true given that it provides the individual variation – in the form of one's ability to engage in mind-reading and manipulation – that is the fuel of natural selection. From an evolutionary perspective, we are better to conceive of language in the same essentially selfish terms as animal communication. The alternative, naïve assumption that language is used to transfer propositional content leads to a series of arguments that the present analysis suggests are unlikely to be true: that, by listening to new and relevant information, listeners receive most, if not all, of the benefit from conversation, and thus that in order to explain our willingness to communicate we must find some justification for massive reciprocated altruism in language use. As

we have seen, this seems unlikely. We are better to conceive of human communication in just the same way as we do the communication of any other animal: as the product of selfish attempts to manipulate and mind-read the behaviour of others.

References

Bates, E. (1994). Modularity, domain specificity and the development of language. *Discussions in neuroscience, X*, 135-156.

Bickerton, D. (in press). Language evolution: A brief guide for linguists. *Lingua*

Burling, R. (2005). *The talking ape: How language evolved.* Oxford: Oxford University Press

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187-276

Dessalles, J-L. (1998). Altruism, status and the origin of relevance. In J. R. Hurford, M. Studdert-Kennedy and C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 130-147). Cambridge: Cambridge University Press.

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of general psychology, 8*(2), 100-110

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137-140

Fitch, W. T. (2000). The evolution of speech: A comparative review, *Trends in cognitive science, 4*. 258-267

Fitch, W. T. (2004). Kin selection and "mother tongues": A neglected component in language evolution. In D. K. Oller and U. Griebel (Eds.), Evolution of communication systems: A comparative approach (pp. 275-296). Cambridge, Mass.: MIT Press.

Gigerenzer, G. and Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43*, 127-171

Gray, R. D. (2003). Evolutionary Psychology and the challenge of adaptive explanation. In K. Sterelny and J. Fitness (Eds.), *From mating to mentality: Evaluating Evolutionary Psychology* (pp. 247-268). London: Psychology Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and semantics, vol. III, Speech acts* (pp. 41-58). New York: Academic.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of theoretical biology, 7*, 1-52

Hauser, M. D., Chomsky, N. and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569-1579

Hauser, M. D. and Fitch, W. T. (2003). What are the uniquely human components of the language faculty. In M. H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 158-191). Oxford: Oxford University Press

Hurford, J. R. (2003). Why synonymy is rare: Fitness is in the speaker. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim and J. Ziegler (Eds.), *Advances in artificial life - Proceedings of the 7th European Conference on Artificial Life (ECAL), lecture notes in artificial intelligence, Vol. 2801* (pp. 442-451). Berlin: Springer Verlag

Krebs, J. R. and Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs and N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 380-402). Oxford: Blackwell.

Lieberman, P. (1984). *The biology and evolution of language.* Cambridge, MA: Harvard University Press.

Locke, J. L. (2001). Rank and relationships in the evolution of spoken language. *Journal of the Royal Anthropological Institute, 7*, 37-50

Ninio, A. and Snow, C. E. (1996). *Pragmatic development.* Boulder, CO: Westview Press.

Miller, G. F. (2000). *The mating mind: How sexual choice shaped the evolution of human nature.* London: Vintage.

Oldfield, R. C. (1966). Things, words and the brain. *Quarterly journal of experimental psychology, 18*, 340-353

Origgi, G. and Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers and A. Chamberlain (Eds.), *Evolution and the human mind: Language, modularity and social cognition* (pp. 140-169), Cambridge: Cambridge University Press

Pinker, S. (2003). An adaptation to the cognitive niche. , In M. H. Christiansen and S. Kirby (Eds.), Language evolution (pp. 16-37). Oxford: Oxford University Press.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection. *Behavioral and brain sciences, 13*, 707-784

Pulvermüller, F. (1999). Words in the brain's language. *Behavioural and brain sciences, 22*, 253-336

Tomasello, M. and Bates, E. (2001). General introduction. In M. Tomasello and E. Bates (Eds.). *Language Development: The essential readings* (pp. 1-11). Oxford: Blackwell.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly review of biology, 46*, 35-57

SEMANTIC RECONSTRUCTIBILITY AND THE COMPLEXIFICATION OF LANGUAGE

ANDREW D. M. SMITH

Language Evolution and Computation Research Unit,
School of Philosophy, Psychology and Language Sciences, University of Edinburgh,
Adam Ferguson Building, 40 George Square, Edinburgh, EH8 9LL, UK
andrew@ling.ed.ac.uk

Much of the current debate about the development of modern language from protolanguage focuses on whether the process was primarily synthetic or analytic. I investigate attested mechanisms of language change and emphasise the uncertainty inherent in the inferential nature of communication. Both synthesis and analysis are involved in the complexification of language, but the most significant pressure is the need for meanings to be reconstructible from context.

1. Introduction

Grammaticalisation is the historical genesis and subsequent development of linguistic functional categories, such as prepositions and case markers, from earlier lexical items such as nouns and verbs. It is often accompanied by phonetic loss, and is regularly characterised by semantic bleaching and generalisation, or the loss of some specificity of meaning, and the use of a form in new, broader contexts. Despite the existence of some counter-examples (see Newmeyer (1998)), grammaticalisation is widely recognised as being an overwhelmingly unidirectional process. Heine and Kuteva (2002a) have proposed, therefore, that we can make use of this unidirectionality to find insights into the nature of early human language. Hurford (2003) has gone further, and suggested that we need posit the existence of only verbs and nouns, and that auxiliaries, prepositions and all the functional paraphernalia of modern language can be derived through well-understood grammaticalisation processes.

At the same time, there is currently a lively debate in the literature concerning the structure of early human language (or protolanguage) itself (for example, see Tallerman (in press)). Protolanguage is often characterised either as a "slow, clumsy, ad hoc stringing together of symbols" (Bickerton, 1995, p.65), or as being "composed mainly of 'unitary utterances' that symbolized frequently occurring situations... without being decomposable into distinct words" (Arbib, 2005, p.108). These accounts lead themselves to opposing visions of the process through which modern language developed from protolanguage: either through

a *synthetic* process in which increasing numbers of words are concatenated to express increasingly complex propositions, or through an *analytic* process of segmentation (Wray, 2000), where the unitary utterances are divided into meaningful sub-units and rules which govern their recombination are created.

Very little of this debate, however, is concerned with how protolinguistic utterances would actually have been used and understood by early humans. In this paper, I aim to redress this omission, by exploring the uncertainty of meaning construction in an inferential communicative system. The development of protolanguage into modern human language, and the complexification of language more generally, can only occur when language users can successfully communicate even while they maintain different internal representations of language. I propose that a focus on meaning inference and reanalysis provides us with exactly this scenario, where stable variation in linguistic structure leads to significant language change (Smith, forthcoming). In section 2, I discuss these processes in more detail, explore the inferential nature of the communicative process, and introduce the concept of semantic reconstructibility. In section 3, I explore the effect that semantic reconstructibility has on the replication of linguistic structures in a hypothetical protolanguage, and finally suggest why the inferential reconstructibility of semantic structure holds the key to the complexification of language.

2. Grammaticalisation Processes

Metaphorical innovation has long been identified as having a major role in the creation and maintenance of concepts, and in semantic change more generally (Trask, 1996; Deutscher, 2005). Metaphors are normally considered in terms of mappings across conceptual domains, and are crucially not random, but motivated by analogy and iconicity (Hopper & Traugott, 2003), and the desire to express abstract concepts by building on socially-constructed semantic schemas. Lakoff (1987), for instance, shows how English has a large range of expressions relating to anger, which are built on various metaphors comparing anger to heat in a container, fire, and a dangerous animal, among others. Cross-linguistically and historically, one of the most pervasive metaphorical schemas is the conversion of spatial terms into temporal terms (Haspelmath, 1997). In English, this can be seen through numerous examples such as the spatial prepositions *behind* and *around* being used both spatially, as in 'behind the house' and 'around the fire', and also temporally, in phrases such as 'behind schedule' and 'around noon'.

More interestingly from a grammaticalisation point of view is the derivation of spatial prepositions themselves, which, in languages throughout the world, consistently develop from an apparently universal metaphorical extension of the relative location of parts of the human body. Heine and Kuteva (2002b), for instance, have collected many such examples from languages across the world, two of which are repeated here for illustration:

(1) a. stomach → in (Mixtec)

ni- kāžáa ini ndúčá
CPL- drown *stomach* water
Someone drowned *in* the water.

b. breast → in front of (Welsh)

ger fy mron
near my *breast*
In front of me.

Reanalysis, on the other hand, occurs when the structure of an utterance which the hearer infers is different from that which the speaker originally intended. For example, the Latin phrase *clara mente* initially meant 'with a clear mind', and was used as a descriptive adverbial phrase. Later, it was reinterpreted to mean 'in a clear manner', and this reanalysis led to its being used in other, non-psychological contexts, and eventually to modern French adverbs such as *lentement* 'slowly' and *doucement* 'sweetly' (Hopper & Traugott, 2003). Over time, the noun *mente* has been grammaticalised into a generalised derivational morpheme *-ment* which can now be attached to almost all French adjectives.

2.1. *The Communicative Process*

It is reasonable to characterise communication as the transfer of some information from a speaker to a hearer, but it is important to recognise that this information is not transferred directly, but indirectly. The speaker wants to convey a meaning, and chooses an utterance which represents this meaning. The hearer, on the other hand, must *infer* a meaning, from pragmatic insights and the wider context in which the utterance is used, and attempt to *reconstruct* the speaker's original meaning. Communication succeeds when this reconstruction succeeds.

This inferential process of meaning reconstruction, however, is fraught with uncertainty, as famously shown by Quine (1960). Individuals can therefore not be certain of inferring exactly the same meanings as each other. The inevitable reanalyses of utterances which take place during meaning construction cause the development of (slightly) divergent internal linguistic representations. Fortunately, however, there is a degree of slack in the communication process as well: it is not usually necessary for the hearer to reconstruct the original meaning *exactly*, in order for the communication to succeed sufficiently. Latin speakers, for instance, could happily use *clara mente* to mean either 'with a clear mind' or 'in a clear manner' in most contexts without any fear of confusion, because only rarely would any significant difference arise.

Speakers and hearers play different roles in the development of a negotiated, language-like, communication system: although utterances are produced by

speakers, their meanings must be successfully reconstructed by hearers if they are to be replicated in future communicative episodes and generations (Croft, 2000). Utterances which cannot be interpreted by hearers will neither succeed in communication nor be replicated. Metaphorical innovation, then, is a speaker-driven innovation, deriving from the speaker's desire to express concepts which lack words. A speaker will not merely invent a random expression, which is unlikely to be understood by the hearer, but will build on an existing system, extending it systematically and predictably, so that the hearer will be able to reconstruct the appropriate meaning from the social and linguistic context. Reanalysis, however, is the unconscious yet inevitable result of the uncertainty involved in the hearer's inferential reconstruction of meaning. As long as the communicative episode succeeds sufficiently, the hearer cannot verify that their reconstructed meaning is exactly the same as the speaker's, and so different representations will inevitably co-exist. Certain kinds of pragmatic inferences are more likely to be made in this process than others (for instance, the inference that travelling somewhere to do X implies that X will happen in the future), and therefore the same kinds of reanalyses will recur, both cross-linguistically and historically. The internal nature of meaning reconstruction, moreover, means that divergent reanalyses can remain hidden in internal linguistic representations for some time, with individuals communicating through utterances which they map to slightly different meanings.

Inferential communication, therefore, and the negotiation and reconstruction of meaning at its heart, results naturally in systematic changes in mappings between utterances and meanings. In order for any utterance to be replicated, it must be able to be reconstructed by hearers; all speaker-driven innovations are therefore tempered by the over-arching need that they be able to pass the test of the inferential reconstructibility of meaning.

3. Holistic Protolanguage

What does the requirement for semantic reconstructibility imply, then, for the nature of early human language? Wray (2000) models the evolution of language from a holistic ancestor through a segmentation process. Example 2 shows part of her hypothetical initial holistic language, in which arbitrary forms are coupled with arbitrary meanings.

(2) a. tebima
give-that-to-her

b. kumapi
share-this-with-her

Neither the forms nor the meanings are initially segmented in any way, so the whole of the utterance corresponds to the whole of the meaning. Language users, however, have the potential to *analyse* their mappings, and so take advantage of

coincidental correspondences between parts of utterances and parts of meanings. For instance, the language user may notice the chance correspondence between the segment *ma* in the utterances and the meaning component 'her', and modify their internal representation to something like that shown in example 3.

(3) a. tebi X
give-that-to Y

b. ku X pi
share-this-with Y

c. X = ma
Y = her

Over time, repeated segmentation leads to a system of word-like sub-units and linguistic rules governing their recombination. Kirby (2002) and others have used computational simulations to demonstrate the emergence of compositional language from a holistic ancestor using this very technique. However, it has also been recognised (Smith, 2003) that the form of the resultant 'emergent' syntax in such models is effectively predetermined by the explicit coupling of utterances and meanings, the initial complex representation of meaning which is chosen, and the kinds of generalisations which are allowed or assumed[a].

Holistic accounts of protolanguage assume that, although the utterances are monomorphemic, they represent an entire, complex proposition, albeit initially unanalysed. Such propositions are supposedly represented in protolanguage because they are 'complex, but frequently important situations' (Arbib, 2005, p.119). Many of the semantic structures suggested in the literature, however, are even more complex than Wray (2000)'s examples; we should be very sceptical of their proposed status as 'frequently important'. Mithen, for example, suggests that early humans might have had a holistic message with a meaning like 'go and hunt the hare I saw five minutes ago behind the stone at the top of the hill' (Mithen, 2005, p.172). I suggest that it is utterly implausible that early humans would have considered such a specific proposition so frequently important that it should have its own utterance. Tallerman (in press), moreover, raises the important question of just how many such unanalysed structures an early human could be expected to memorise, though it is difficult to see how this could be conclusively answered.

More importantly, however, it is surely wholly unlikely that any hearer could possibly reconstruct such a complex meaning from context, without any help at all from the structure of the utterance, which is of course both holistic and arbitrary in

[a]For instance, if predicate-argument structure is used to represent meanings (Kirby, 2002), then the resultant syntax consists of sub-units corresponding directly to 'predicates' and 'arguments'; if meaning is represented as a multi-dimensional matrix (Brighton, 2002), then the resultant syntactic units correspond directly to the dimensions of the matrix.

its form. But without such reconstructibility, the utterance could not be replicated, and thus would become extinct almost immediately it was born. The putative semantic complexity of holistic protolanguage, therefore, seems to be on the one hand the driving force behind the analytic development of modern language, but on the other, presents a major credibility problem of semantic reconstructibility for these same holistic accounts.

3.1. *Meaning Reconstruction*

The problem may be overcome, however, if we consider what actually *is* reconstructible with any degree of accuracy from an unstructured signal. Even if it were conceivable that a speaker might wish to produce an utterance corresponding to 'go and hunt the hare I saw five minutes ago behind the stone at the top of the hill', it is not plausible to assume the hearer either 'receives' this meaning accurately, or reconstructs it to such a highly complex degree. In fact, I would suggest that hearers would only need to reconstruct the meaning to a level of detail and complexity which is sufficient for them to understand the utterance in context, and contrast it with others in their communication system. Inevitably, the meanings of protolanguage utterances would have been rather simple and easily inferred.

It may be useful to consider an analogy with the famous vervet monkey call system (Cheney & Seyfarth, 1990) at this point. The vervets make three different calls, which correspond to their noticing the presence of three different groups of predators; these situations are therefore clearly analogous to Arbib's 'frequently important situations' above. But what do their calls mean? They could correspond, in a Mithen-esque account, to very complex propositions such as 'Everybody! Quick! I think I saw an adult male snake over there by the trees where we normally eat. Let's cluster together into a big group and look in the grass!'. But in reality it's more likely that the inferred meaning will only be reconstructed to a level of detail which is just enough to allow it to be understood, and to contrast with the other utterances in the system; in fact something very simple, rather like 'snake'. Similarly, early humans are likely to infer that the meaning of Mithen's protolinguistic utterance is simply 'there's a hare' or 'I'm hungry', depending on the context in which the utterance was heard, and on the existing meanings in their communication system, from which this inferred meaning must be disambiguated. Even if we accept that early humans were capable of conceiving complex meanings, therefore, we should not assume that such complexity was needed for communication. Simple meanings, by virtue of their better reconstructibility, are much more likely to be used and to be maintained in the language.

3.2. *Complexification*

By default, therefore, the inferred meaning of protolanguage utterances would in fact be very simple, probably referential, and, crucially, reconstructible from the context in which they were uttered. As the number of utterances increased, it is

possible that the reconstructed meanings could become slightly more complex, in order to maintain contrast with the others in the system, yet still remain communicatively viable. Even a slight increase in complexity would open the door for reanalysis and segmentation, by taking advantage of coincidental co-occurrences across multiple utterance-meaning pairs, as Wray (2000) describes.

The involvement of synthetic processes, however, cannot be ruled out. Unless there is a very strict convention of role-taking, indeed, natural discourse processes will ensure that consecutive simple utterances are inevitably concatenated together and processed as a whole by hearers, whether or not this was the speaker's intention. As always, however, the continued propagation of any such complex utterance through a linguistic community is completely dependent on the reconstruction of its meaning by the hearer. The hearer may be prompted by their existing knowledge of the meanings of the two individual (sub-)utterances to reconstruct a combined, complex meaning for the whole, or they may reconstruct a simple meaning, and thus lose the potential innovation introduced by the speaker. At some point, however, some useful and slightly more complex meanings may well become established in the negotiated system. Coincidental co-occurrences will allow such meanings to be eventually decomposed into their sub-parts, and then the resulting constructions can be analogically and metaphorically extended, to be used in other utterances. If the compositional constructions are productive, and their meanings remain reconstructible, then they will be replicated faster than holistic mappings (Kirby, 2002), and a structured system will develop.

4. Summary

Utterances are produced by speakers, but their replication depends on them being reconstructible by hearers. Any speaker-led innovations in language, therefore, must be as predictable and natural as possible, building on analogy, iconicity, and existing socially-constructed schemas. Both synthetic and analytic processes are implicated in the development of modern languages from ancestral protolanguage. The most significant pressure, however, comes from the need for meanings to be inferable, and reconstructible from context.

Acknowledgements

Andrew Smith is funded by AHRC grant AR112105.

References

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*(2), 105–124.

Bickerton, D. (1995). *Language and human behavior.* Seattle: University of Washington Press.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*(1), 25–54.

Cheney, D., & Seyfarth, R. (1990). *How monkeys see the world: Inside the mind of another species.* Chicago, IL: University of Chicago Press.

Croft, W. (2000). *Explaining language change: an evolutionary approach.* Harlow: Pearson.

Deutscher, G. (2005). *The unfolding of language: an evolutionary tour of mankind's greatest invention.* New York: Metropolitan Books.

Haspelmath, M. (1997). *From space to time: temporal adverbials in the world's languages.* LINCOM Europa.

Heine, B., & Kuteva, T. (2002a). On the evolution of grammatical forms. In A. Wray (Ed.), *The transition to language.* Oxford University Press.

Heine, B., & Kuteva, T. (2002b). *World lexicon of grammaticalization.* Cambridge: Cambridge University Press.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge: Cambridge University Press.

Hurford, J. R. (2003). The language mosaic and its evolution. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 38–57). Oxford: Oxford University Press.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173–203). Cambridge University Press.

Lakoff, G. (1987). *Women, fire and dangerous things: what categories reveal about the mind.* University of Chicago Press.

Mithen, S. (2005). *The singing Neanderthals: the origins of music, language, mind and body.* London: Weidenfeld & Nicolson.

Newmeyer, F. J. (1998). *Language form and language function.* Cambridge, MA: MIT Press.

Quine, W. v. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Smith, A. D. M. (2003). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, *9*(2), 175–190.

Smith, A. D. M. (forthcoming). Language change and the inference of meaning. In C. Lyon, C. Nehaniv, & A. Cangelosi (Eds.), *The emergence and evolution of linguistic communication.* Springer.

Tallerman, M. (in press). Did our ancestors speak a holistic protolanguage? *Lingua.*

Trask, R. L. (1996). *Historical linguistics.* London: Arnold.

Wray, A. (2000). Holistic utterances in protolanguage. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language: social function and the origins of linguistic form* (pp. 285–302). Cambridge: Cambridge University Press.

THE PROTOLANGUAGE DEBATE: BRIDGING THE GAP?

KENNY SMITH

Language Evolution and Computation Research Unit,
School of Philosophy, Psychology and Language Sciences, University of Edinburgh,
Adam Ferguson Building, 40 George Square, Edinburgh, EH8 9LL, UK
kenny@ling.ed.ac.uk

Synthetic and holistic theories of protolanguage are typically seen as being in opposition. In this paper I 1) evaluate a recent critique of holistic protolanguage 2) sketch how the differences between these two theories can be reconciled, 3) consider a more fundamental problem with the concept of protolanguage.

1. Introduction

Humans have language. It is hypothesised that the common ancestor of chimpanzees and humans did not. Evolutionary linguists therefore have to explain how the gap between a non-linguistic ancestor and our linguistic species was bridged. It has become common to invoke the concept of a *protolanguage* as a stable intermediary stage in the evolution of language: "[t]he hypothesis of a protolanguage helps to bridge the otherwise threatening evolutionary gap between a wholly alingual state and the full possession of language as we know it" (Bickerton, 1995, p51).

What was protolanguage like? Under the *synthetic* account, advanced by Bickerton (see, e.g., Bickerton, 1990, 1995), protolanguage had symbols which could be used to convey atomic meanings, and these proto-words could be strung together in ad-hoc sequences. Language developed from such a protolanguage through the *synthesis* of these words into more and more complex, formally-structured utterances.

Under the (competing) *holistic* account, (see, e.g., Wray, 1998), protolanguage was a system in which individual signals, lacking in internal morphological structure, conveyed entire complex propositions, rather than semantic atoms. The transition from a holistic protolanguage to language was by a process of *analysis*, by which holistic utterances were broken down to yield words and complex structures.

Recent times have seen a number of critiques of holistic theories of protolanguage (most notably Bickerton, 2003; Tallerman, 2004, 2005). I will (briefly) review some of these criticisms in section 2. This review suggests that these com-

peting theories actually have rather different targets of explanation, and the apparent conflict between them can potentially be resolved. Such a unified account is sketched in section 3. However, this reconciliation highlights a more fundamental problem with theories which appeal to protolanguage as an intermediary stage in the evolution of language, namely that such theories are in danger of merely labelling the gap between alingual and lingual states, rather than bridging it.

2. Some criticisms of holistic protolanguage, and some responses

Bickerton (2003) and Tallerman (2004, 2005) highlight a number of potential problems with holistic protolanguage. The most thorough critical evaluation is Tallerman (2005), which provides a series of roughly 30 criticisms. I will outline and evaluate five of these here. The reader should appreciate that this is only a partial presentation and examination of Tallerman's arguments, the fuller consideration which her paper deserves requiring a rather longer treatment than this.

2.1. *Problems with learnability*

A first line of attack on holistic protolanguage is that it is not a viable communication system in its own right. I will focus on two such criticisms here. The suggestion in both cases is that *Homo erectus* (the species linked to protolanguage by Bickerton, Tallerman, and Wray) could not plausibly have learned a sufficient number of utterances to make a holistic protolanguage work.

2.1.1. *Argument 1: limited inventory size*

Tallerman's first argument to this effect is that *Homo erectus* would simply have a limited capacity for learning holistic utterances:

> "How many holistic utterances is it reasonable to assume that the hominid could learn over the course of a lifetime (of maybe 25 years)? ... [For human infants] a reasonable estimate of learning rate is an average of 9–10 words a day from 18 months onwards. Assuming that the input was a set of holistic utterances, could this feat conceivably have been matched, even approached, by the smaller-brained erectus...? I submit not."(p16–17)[a]

Is this a valid criticism? Firstly, as Tallerman herself acknowledges, it is unclear how many utterances are required to create a viable protolanguage, holistic or otherwise. This makes it difficult to evaluate how damaging this type of criticism actually is. Would a holistic protolanguage require, say, 1000 utterances to

[a]Unattributed citations refer to Tallerman (2005). Page numbers refer to the in press version of this paper, available online at http://dx.doi.org/doi:10.1016/j.lingua.2005.05.004.

work? Or is less than 1000 actually sufficient? Or less than 100? How does this correspond to the numbers required for a synthetic protolanguage?

Secondly, why not assume that the capacity of *Homo erectus* to memorise signals is approximately the same as that of modern humans, e.g. on the order of 10^4 items (although, again, we can't say if this would be enough to make holistic protolanguage viable). Tallerman discounts this possibility because of the (relatively) small brain size of *Homo erectus*.[b] In order for this to be a factor, however, we need to know what, if any, relationship exists between brain size and maximum inventory size. Jackendoff (2002, p241–242), for example, speculates that there is *no* link between brain size and capacity for lexical memorisation. Much work remains to be done if Tallerman's hunch is to be vindicated and this criticism established as significant.

2.1.2. *Argument 2: holistic signals are harder to learn*

A further factor suggested by Tallerman as reducing the maximum inventory size of a holistic protolanguage, and possibly forcing it below the (unknown) viability threshold, is that holistic lexical items are harder to learn than their synthetic counterparts:

> "whereas lexical vocabulary can be stored by pairing a *concept* with the arbitrary sound string used to denote it, holistic utterances must be stored by memorizing each complex *propositional event* and learning which unanalysable string is appropriate at each event. This task is harder"(p17).

The simple response to this argument is "why?". Why is it harder to memorise an association between a signal and an atomic concept (a predicate or argument, say) rather than a proposition involving both a predicate and an argument? Is it twice as hard to memorise the latter? Or does difficulty of learning increase exponentially with number of semantic atoms attached to lexical items? How does this putative increase in difficulty compare with the difficulty of identifying the individual semantic contribution of words in a synthetically-constructed protolanguage utterance? Tallerman offers no insight on the basis for this claim, or on any tradeoff between the two alternative tasks, or means in which it might be investigated. Without further support, this criticism seems mainly a matter of assumption.

[b] She actually offers several objections, the full quote being "could this feat conceivably have been matched, even approached, by the smaller-brained erectus, *lacking any linguistic cues, no fixed phonemic inventory, and with only the vaguest idea of the intended meaning of the holistic string*?" The proposed deficiencies are all outcomes of earlier argumentation in Tallerman (2005), and are themselves open to dispute. Given the limited scope of this paper, this argumentation will be omitted.

2.2. *Problems with analysis*

Analysis, also sometimes referred to as segmentation or fractionation, is the process by which holistic utterances are broken down into component words plus rules which govern their combination. Wray (1998) describes a scenario under which chance co-occurrences of meaning and surface form between holistic utterances lead protolanguage learners/users to segment out words, leaving behind a residual template. The accumulation of such analyses over time eventually leads to a system of words and grammatical structures. Computational models have shown that a similar process can, in principle, lead to a transition from holistic protolanguage to compositionally-structured linguistic systems (see, e.g., Kirby, 2002).[c]

Tallerman provides two arguments suggesting that a holistic protolanguage is not a plausible precursor to language — that the transition from a holistic protolanguage to language via a process of analysis would not be possible.

2.2.1. *Argument 1: The problem of counterexamples*

Tallerman states the problem as follows, classing it as "major":

> "logically, similar substrings must often occur in two (or more) utterances which do *not* share any common elements of meaning at least as many times as they occur in two utterances which *do* share semantic elements. ... The holistic scenario is, therefore, weakened by the existence of at least as many counterexamples as there could be pieces of confirming evidence for each putative word." (p19–20)

Were this accepted, we might indeed doubt any account requiring transition via analysis from holistic protolanguage to language. There are, however, two problems.

Firstly, it is not a *logical necessity* that counter-examples outnumber confirming cases for any possible segmentation — this is certainly a possibility, but we can trivially construct a case where there are *no* counter-examples to a particular segmentation. The number of counter-examples to a segmentation depends on the set of utterances under consideration, and cannot be deduced *a priori*.

[c] A frequent criticism of these models is that, typically, learners are provided with meaning-signal pairs during learning: "If the problem space were not limited in this way, the simulations simply wouldn't work — the agents would never converge on a workable system. But such unrealistic initial conditions are unlikely to have applied to our remote ancestors" (Bickerton, 2003, p86). Such comments reveal two regrettable, though common, errors. Firstly, this modelling decision does not embody an (unrealistic) assumption about "initial conditions", but rather an idealisation which allows another aspect of the process to be addressed and understood. Secondly, the fact that the analysis process works in models which make this idealisation does *not* demonstrate that analysis would not work if this idealisation was relaxed — in order to make this point, such a model must be shown not to work. This has not been done, to my knowledge.

What if *in practice* we find that, in any holistic system of a reasonable size, counter-examples tend to outnumber supporting cases? Does that mean that all possible segmentations will be blocked, and the analysis process will never get started? This depends how the analysing learner/user deals with counter-examples. One possibility, as suggested by Tallerman, is only to segment if the evidence for a given segmentation outweighs the evidence against. An alternative approach is to segment at the earliest opportunity, on the basis of local pairwise comparison (as in Kirby, 2002), in which case the number of counter-examples to a given segmentation is irrelevant.

What do human language learners do — do they weigh up the number of possible counter-examples to an apparent regularity, or do they work on purely local comparison, or do they do something more sophisticated? Tallerman offers no comment on this, nor on a more directly relevant question: what did *Homo erectus* do? Until that question can be answered (and assuming an answer is possible), we cannot use the possibility of counter-examples to argue that analysis of a holistic protolanguage is impossible.

2.2.2. *Argument 2: The problem of surface instability*

Tallerman's second criticism of the analysis process is to argue that (premise 1) the analysis process requires *consistency of expression* (forms which are underlyingly the same are recognisably the same in surface form), and (premise 2) holistic protolanguage could not plausibly exhibit consistency of expression.

Tallerman offers several persuasive arguments in support of premise 2: synchronic consistency is unlikely due to factors such as allophonic variation, and allomorphic variation in any emerging semi-analysed system; diachronic inconsistency will inevitably arise as a consequence of processes of sound change. To summarise, "variation cannot help but exist because once hominids have a vocal tract in anything approaching its modern form, then specific phonetic tendencies appear spontaneously." (p9). Premise 2 therefore seems secure. What about premise 1 — does analysis really require synchronic and diachronic consistency of expression? Tallerman's three arguments here are considerably weaker.

Her first argument is that chance similarities cannot occur in a system which does not exhibit consistency of expression: "if the emerging stems aren't consistently audible in a fixed form, how can the chance similarities ... ever arise?" (p12). This is simply incorrect: chance similarities can of course occur in such a changing system, just as they can in a system where stems are audible in a fixed form. To give a concrete example, chance similarities between the lottery draw and the numbers on your lottery ticket are possible even if you change your numbers every week.

The second argument is that inconsistency in surface form may somehow obscure the intended *meaning* of a holistic utterance: "it's even harder for the speakers to decide on an agreed holistic message for any given string, because any given

string is constantly being eroded, assimilated, and so on" (p12). This suggestion needs more support. Why does sound change inhibit the acquisition or negotiation of meaning for an utterance? Is a similar process known to occur in attested instances of language change, such that words which undergo sound change have an increased likelihood of undergoing subsequent semantic change? Given the current lack of support for this claim, we may have to remain sceptical.

The third argument has to do with the damage done by sound change: "How, then, could the fractionation have proceeded successfully over ... hundreds of thousands of years, when the material the speakers were working on was continually slipping out of their grasp, changing the validity of any hypothesis formed by one generation and demolishing the emerging system?" (p11). This is an interesting question — can analysis proceed when an emerging regularity may be obscured by sound change? There are, however, grounds to think that this final argument is also incorrect. In attested language change, paradigms which have been damaged by sound change can be repaired by analogical levelling (see, e.g., Trask, 1996, for examples). Kirby (2001) uses a computational model to demonstrates that analysis can, in principle, still work despite destructive sound change. Tallerman's premise 1 therefore seems rather shaky: analysis can derive structure from a holistic system despite synchronic and diachronic inconsistency of expression, and Tallerman's position that it cannot remains to be demonstrated convincingly.

2.3. *Uniformity of process*

Tallerman mounts a more damaging criticism of holistic protolanguage in relation to uniformity of process:

> "We have a very good idea where [for example] grammatical morphemes come from in fully-fledged language: they are formed from lexical morphemes, specifically from nouns and verbs, via the bundle of processes known as *grammaticalization* ... The null hypothesis is that the same processes were at work in the earliest forms of language ... to propose a holistic strategy involving fractionation is to ignore the known processes by which words come into being in language"(p18)

This is potentially a serious problem for holistic protolanguage, and one which its proponents must address. One possible avenue of response is to attribute the apparent discontinuity to radically different *inputs* to a single mechanism. A recent trend has been to view children's' acquisition of syntax as the conservative extraction of regularities and generalisations from utterances which are initially under-analysed (see, e.g., Tomasello, 2003). These theories of acquisition are compatible with an account of analysis of holistic protolanguage. The difference in outcomes (segmentation and analysis versus synthesis and grammaticalisation) can then be attributed to differences in input — when presented with an input

which has undergone thousands of generations of analysis already, subsequent analysers are at least likely to proceed more rapidly and further than early analysers, and may proceed in a rather different direction altogether. This possibility, however, needs to be developed considerably if it is to constitute a valid response to Tallerman's criticism.

3. Bridging the gap?

What is Tallerman's own position on the nature of protolanguage, and its role in theories of language evolution? Firstly, for Tallerman protolanguage had nouns and verbs: "Once nouns and verbs come into being, well-understood linguistic processes will do the rest" (p18). Furthermore, a theory of protolanguage is not required to explain the origins of these categories: "Nouns and verbs more or less invent themselves, in the sense that the the protoconcepts must be in existence before hominids split from the (chimpanzee) genus *Pan*" (p18). Secondly, Tallerman suggests that protolanguage users had available to them pre-grammatical ordering and grouping principles, and that the origins of such principles do not require much explanation: "Given that it is well known that apes in language training experiments can spontaneously adopt ordering ... and even parrots can be trained to pay attention to sequencing of symbols ..., it would be very surprising if our hominid ancestors did not share that same skill"(p21).

Tallerman is therefore unconcerned with the origins of words (nouns and verbs, at least) and ordering constraints. Wray offers an explanation for the origins of such features, via the analysis of a holistic protolanguage. The two theories therefore seek to explain different aspects of linguistic structure and seem to be compatible, at least potentially. To give the bare bones of one possible unified account: a holistic protolanguage undergoes analysis to deliver up nouns, verbs, and some conventionalised ordering principles; the resulting synthetic protolanguage then feeds into known processes, such as grammaticalisation, to deliver fully modern language. Insisting on an account involving only one "true" protolanguage either risks assuming away part of the phenomenon to be explained (as Tallerman does), or ignoring known process acting in the formation of linguistic structure (as Wray does).

What follows if we relax this constraint, and allow room for *two* protolanguages, rather than merely one, as in the unified account sketched above? If we admit this first subdivision of "protolanguage" into two stages, is there any reason to reject further subdivisions, reflecting the development of phonological systems, emerging paradigmatic structure, evolution of function words, and so on (as in, e.g., Jackendoff, 2002)? The division into holistic and synthetic protolanguage is then rather a simplistic one — there are other alternative labellings of the stages, based on the presence or absence of other characteristic features of language which must be explained. In this case we might reserve the single term "protolanguage" to cover the series of stages, rather than reifying any one

particular stage as *the* protolanguage.

Of course, there is no requirement that these sub-stages be strictly segregated — for example, new segmentations delivered up by analysis might enter immediately into grammaticalisation processes, while other holistic utterances and parts of utterances are further broken down. In such a scenario, where different processes overlap temporally and interact, it makes little sense to see the process as consisting of a series of two or more discrete, stable steps.

Is there still a useful place for "protolanguage" in this more pluralistic conception of the evolution of language? If there is no single, steady state corresponding to protolanguage, but rather a continuous transition to language, with multiple aspects of linguistic structure being at different stages of development and entering into different interactions at any one time, then the concept of protolanguage is not really bridging the gap between alingual and lingual states, but rather labelling it.

Acknowledgements

Kenny Smith is funded by a British Academy Postdoctoral Research Fellowship.

References

Bickerton, D. (1990). *Language and species.* Chicago, IL: University of Chicago Press.

Bickerton, D. (1995). *Language and human behaviour.* London: University College London Press.

Bickerton, D. (2003). Symbol and structure: A comprehensive framework for language evolution. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 77–93). Oxford: Oxford University Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation, 5*(2), 102–110.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173–203). Cambridge: Cambridge University Press.

Tallerman, M. (2004). *Analysing the analytic: problems with holistic theories of the evolution of protolanguage.* Presented at Evolang V.

Tallerman, M. (2005). Did our ancestors speak a holistic protolanguage? *Lingua.*

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Trask, R. L. (1996). *Historical linguistics.* London: Arnold.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication, 18*, 47–67.

HOW TO DO EXPERIMENTS IN ARTIFICIAL LANGUAGE EVOLUTION AND WHY

LUC STEELS
VUB AI Lab, Pleinlaan 2
1050 Brussels, Belgium
steels@arti.vub.ac.be, and,
SONY Computer Science Laboratory Paris

The paper discusses methodological issues for developing computer simulations, analytic models, or experiments in artificial language evolution. It examines a few examples, evaluation criteria, and conclusions that can be drawn from such efforts.

1. Introduction

The problem of the origins and evolution of language is notoriously difficult to approach in a scientific way, simply because solid data are lacking of the earliest human languages or of the neurobiological changes that enabled language. But that does not mean that scientific theorising is impossible. After all, there are many scientific fields where direct observation is not feasible, for example studies of the origins of the cosmos, and despite of this, concrete theories have been developed by analytic models, computer simulations and experiments. The same approach is possible for studying the origins of language, at least for certain aspects of this question.

In what follows, a communication system is said to be 'natural language like' if it has features such as: compositionality, marking of predicate argument structure in terms of abstract semantic roles and cases, use of perspective in the conceptualisation and marking of perspective, use of hierarchy and recursion, for example for grouping lexical items that share semantic functions (like the words in a noun phrase), use of pronouns or other elliptic expressions for reference to entities already introduced in earlier discourse, conceptualisation of events and marking in terms of Tense-Aspect-Mood systems, marking of information structure through syntax (e.g. a topic-comment distinction), etc.

The work considered here assumes that the needs for a complex communication system with such features are there and that at least the basic neurobiological machinery to configure a language faculty are there as well, but then asks *how*

a complex, natural language like communication system might develop, specifically: (i) What kind of cognitive mechanisms individuals need to develop and sustain such a system, (ii) what factors make these mechanisms relevant for communication, and (iii) by what processes the mechanisms get configured into a language faculty. It is for this type of investigation that computer simulations and experiments in artificial language evolution are appropriate, particularly if one seeks a theoretical explanation which constrains how language evolution *might* have happened.

For over a decade now, our team has been doing computer simulations and experiments in artificial language evolution to try and explain the origins of such natural-language like features, starting from agent-based models of a spatial language game (Steels, 1995), and then branching into experiments with robotic agents able to self-organize communication systems grounded in reality through their sensori-motor apparatus (Steels, Kaplan, McIntyre, & Looveren, 2002), (Steels, 2004). Other representative work is found in collections by (Briscoe, 2002), (Cangelosi & D.Parisi, 2003), (Minett & Wang, 2005), a.o. These collections also contain various attempts to develop analytic models for aspects of language evolution.

Although those engaging in these kinds of studies feel that there is steady progress with very profound results, the impact on other disciplines interested in the origins and evolution of language has so far been limited. Reactions vary from fascination and incomprehension to scepticism or downright rejection. These reactions are partly due to a lack of explanation from those of us using these approaches. It is perhaps not clear how the methodology works and why it is relevant. Moreover the criticisms are to some extent justified because the model assumptions are not always very clear or downright unrealistic, and often conclusions are drawn which are not warranted based on the models that have been proposed. This paper is intended to clarify methodological issues and sharpen the criteria for their sound application. I discuss first computer simulations, then analytic models and then experiments in artificial language evolution.

2. Computer Simulations

Three steps are involved in setting up computer simulations: (1) The researcher hypothesises that a certain set of cognitive mechanisms and external factors are necessary to see the emergence of a specific feature of language. (2) The mechanisms are operationalised in terms of computational processes, and (simulated) 'agents' are endowed with these processes, (3) A scenario of agent interaction is designed, possibly embedded in some simulation of the world. The scenario and the virtual world capture critical properties of the external factors as they pose specific communicative challenges. (4) Systematic computer simulations are performed, demonstrating that the feature of interest indeed emerges when agents endowed with these mechanisms start to interact with each other. Ideally a com-

parison is made between simulations where a mechanism or factor is included and others where it is not, in order to prove that the mechanisms or factors are not only sufficient but also necessary. This still does not prove anything about human language evolution because there may be multiple mechanisms to handle the same communicative challenges, but at least it shows a possible evolutionary pathway.

Here is one example of this approach: the Naming Game (Steels, 1995). Every human language features proper names for individual objects, and this must have been an obvious first use of language, for example to call or designate members of the group. A crucial question is then: How can a population converge on a consistent set of names for a particular set of objects, without a prior system, a central authority, or telepathy (one individual having access to the internal brain state of another one). The Naming Game studies this question by framing interactions in terms of language games. The speaker uses a name to identify some topic in the context, and the hearer guesses the topic based on the name. The game is a success if the hearer was able to identify the same topic as chosen by the speaker.

It is now known that agents can use a wide variety of strategies to play the Naming Game, each implying particular cognitive mechanisms. For example, computer simulations (as shown in figure 1) have shown that using an associate memory of object-name pairs with weights and lateral inhibition is a good strategy. For those unfamiliar with computer simulation, it is perhaps important to stress that such simulation results do not depend on a specific computer implementation nor on the programming language used, nor even on the fact that a computer is used. The simulations simply show the behavior of a dynamical system. The assumption underlying this work (which is a fundamental assumption of science) is that the properties of the dynamical system constitute an explanation of the emergent phenomenon, the same way oscillations in predator-prey populations are explained by the dynamics of the Lottka-Volterra equations and depend in no way on the specific organisms involved.

For the computer simulation to have value, some conditions must be met: (1) It must be clear what language features are supposed to be emergent and what features are assumed. It is simply not possible to explain everything at once. A lot of scaffolds in terms of assumed cognitive abilities, interaction patterns or environmental constraints must be introduced. For example, the Naming Game strategies discussed earlier assume that both agents are able to individually recognise the objects they are naming, that the hearer has a way to indicate what topic he has guessed, that agents can recognise and reproduce the names used by others, and so on.

(2) There must be no hidden 'global hand', in other words effects of global properties not observable by individual agents, nor any direct causal link between a mechanism and the feature to be explained. For example, genetic models of lexicon convergence (as opposed to cultural models as discussed above) often introduce a fitness function which is calculated in terms of the similarity of an

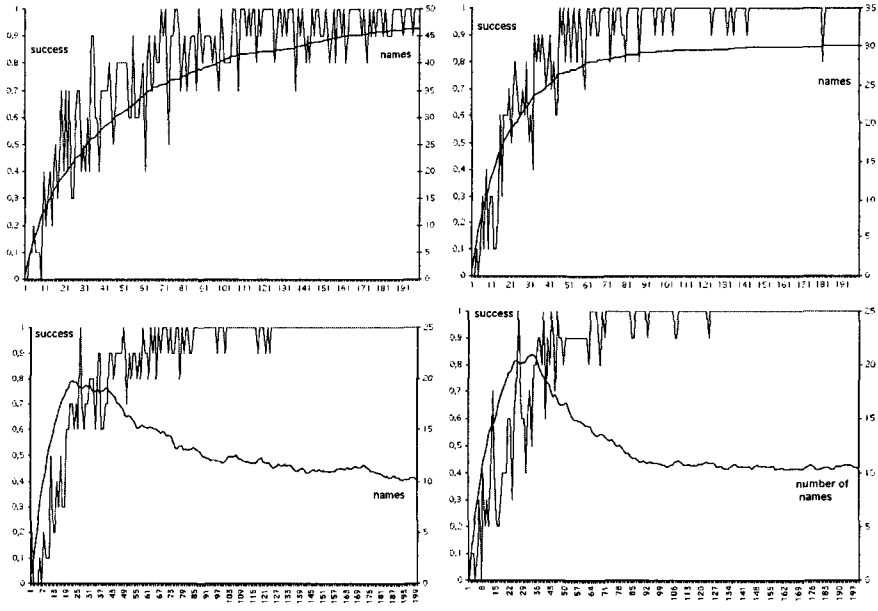


Figure 1. Effect of different strategies for playing the Naming Game. The size of the population $\mathcal{N}$ and the number of objects $\mathcal{O}$ is always equal to 10. The evolution in communicative success (left y-axis) and average inventory size (right y-axis) is shown for 2000 games (x-axis). Top left shows a strategy where agents simply adopt the word used by others. After a while everybody knows all words and hence there is complete communicative success but the inventory is large (45 words). Top right shows a strategy where success translates to enforcement (weight increase) of the word used. Success is reached more quickly and the inventory size goes down (30 words). Bottom left adds lateral inhibition (decrease of weight of competitors) and bottom right adds damping (weight decrease in case of failure). The last strategy leads to an optimal inventory (10 words) and fastest convergence, while tolerating homonymy.

agent's lexicon with that of others in the population. The computation of this fitness function requires a global view which none of the agents can have. The same sort of models also try to explain convergence by setting up a selection process that is based on greater fitness, but this fitness is calculated in terms of similarity of lexicons, in other words on how well the lexicon of the agent converges to that of the group. So there is an undesirable direct causal link between a (global) mechanism and the feature being explained.

(3) It is crucial to consider not only configurations that 'work' but also those that do not work or work less well, both to understand the causal role of each specific component integrated in the language faculty of the individual agents, and the role of parameter choices for the different mechanisms (as shown in figure 1) or the environmental factors. All this is standard scientific practice (Platt, 1964) and can be applied easily here.

3. Analytic Models

Computer simulations are an effective way to test claims about the sufficiency and necessity of certain cognitive mechanisms or how communicative challenges impact the evolution of a language, and are very valuable because it is notoriously difficult (even for computer scientists) to understand how specific computational mechanisms affect the outcome of observed (collective) behavior. But computer simulations have a major limitation. They cannot predict the general long-term behavior of a system. This is where analytic models come in. They aggregate the state of individual agents or agent behaviors by postulating global quantities with which a series of master equations is formulated. Then the standard mathematical techniques for solving these equations can be used to predict the global time course of the system. Of particular relevance is the search for scaling laws, which capture how increase in certain system parameters (for example the number of agents in the population, the number of objects they have to name, etc.) impact other system properties (such as time to reach convergence, size of the lexicon, etc.). Normally, the global quantities used in analytic models are measured by empirical observation, but, if data is missing, as in the case of language evolution, the approach can be applied to the outcome of computer simulations.

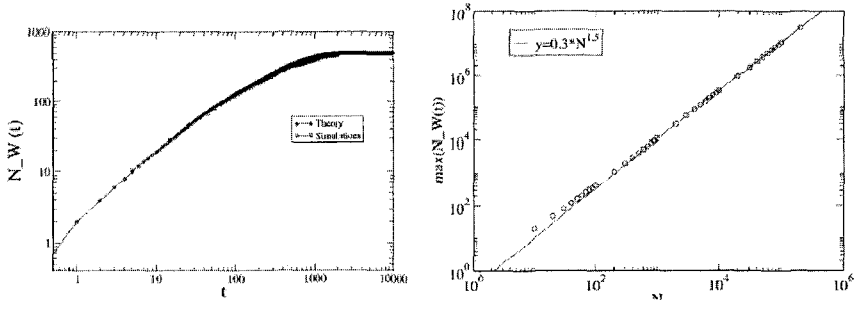


Figure 2. Very close fit between a simulation and an analytic model of the Naming Game (left). Power law behavior of the Naming Game is shown in log-log plot (right). The maximum number of words (y-axis) has a power relation with population size (x-axis) with exponent 1.5. It is not only observed in computer simulations but also predicted by the analytic model.

A recent example of this approach for the Naming Game in very large populations is discussed in (Baronchelli, Felici, Caglioti, Loreto, & Steels, 2005). It focuses only on naming one object and uses global quantities like the number of agents N_a, the total number of words at time t $N_w(t)$, the number of different words $N_d(t)$, the success rate $S(t)$, and the overlap function $O(t)$ which monitors lexical coherence in the system. It is possible to analytically predict the behavior of these global quantities from master equations using a mean field approach (fig-

ure 2(left)) and to identify power laws, such as the one shown in figure 2 (right), and prove why they have these exponents. In this type of investigation, the role of the computer is restricted to calculating the graphs that display the mathematical functions derived from the equations. These are not computer simulations. Models of agents have completely disappeared.

There are some criteria that analytic models must meet in order to be relevant: (1) The models must in one way or another relate to data, ideally from empirical sources but otherwise at least from computer simulations. Otherwise, any kind of relation can be claimed and any kind of conclusion can be drawn. Unfortunately most analytic models of language evolution that have been published so far do not meet this criterion (although the work reported above does albeit only w.r.t. simulated data).

(2) Realistic assumptions must be made about the cognitive capacities of the agents or the effects of natural or cultural selection. Human beings, as embodied autonomous agents, have strong limitations, for example, they cannot perceive the world exactly from the viewpoint of another agent and so equal perception is excluded, direct meaning-transfer is not possible, no agent can have a global overview of the language in the total population, grammar induction is always influenced by the available data, etc.

There are strong limitations to the analytic method, partly because the aggregate quantities and master equations must be found, which is very non-trivial, but more importantly because for a large number of non-linear dynamical systems (and language definitely falls into this category), no solution method is available or can ever be found. New techniques from statistical physics, such as network analysis offer nevertheless hope that much more is possible than so far achieved.

4. Experiments

Many empirical sciences use a third method for investigating natural systems, namely experiments. Normally, an experiment takes an existing natural system (for example a cell or a block of ice) and examines what happens when certain environmental parameters or system components are changed. An experiment therefore generates new data that would otherwise not be observable. The method is particularly appropriate to understand and prove which causal relations exist between the changed parameters and the observed system behavior. For example, between the surrounding temperature and the phase transitions of the block of ice into water and steam.

We might in principle invent experiments for language origins and evolution as well, although it is not so obvious. It is not possible to selectively turn on and off components in the brains of groups of humans and see the effect on the language that emerges in the group, or to make a group forget some aspect of their language (like the Tense-Aspect-Mood system) and see whether they evolve a new TAM system. Sometimes there are natural experiments: Brain disorders

due to genetics or aging may lead to language disorders. Unusual social circumstances like rapid population change in highly multi-lingual settings may give rise to new languages or language features as in creoles. But these natural experiments are generally not sufficiently controllable for being a solid basis for doing science. Quite recently some psychologists have begun to study the emergence of communication systems in dialog by constraining normal communication or creating unusual challenges (Healey, Swoboda, Umata, & Katagiri, 2002). These experiments are more controlled and yield fascinating data that are highly relevant to the question of language origins. They show for example that humans can quite quickly negotiate new communication systems and that they constantly adapt their language systems at all levels to those used by others involved in the same dialog.

However the state of the art in robotics and Artificial Intelligence makes it now possible to do non-trivial experiments with physically embodied agents (robots). Rather than selectively adding or removing components in the language faculty of humans, we do it with robots. Moreover we can control the robot's perception of the world, progressively introduce communicative challenges, control the in- and outflow of the population, the degree of noise and stochasticity in sound transmission and reception, and so on. In addition we can completely monitor both the external behavior, the emergent language system, and the internal states of the agents, even for very large populations. Such experiments in artificial language evolution have some characteristics in common with computer simulations, but they go far beyond them. Computer simulations can introduce all sorts of scaffolds and make various kinds of assumptions which can no longer be made in these experiments. For example, if we require that agents can identify objects to play the Naming Game, then we must implement the necessary perception and memory functions to achieve this - a very non-trivial task in itself. So the experiments are the most powerful and stringent way to test the realism of model assumptions.

Here is an example experiment discussed in more detail in (Steels, Loetzsch, & Bergen, in review). The experiment focuses on perspective reversal, a clear universal feature of human languages. A communication system with perspective reversal allows that a scene is conceptualised from different points of view (the speaker, the hearer, other participants, landmarks) and that the perspective is possibly marked explicitly, as in English *your left* versus *my left*. The perspective reversal experiment uses two autonomous AIBO robots that move around in search for a ball and, if they have found one, play a description game, describing to each other the movement of the ball, such as 'the ball was far away to my right and then rolled to your left' (see figure 3). The population starts without any perceptual categories (like left/right or close/far) and without any lexicon, but has to evolve sufficiently shared ontologies and lexicons to be successful in the game.

The perspective reversal experiment examines three issues: (1) Why is perspective reversal needed. It turns out that agents can develop an adequate system

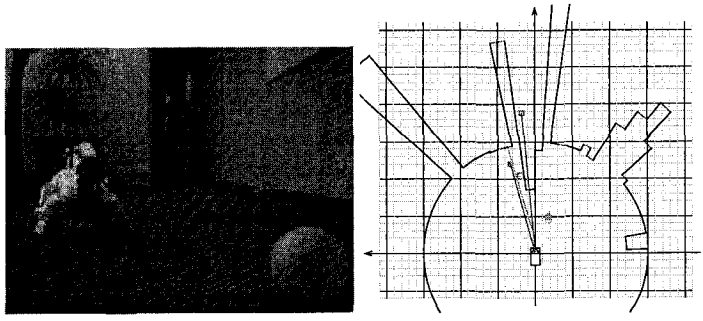


Figure 3. AIBO robot used in perspective reversal experiment (right). The dynamic world model of the robot as it is tracking the ball, obstacles and other robots. The description game is based on such world models.

if they (unrealistically) share exactly the same perception (figure 4a) but as soon as they see the world from their own perspective - which is always the case in embodied agents - their communication system collapses (figure 4b). (2) Then agents are given the ability to perform egocentric perspective transformation, which means that they can geometrically transform their perception of the world to see the scene from the viewpoint of the other agent, and they use that in conceptualisation. Communicative success goes up again (figure 4c). (3) Next agents also mark perspective which means that information flows from the egocentric perspective transformation component to the lexical component. We see that cognitive effort goes down (figure 4d). This experiment therefore demonstrates why we see perspective reversal and marking in human language: it increases communicative success in the case of embodiment and decreases cognitive effort.

As before we list a few criteria that such experiments need to take care of: (1) We first of all get the same caution as with computer simulations: It must be clear what feature is supposed to emerge and experiments must be done to compare configurations with and without cognitive components held responsable for their emergence and in different environmental circumstances (as in the experiment in figure 4).

(2) There should obviously be no global hand nor any direct causal link between mechanisms and features. It is more difficult now to make errors (compared to simulations) because certain short-cuts are no longer possible, even though scaffolds are still necessary and no problem if they do not impinge on the basic point of the experiment (for example give all robots the same perceptual system or implement the script for playing the language game).

For those not familiar with robot experiments, we need to stress that results (as those shown in figure 4) are on the one hand related to the specific embodiments (the shape, perceptual capabilities, computer processors, etc. of the robots used) however those details are not crucial, they are only instantiations of basic princi-

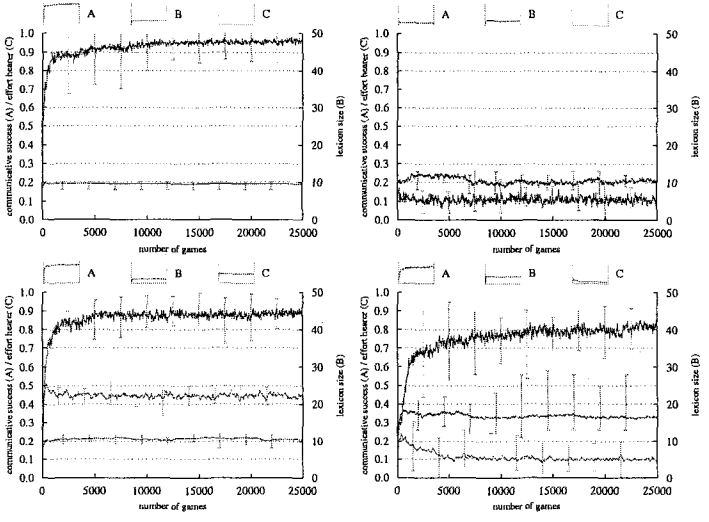


Figure 4. *Experiments in perspective reversal with same and different view on scene (top), and with* egocentric perspective transformation for conceptualisation (bottom left) and with marking (bottom right).

ples. Just like one can study a fruitfly to study genetic mutation rates in general. The same experiments could be carried out on other kinds of robots or even for other sensory domains or perceptually grounded categories, as long as the agents get differents views and hence different perceptions of the world so that perspective reversal becomes necessary. The specific implementation of the cognitive components is irrelevant, it is the functionality of the component that counts, and the experiment proves that these functionalities can be operationalised and that they can be put together in a way that effectively leads to an emergent communication system with this specific feature.

5. Conclusions

There is a growing number of computer simulations, analytic models, and experiments in artificial language evolution which shine new light on the age-old question of the origins of communication systems with the features of human natural languages. A large number of issues has not been tackled yet and we only have solid results so far for some of the most basic questions, such as how can a population develop a shared set of names. So this presents enormous opportunities for young researchers coming in the field. At the same time useful dialog is already possible and ongoing with the other approaches to language evolution, that emphasise the linguistic and anthopological data or constraints from neurobiology.

Acknowledgements

This research was funded and carried out at the Sony Computer Science Laboratory in Paris with additional funding from the EU FET ECAgents Project IST-1940.

References

Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., & Steels, L. (2005). *Sharp transition towards shared vocabularies in multi-agent systems.*

Briscoe, T. (2002). *Linguistic evolution through language acquisition: Formal and computational models.* Cambridge, UK: Cambridge University Press.

Cangelosi, A., & D.Parisi. (2003). *Simulating the evolution of language.* Berlin: Springer Verlag.

Healey, P., Swoboda, M., Umata, I., & Katagiri, I. (2002). Graphical representation in graphical dialogue. *International Journal of Human-Computer Studies, 57,* 375–395.

Minett, J., & Wang, W. S.-Y. (2005). *Language acquisition, change and emergence: Essays in evolutionary linguistics.* Hong Kong: City University of Hong Kong Press.

Platt, J. (1964). Strong inference. *Science, 146,* 347–353.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life, 2*(3), 319–332.

Steels, L. (2004). Constructivist development of grounded construction grammars. In D. Scott, W. Daelemans, & M. Walker (Eds.), *Proceedings of the 42nd annual meeting of the association for computational linguistics* (pp. 9–16). Barcelona: ACL.

Steels, L., Kaplan, F., McIntyre, A., & Looveren, J. V. (2002). Crucial factors in the origins of word meaning. In A. Wray (Ed.), *The transitions to language* (pp. 252–271). Oxford: Oxford University Press.

Steels, L., Loetzsch, M., & Bergen, B. (in review). *Why human languages mark perspective.*

THE IMPLICATIONS OF BILINGULISM AND MULTILINGUALISM FOR POTENTIAL EVOLVED LANGUAGE MECHANISMS

DANIEL A. STERNBERG
Department of Psychology, Cornell University
Ithaca, New York

MORTEN H. CHRISTIANSEN
Department of Psychology, Cornell University
Ithaca, New York

Simultaneous acquisition of multiple languages to a native level of fluency is common in many areas of the world. This ability must be represented in any cognitive mechanisms used for language. Potential explanations of the evolution of language must also account for the bilingual case. Surprisingly, this fact has not been widely considered in the literature on language origins and evolution. We consider any array of potential accounts for this phenomenon, including arguments by selectionists on the basis for language variation. We find scant evidence for specific selection of the multilingual ability prior to language origins. Thus it seems more parsimonious that bilingualism "came for free" along with whatever mechanisms did evolve. Sequential learning mechanisms may be able to accomplish multilingual acquisition without specific adaptations. In support of this perspective, we present a simple recurrent network model that is capable of learning two idealized grammars simultaneously. These results are compared with recent studies of bilingual processing using eyetracking and fMRI showing vast overlap in the areas in the brain used in processing two different languages.

1. Introduction

In many parts of the world, fluency in multiple languages is the norm. India has twenty-two official languages, and only 18% of the population is a native Hindi speaker. Half of the population of sub-Saharan Africa is bilingual as well. Though bilingualism (or multilingualism, as is often the case) has been investigated in some detail within linguistics and psycholinguistics, it has to date received scant attention from researchers studying language evolution. An extremely important issue remains undiscussed. Whatever theoretical framework one chooses to subscribe to, it is clear that the mental mechanisms used for language processing allow for the native acquisition of multiple distinct languages nearly simultaneously. What is not immediately evident is why they can be used in this way.

On the simplest level, there are two opposing possibilities: either the ability to acquire, comprehend and produce speech in multiple languages was selected for or it came for free as a by-product of whatever mechanisms we use for language. In this paper, we consider a number of the contending theories of language evolution in terms of their compatibility with bilingual acquisition. We test one particular type of general learning mechanism, namely sequential learning, which has been considered a potential mechanism for much of language processing. We propose a simple recurrent network model of bilingual processing trained on two artificial grammars with substantially different syntax, and find a great deal of fine-scale separation by language and grammatical role between words in each lexicon. These results are substantiated by recent findings in neuroimaging and eye-tracking studies of fluent bilingual subjects. We conclude that the bilingual case provides support for the sequential learning paradigm of language evolution, which posits that the existence of linguistic universals may stem primarily from the processing constraints of pre-existing cognitive mechanisms parasitized by language.

2. Potential selectionist theories

Research on bilingualism and natural selection is rather scant, thus selectionist theories on the existence of language diversity may be a good starting point for considering how a selectionist might account for the bilingual case. Interestingly, Pinker & Bloom (1990) argue against a selectionist approach to grammatical diversity, stating that "instead of positing that there are multiple languages, leading to the evolution of a mechanism to learn the differences among them, one might posit that there is a learning mechanism, leading to the development of multiple languages." This argument rests on the conjecture that the Baldwin effect leaves some room for future learning. Because the previous movement via natural selection toward a more adaptive state increases the likelihood of an individual learning the selected behavior, further distillation of innate knowledge is no longer required after a point (e.g. when the probability nears 100%).

Baker (2003) objects to the claim that the idiosyncrasies of the Baldwin Effect account for the diversity of human languages. He argues that the formidable differences in surface structure between languages should not be glossed over by reference to some minor leftover learning mechanisms. Instead, he suggests that the ability to conceal information from other groups by using a language with which they are unfamiliar could drive the creation of different languages. Like Pinker & Bloom, Baker does not directly argue for a

selectionist model of language differentiation as such, but gives a reason for language differentiation after selection for the linguistic ability has already taken place. What both theories are lacking, however, is an explanation for how this language system can not only accommodate language variation across groups of individuals, but also the instantiation of multiple languages within a single individual.

3. Sequential learning and language evolution

An alternative to the selectionist approach to language evolution can be found in the theory that languages have evolved to fit preexisting learning mechanisms. Sequential learning is one possible contender. There is an obvious connection between sequential learning and language: both involve the extraction and further processing of elements occurring in temporal sequences. Recent neuroimaging and neuropsychological studies point to an overlap in neural mechanisms for processing language and complex sequential structure (e.g., language and musical sequences: Koelsch et al., 2002; Maess, Koelsch, Gunter & Friederici, 2001; Patel, 2003, Patel et al., 1998; sequential learning in the form of artificial language learning: Friederici, Steinhauer & Pfeifer, 2002; Peterson, Forkstam & Ingvar, 2004; break-down of sequential learning in aphasia: Christiansen, Kelly, Shillcock & Greenfield, 2004; Hoen et al., 2003). We have argued elsewhere that this close connection is not coincidental but came about through linguistic adaptation (Christiansen & Chater, in preparation). Specifically, linguistic abilities are assumed to a large extent to have "piggybacked" on sequential learning and processing mechanisms existing prior to the emergence of language. Human sequential learning appears to be more complex (e.g., involving hierarchical learning) than what has been observed in non-human primates (Conway & Christiansen, 2001). As such, sequential learning has evolved to form a crucial component of the cognitive abilities that allowed early humans to negotiate their physical and social world successfully.

4. Sequential learning and bilingualism

Distributional information has been shown to be a potentially crucial cue in language acquisition, particularly in acquiring knowledge of a language's syntax (Christiansen, Allen, & Seidenberg, 1998; Christiansen & Dale, 2001; Christiansen, Conway, and Curtain, in press). Sequential learning mechanisms can use this statistical cue to find structure within sequential input. The input to a multilingual learner may contain important distributional information that

would also be useful in acquiring and separating different languages. For example, a given word in one language will, on average, co-occur more often with another word in the same language than a word in another language. Thus an individual endowed with a sequential learning mechanism might be able to learn the structure of the two languages. We decided to test this hypothesis using a neural network model that has been demonstrated to acquire distributional information from sequential input (Elman, 1991, 1993).

5. A simple recurrent network model of bilingual acquisition

We used a simple recurrent network (Elman, 1991) to model the acquisition of two grammars. An SRN is essentially a standard feed-forward neural network equipped with an extra layer of so-called "context units". At a particular time step t an input pattern is propagated through the hidden unit layer to the output layer. At the next time step, t+1, the activation of the hidden unit layer at time t is copied back to the context layer and paired with the current input. This means that the current state of the hidden units can influence the processing of subsequent inputs, providing a limited ability to deal with integrated sequences of input presented successively. This type of network is well suited for our simulations because they have previously been successfully applied both to the modeling of non- linguistic sequential learning (e.g., Botvinick & Plaut, 2004; Servan- Schreiber, Cleeremans & McClelland, 1991) and language processing (e.g., Christiansen, 1994; Christiansen & Chater, 1999; Elman, 1990, 1993).

Previous simulations of bilingual processing employing simple recurrent networks have come to somewhat opposing conclusions. French (1998) demonstrated complete separation by language and further separation by part of speech. Scutt & Rickard (1997) found that their model separated each word by part of speech, but languages were intermixed within these groupings. The languages differed in their size (Scutt & Rickard's contained 45 words compared to French's 24), however both sets contained only declarative sentences and both used only SVO grammars in their main study. We set out to create a simulation that would more realistically test the ability of this sequential learning model to acquire multiple languages simultaneously. To accomplish this, we used more realistic grammars with larger lexicons and multiple sentence types. We also chose grammars that differed in their word order system.

5.1. *Languages*

We used two grammars based on English and Japanese, which were modeled on child-directed speech corpora (Christiansen & Dale, 2001). Both grammars contained declarative, imperative and interrogative sentences. The two grammars were chosen because of their different systems of word order (SVO vs. SOV). The English lexicon contained 44 words, while the Japanese was slightly smaller (30 words) due to the language's lack of plural forms.

5.2. *Model*

Our network contained 74 input units corresponding to each word in the bilingual lexicon, 120 hidden units, 74 output units, and 120 context units[1]. The network's goal was to predict the next word in each sentence. It was trained on ~400,000 sentences (200,000 in each language). Following French (1998), languages would change with a 1% probability after any given sentence. The learning rate was set to .01 and momentum to .5.

5.3. *Results & Discussion*

To test for differences between the internal representations of words in the lexicon, a set of 10,000 test sentences was used to create averaged hidden unit representations for each word. As a baseline comparison, the labels for the same 74 vectors were randomly reordered so that they corresponded to a different word (e.g. the vector for the noun X in English might instead be associated with the verb Y in Japanese). We then performed a linear discriminant analysis on the hidden unit representations and compared the results in chi-square tests for goodness-of-fit. Classifying by language resulted in 77.0% accuracy compared to 59.5% for the randomized vectors [$\chi^2(1,n=74)=5.26$, $p<.05$]. We also created a crude grouping by part of speech. Though nouns, verbs and adjectives were easy to group, there were a number of words that served a more functional purpose in the sentence, such as determiners, common interrogative adverbs (e.g. "when", "where", "why"), and certain pronouns (e.g. "that"). We classified this set as "function" words. This part of speech classification resulted in 48.65% correct classification, compared with 35.14% for the randomized vectors, but this result was not significant

[1] One reviewer asked about the significance of the number of hidden units used in the model. Generally speaking, learning through back-propagation is rather robust to different quantities of hidden units. It is unlikely that choosing any number of hidden units slightly below or even quite a bit above the number of inputs units would yield different results other than on the efficiency of training (in this case the amount of training required to reach a proficient state).

[$\chi^2(1,n=74)=2.78$, $p=.099$]. When words were grouped by language and part of speech combined (thus creating eight categories), accuracy rose to 68.92%, compared with 17.57% for the randomized version [$\chi^2(1,n=74)=39.8$, $p<.001$]. These discriminant analysis results indicate that the net places itself in different internal states when processing English and Japanese. Importantly, the network is sensitive to the specific constraints on parts of speech within each language as indicated by the last analysis which demonstrates a highly significant difference between the trained and baseline accuracy.

These results seem to support local-scale language separation rather than the emergence of two completely distinct lexicons. Though the ambiguous "function" grouping might have created some noise in the data, grouping by language and part of speech gave a highly significant result, seeming to imply that the network attends to both language and part of speech, rather than primarily focusing on one.

6. General Discussion

The bilingual case, as the most prevalent form of language fluency in the world, must be considered in any explanation for the existence of human language. We have argued that it seems difficult to develop a selectionist account of bilingualism. In contrast, a theory of language origins and evolution via sequential learning may be more parsimonious in this regard because it seems to account for bilingualism without needing any major post-hoc revisions. Our simulation of bilingual acquisition via sequential learning demonstrated language separation at a very local scale (i.e. within part of speech and language), rather than the creation of two completely separate lexicons. Converging evidence from neurological and low-level perceptual studies of bilingual processing seem to support this finding. Recent neuroimaging data points to a great deal of overlap in the brain areas used to process different languages in fluent bilinguals (Chee et. al, 1999a, 1999b; Hasegawa et. al, 2002). Eye-tracking studies of fluent bilinguals have also demonstrated partial activation for phonologically-related words in a language not used in the experimental task (Spivey & Marian, 1999)

There are many aspects of language that need to be considered in a final model of bilingual acquisition that were not included in our first model. However, there are at the moment few contending explanations for how this ability came to exist. Our work thus far serves as a first step in demonstrating that sequential learning might be able to account for the ability to process not

only a single language as shown in previous work, but also the ability to process multiple languages simultaneously.

Acknowledgements

We thank Rick Dale for providing his *sentgen* script as well as his English and Japanese grammars, which were used to create the sentences in the simulation. We also thank Luca Onnis and three anonymous referees for their helpful comments and feedback on earlier drafts of this paper.

References

Baker, M.C. (2003). Linguistic differences and language design. *Trends in Cognitive Sciences, 7,* 349-353.

Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review, 111,* 395-429.

Chee, M.W.L., Tan, E.W.L., & Thiel, T. (1999). Mandarin and English single word processing studied with functional magnetic resonance imaging. *Journal of Neuroscience, 19,* 3050-3056.

Chee, M.W.L., Caplan, D., Soon, C.S., Sriram, N. Tan, E.W.L., Thiel, T., & Weekes, B. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron, 23,* 127-137.

Christiansen, M.H., Allen, J. & Seidenberg, M.S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13,* 221-268.

Christiansen, M.H. & Chater, N. (Eds.). (2001). *Connectionist Psycholinguistics.* Westport, CT: Ablex.

Christiansen, M.H. & Chater, N. (in preparation). *Language as an organism: Language evolution as the adaptation of linguistic structure.* Unpublished manuscript, Cornell University.

Christiansen, M.H., Conway, C.M. & Curtin, S.L. (in press). Multiple-cue integration in language acquisition: A connectionist model of speech segmentation and rule-like behavior. In J.W. Minett & W.S.-Y. Wang (Eds.), *Language Evolution, Change, and Emergence: Essays in Evolutionary Linguistics.* Hong Kong: City University of Hong Kong Press.

Christiansen, M.H. & Dale, R. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 220-225). Mahwah, NJ: Lawrence Erlbaum.

Christiansen, M.H., Kelly, L., Shillcock, R., & Greenfield, K. (2004). *Artificial grammar learning in agrammatism.* Unpublished manuscript, Cornell University.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. Trends in Cognitive Sciences, 5(12):539--546.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science, 14,* 179-211.

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71-99.

French, R.M. (1998). A simple recurrent network model of bilingual memory. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum.

Friederici, A.D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures in artificial language processing. *Proceedings of the National Academy of Sciences, 99,* 529-534.

Hasegawa, M., Carpenter, P.A., & Just, M.A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage, 15,* 647-660.

Hoen, M., Golembiowski, M., Guyot, E., Deprez, V., Caplan, D., & Dominey, P.F. (2003). Training with cognitive sequences improves syntactic comprehension in agrammatic aphasics. *NeuroReport,* 495-499.

Koelsch, S., Schroger, E., & Gunter, T.C. (2002). Music matters: preattentive musicality of the human brain. Psychophysiology, 39, 38-48.

Maess, B., Koelsch, S., Gunter, T., & Friederici, A.D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience, 4,* 540-545.

Marian, V., Spivey, M.J., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language, 86,* 70-82.

Patel, A.D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6,* 674-681.

Patel, A.D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P.J. (1998). Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience, 10,* 717-733.

Petersson, K.M, Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science, 28,* 383-407.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707-784.

Scutt, T., & Rickard, O. (1997). Hasta la vista, baby: 'bilingual' and 'second-language' learning in a recurrent neural network trained on English and Spanish sentences. In *Proceedings of the GALA '97 Conference on Language Acquisition.*

Spivey, M.J. & Marian, V. (1999). Crosstalk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science. 10,* 281-284.

SELECTION DYNAMICS IN LANGUAGE FORM AND LANGUAGE MEANING

MONICA TAMARIZ

Linguistics and English Language, The University of Edinburgh, 14 Buccleuch Place, Edinburgh EH8 9LN, UK

This paper describes evolutionary dynamics in language and presents a genetic framework of language akin to those of Croft (2000) and Mufwene (2001), where language is a complex system that inhabits, interacts with and evolves in communities of human speakers. The novelty of the present framework resides in the separation between form (phonology and syntax) and meaning (semantics), which are described as two different selection systems, connected by symbolic association and by probabilistic encoding of information.

1. Selection systems

General frameworks for complex adaptive systems, or selection systems (Gell-Mann, 1994; Hull, Langman & Glenn, 2001) fit systems as diverse as biology, immunology, the history of science, and language. *Selection* consists of iterated cycles of replication, variation and adaptation so structured that adaptation causes replication to be differential. *Replication* involves the (mostly faithful) iteration of the information contained in *replicators* (also called schemata and vehicles), which encodes the structure of the *interactors*. The principle of *variation* says that selection needs variants of the replicators to select from. These variants encode adaptations to the environment. *Adaptation* refers (a) to the effect of the developmental pressures on the replicators that affect development of the interactor and (b) to the effects of the environmental pressures on the interactors that affect replication.

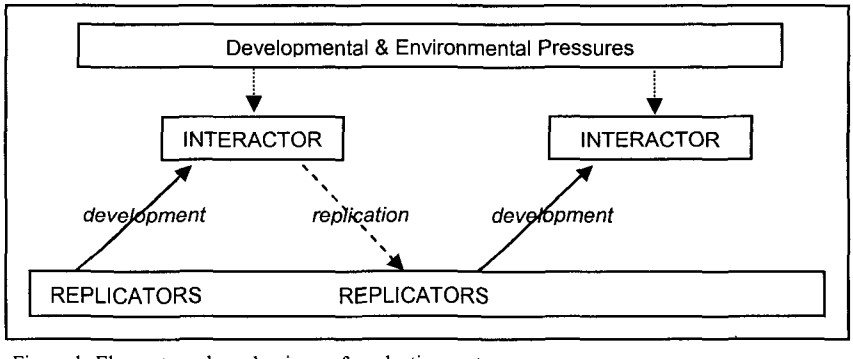


Figure 1. Elements and mechanisms of a selection system.

As shown schematically in Figure 1, during development, the information contained in the replicators unfolds to produce an interactor. Normal replication results in copies of the same replicators being produced into the replicator pool.

I propose that there are two instantiations of this selection system in language, one related to phonology and syntax (PS) and another one related to semantics. In the PS system, the interactor is a speaker's ability to process phonology and syntax (PS) in his or her native language, specifically the set of learned PS concept-to-form mappings and the replicators are tokens of PS use in speech. In the PS system, semantics plays the role of an environmental pressure providing concepts to be mapped onto forms by the interactor -the PS interactor is adapted to concepts.

In the semantic system, the interactors are linguistic utterances and the replicators are the concepts that exist in speakers' brains and that are replicated, or copied, in other speakers' brains by means of the interactors. Here, the PS system is an environmental factor determining how concepts are encoded into and decoded from utterances. It is important to emphasize that while PS replicators are found in speech, semantic replicators exist in speakers' brains (and while PS interactors reside in the brain, semantic interactors exist as speech).

The asymmetry between form and meaning in language has been pointed out by several authors (e.g. Tomasello, 2003; Davidson, 2003) and several facts support the evolutionary distinction between PS and semantics. One is the timescale of their evolution: PS patterns of change are slower and more systematic than semantic ones, for instance change in one sound induces change in the rest of the phonological space over decades, which has lead to systematic sound change patterns informing comparative method language phylogenetic classifications. PS patterns of change seem to be, then, language-internal. Semantic change, on the other hand, occurs much faster, with words changing meaning, new words being introduced in a language, and replacing old ones all the time, without systematic effects on the lexicon (Aitchison 2001), reflecting the interaction of humans with their environment.

According to the proposed framework, PS is learned through long-term, repeated exposure to a probabilistically structured input, whereas semantics (symbolic associations) is learned through other mechanisms, which may only involve a single exposure to a word. Evidence of the possibility of learning PS without semantics include Pierrehumbert (2003) and Monaghan, Chater and Christiansen (2005)'s studies showing that exposure to language-internal probabilistic cues such as acoustic and/or distributional patterns can lead to learning phonological and syntactic categories, respectively. Also, musical syntax learning relies on input-internal probabilistic patterns - and it seems to be processed in the same neural areas as auditory language comprehension (Maess et al., 2001). Cultural learning of birdsong syntax in oscines relies on song-

internal cues from tutors (Beecher & Brenwitz, 2005). Patients suffering from fluent aphasia can produce syntactically complex speech, but their processing of meaning is impaired.

In contrast, symbolic association can be learnt without language-internal probabilistic cues: apes are able to learn symbolic associations, but there is no evidence that they need to be sensitive to language-internal probabilistic cues or that they using PS-structured language forms (Terrace et al., 1979; Savage-Rumbaugh, 1993). Learning of naming in humans seems to depend on consistent cooccurrence of words with objects or actions in the environment as well as other language-external cues such as social ones (Hollich et al., 2000). And patients of Broca's aphasia have difficulties with sounds and syntax, but their comprehension (and therefore, their word form-meaning associations) remains relatively intact.

PS and semantics are, then, evolutionarily independent and show different evolutionary timescales and so can arguably be treated as a separate selection system. In the proposed framework, however, a semantic system is assumed to pre-date and to be a pre-requisite for human language emergence, and the two systems are intimately linked in a symbiotic relationship where each system provides necessary environmental requirements for the other.

2. Phonology and syntax

This section deals with an instantiation of the general selection system in the case of PS. Figure 2 illustrates this instantiation. Following Croft (2000) and Mufwene (2001), the level of the species is the language spoken in a community.

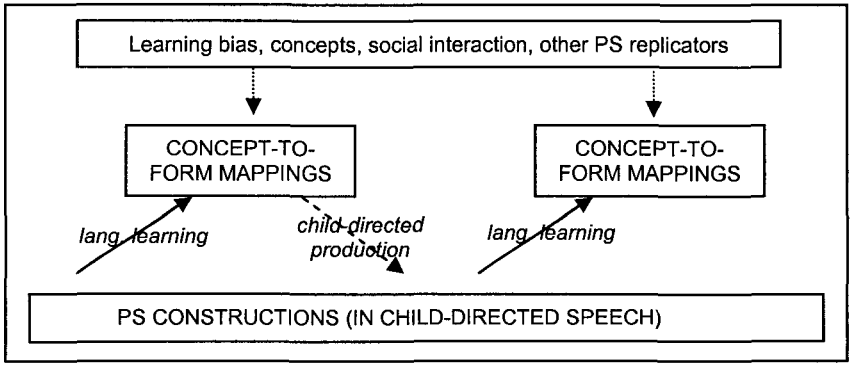


Figure 2. Dynamics of the Phonology and Syntax selection system.

The *interactors* are individual speakers' PS capacities, or the set of concept-to-form mappings that a speaker has learned. These interactors develop from the interaction between the PS replicators present in the speech that speakers have

been exposed to and pressures such as the learning bias, the structure of concepts and social factors. We can describe the interactor as the PS structure that develops around concepts to form a multi-level lexicon. PS contributes to that lexicon several layers of organisation, such as phonological, morphological and syntactic categories. It can also be described as symbolic association: the links or mappings between concepts and forms. The *replicators* are PS constructions found in speech, particularly in child-directed speech. Examples of replicators include sounds (phonetic realisations) and sound combinations that have a frequency or a conditional dependency, for instance frequent vs. infrequent phoneme combinations or long-distance sound combinations marking agreement.

As for the encoding of PS replicator information, while in biology genetic information is encoded digitally in the chemically (and temporally) stable sequence of bases in DNA molecules, in the case of PS, replicators are encoded statistically in the more imprecise and temporally unstable speech stream. Unlike spatial DNA, speech unfolds over time, making it impossible to go back to retrieve a piece of information obscured by noise. Statistical encoding solves this by providing information that becomes increasingly robust as the input sample grows larger. Moreover, statistical encoding is an adaptation to the developmental pressure on PS replicators to be learned by humans, and matches human probabilistic learning abilities. Mechanisms for *variation* in the replicator pool include language contact (Mufwene, 2001) and Lass's (1990) linguistic exaptation. Mechanisms for propagation of variation include social and prestige factors (Labov, 1972; Croft, 2000).

In PS *replication* the interactors copy their input replicators in their output speech, and this speech contributes to the development of a new PS interactor (in the brain of a new child). In this system, the interactor begins to "reproduce" before its development is complete - children begin to speak before they have a stable PS interactor. Notwithstanding the effects of horizontal transmission of unconventional speech from child to child, I assume that they are normally reversed by a larger amount of conventional speech from adults. Also, speakers continue to be exposed to speech over their whole life, however, I assume that, the PS system develops during the sensitive period for language learning in humans and reproduces during child-directed speech, when a suitable stimulus (an infant) elicits speech containing replicators that are optimally fitted to the learning biases. One prediction of this framework to be tested empirically is that, because the learning bias does not change over the cultural timescale, the PS of child-directed speech should show less variation between speakers both synchronically and diachronically than adult-directed speech, where other more labile pressures such as communication or prestige factors are at play.

A *developmental pressure* affecting PS interactors and acting on the structure of PS replicators in speech is the learning bias, that is assumed to include a sensitivity to probabilistic PS patterns in speech (for a mechanism

underlying such sensitivity see e.g. Maye, Werker & Gerken, 2002). This pressure is usually masked in a situation of normal language transmission because the structure of speech is already adapted to it, and for a given speaker, the PS replicators in her output speech are the same as those of her input speech. Only in situations of strong language contact, or during language emergence, when the input to a new generation is not already adapted to the learning bias, is the pressure's effect unmasked. (This can be studied by examining the outcome of replication when the input contains two different probabilistic replicators, for instance by adding mixed stimuli to Maye, Werker and Gerken's 2002 experiments, or by revisiting data from pidgins and creoles).

3. Semantics

An *environmental pressure* affecting PS replication and acting on PS interactors is the structure of the concepts. I argue that semantics is itself a selection system (see Figure 3).

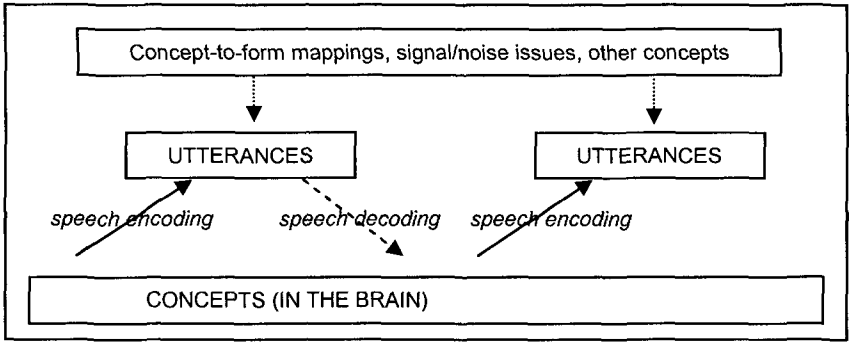


Figure 3. Dynamics of the Semantics selection system.

Moreover, I propose a symbiotic relationship between the PS and the semantic systems as each provides the environmental conditions necessary for the existence of the other.

In the semantic system, the *interactors* are speech utterances. Utterances develop from the interaction between pressures like the speaker's PS skill, the information capacity of the acoustic channel in the face of potential noise, and semantic replicators. The semantic *replicators* are concepts, specifically those transmissible through language, that exist in people's brains. They include the concepts behind words and constructions, and the relationships between them. *Variation* in the concept pool may arise for instance from contact between concepts in the brain.

Replication, or transmission of one concept from one brain to another, is mediated by the utterance. The encoding (development) of an utterance and its

subsequent decoding (replication of the concept) is carried out thanks to the PS interactor's mappings between concepts and forms. So the PS interactor is an environmental pressure affecting the semantic system. This illustrates the symbiotic relationship between the PS and the semantic systems, where each poses *pressures* on the other. Concepts can only be mapped onto utterances (semantic system) thanks to the PS interactor (the concepts-to-forms mappings, or symbolic association). Indeed, the human PS interactor would not exist in the first place if there were no concepts (semantic replicators) to be mapped onto forms. Additionally, there is a relationship of the PS-plus-semantics symbiotic system and its human hosts: language is an adaptation that increases human fitness, so natural selection favours the genes that provide language with the neural substrate it needs.

There are two meeting points between the PS and the semantic selection systems. In the *brain*, the concept-to-form mappings (the PS interactor) are adapted to the concepts that need to be communicated, and to how they are structured. This adaptation is embodied in symbolic association. If the PS system were not able to capture concepts, it would not increase human fitness and would not have been favoured by natural selection. In *speech*, utterances (as semantic interactors) need to be adapted to their substrate, namely the (probabilistically encoded) structure of the PS replicators, which is necessary for the easy acquisition of PS by humans. Again, if the PS replicators' encoding did not match human infants' learning biases, the PS system could not be replicated or transmitted over human generations.

4. Conclusion

I have presented a novel genetic framework to study the evolutionary dynamics of language. In this framework, phonology and syntax on the one hand and semantics on the other are best understood as two separate selection systems with different evolutionary dynamics and timescales, yet intimately intertwined in a symbiotic relationship where each system provides environmental factors that are crucial to the other system's existence. This symbiosis between PS and semantics is based on symbolic association and probabilistic encoding. Considering the two systems as separate in this way helps to explain the mutual influences between form and meaning in language and formalizes aspects of the relationships between linguistic representations in the brain and in speech. Finally, the proposed framework generates a prediction that can be tested empirically, namely the reduced PS-replicator variation in child-directed speech with respect to adult-directed speech.

References

Aitchison, J. 2001. *Language change; Progress or decay?* Cambridge: CUP.

Beecher, M. D., & Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends in Ecology and Evolution, 20(3)*, 143-149.

Croft, W. (2000). *Explaining language change.* Harlow: Longman.

Davidson, I. (2003). Archaeological evidence. In M.H. Christiansen and S. Kirby (eds.) Language Evolution, pp. 140-157. Oxford: OUP.

Gell-Mann, M. (1994). *The quark and the jaguar.* New York: Freeman & Co.

Hollich, G., Kathy Hirsh-Pasek, K. & Michnick Golinkoff, R. 2000. What does it take to learn a word? *Monographs of the Society for Research in Child Development. 65(3)*, 1-17.

Hull, D. L., Langman, R. E., & Glenn, S. S. (2001). A general account of selection: biology, immunology and behavior. *Behavioral Brain Sciences, 245*, 11-28.

Labov, W. (1972). *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

Lass. R. 1990, How to do things with junk: exaptation in language change. *Journal of Linguistics, 26*, 79-102.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience, 4(5)*, 540-545.

Maye, J., Werker, J. F. & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82(3)*, B101-B111.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential contribution of phonological and distributional cues in grammatical categorisation. *Cognition, 96*, 143-182.

Mufwene, S. S. (2001). *The ecology of language evolution.* Cambridge, CUP.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning and acquisition of phonology. *Language and Speech, 46(2-3)*, 115-154. .

Savage-Rumbaugh, E. S. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development, 233*, 58(3-4).

Terrace, H. S., Petito, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence? *Science, 206(4421)*, 891-902.

Tomasello, M. (2003). Different origins of symbols and grammar. In M.H. Christiansen and S. Kirby (eds.) Language Evolution, pp. 94-110. Oxford: OUP.

A STATISTICAL ANALYSIS OF LANGUAGE EVOLUTION

MARCO TURCHI

Department of Information Engineering , University of Siena, Via Roma 36, Siena, 53100, Italy
turchi@dii.unisi.it

NELLO CRISTIANINI

Department of Statistics, University of California Davis,One Shields Ave, Davis CA,95616, US
nello@support-vector.net

We propose to address a series of questions related to the evolution of languages by statistical analysis of written text. We develop a "statistical signature" of a language, analogous to the genetic signature proposed by Karlin in biology, and we show its stability within languages and its discriminative power between languages. Using this representation, we address the question of its trajectory during language evolution. We first reconstruct a phylogenetic tree of IE languages using this property, in this way showing that it also contains enough information to act as a "tracking" tag for a language during its evolution. One advantage of this kind of phylogenetic trees is that they do not depend on any semantic assessment or on any choice of words. We use the "statistical signature" to analyze a time-series of documents from four romance languages, following their transition from latin. The languages are italian, french, spanish and portuguese, and the time points correspond to all centuries from III bC to XX AD.

1. Introduction

In this paper we consider an aspect of language evolution, namely the process by which a language slowly changes by accumulation of many "neutral mutations", that is mutations that do not affect its effectiveness as a means of communication. The resulting "drift" can be studied as a trajectory in a space, as we will describe below.

Biological evolution is the process by which all forms of life change slowly over time because of slight variations in the genetic sequences that one generation passes down to the next. It has been known for some time, now, that the majority of molecular mutations are selectively neutral, that is do not affect the fitness of the phenotype and hence are free to accumulate. The corresponding statistical model of sequence evolution (The Neutral Theory of Evolution, by Motoo Kimura) is a centerpiece of modern genomics. In that model, evolution corresponds to a trajectory in the space of all possible DNA sequences, with most steps being neutral

with respect to selection, and mostly equivalent to a random walk. That neutral mutations can reach fixation for purely statistical reasons has been known for a long time.

Similar considerations can be made for the evolution of languages: neutral mutations accumulate, and some can become fixed in the population, over time. This creates a random walk, that can partly be reconstructed by simply keeping track of some statistical markers in the sequence, as done in DNA sequence evolution.

In this paper we investigate the use of statistical properties of languages to analyze linguistic evolution. We call them statistical language signatures (SLS) and we investigate how they evolve over time, how well their reflect ancestral relations between languages, and if they can be used to obtain language trees that are independent of any subjective choice. This approach by-passes any semantic assessment of word similarity or any arbitrary choice of words to be compared. It is repeatable automatically and hence objectively by simply performing statistical comparisons between text documents. Then we use SLS representation to analyze a time series of romance languages, from early latin to modern times. The approach is entirely data-driven. We make use of 3 datasets to independently validate our choice of features (SLS) and to analyze aspects of language evolution. A first dataset (containing 50 news stories written in 5 languages) is used to test the hypothesis that out representation is sufficiently stable and sensitive to characterize a language, at least within the domain of the indo-european (IE) family. The second corpus contains translations of the same document ("The universal declaration of human rights") in 34 modern languages. And the third dataset contains literary works from early latin to modern romance languages, covering the past 22 centuries.

The fundamental observation is that the SLS of a text does not depend on its semantic content, but rather on the language in which it is written. In other words, all documents in a language have similar statistical signature. Another key observation is that all languages we examine have their characteristic SLS, and that they can be reliably identified by it. We test both these observations on the first dataset, with high statistical confidence.

The consequence of these two - apparently conflicting - observations is that the SLS evolves slowly, drifting over time, and diverging as the languages diverge from a common ancestor. In this, it behaves similarly to the genomic signatures introduced by Karlin and on which our analysis is based (Karlin, Mrzek, & Campbell, 1997). To test this hypothesis, we used the second corpus, and standard phylogenetic reconstruction algorithms, to reconstruct a tree of the IE family. The resulting tree, entirely based on statistical properties, is generally in agreement with the commonly accepted view of the IE family, although some exceptions are discussed in the Conclusions.

Finally, we focus on the process of drift of a language in statistical space. We

model language evolution as a trajectory in the space of all possible statistical signatures, from an ancestral state to the current one. Modeling this drift is an important long term research goal, and we can only outline our approach in this paper. We use the third dataset to measure the distance covered by certain romance languages in the past 22 centuries. We notice some abrupt change points corresponding to known transitions from latin to national languages. At the end we outline a series of open problems, or research objectives, for this project.

In our current analysis we are limited by the use of texts available in the latin alphabet, and hence we focus mostly on european languages. However we believe that the methods can be exported to more general situations, perhaps using standard transliteration methods or - later - even phonetic representations.

1.1. *Statistical Language Signature*

It has been known for a long time that the probability of observing a certain character in a linguistic sequence depends strongly on the previous characters, and also is highly dependent on the language in consideration (Shannon, 1951). The frequency with which di-grams (pairs of letters) appear in a language is a very stable property of that language, as is a related quantity known as Karlin's odds ratio in genome analysis. If we remove all punctuation from a text document, all that is left is 26 letters and blank spaces separating them. So every document is a sequence from an alphabet of 27 letters. We denote by $C(i,j)$ the number of times that the di-gram (ij) is observed in the document. We can then define a di-gram frequency matrix as the matrix whose entry $D(i,j) = \frac{C(i,j)}{(n-1)}$ (where n is the document length). The odds-ratio matrix is defined as follows:

$$K(i,j) = \frac{C(i,j)}{C(i)C(j)}$$

where $C(i) = \sum_j C(i,j)$.

We want to investigate the use of D and K as statistical signatures of a language. We will also use them to assess the proximity between languages, and this means that we need to introduce a concept of distance that is appropriate in the space of matrices $\Re^{27\times 27}$. We are in this way defining a metric space where we "embed" a language, and we model language evolution as a trajectory in that space.

We will use two simple distances. Other choices are naturally possible, and should be investigated separately.

- Frobenius Distance:
  $D_F(M^1, M^2) = \|M^1 - M^2\|_F = \langle (m^1_{i,j} - m^2_{i,j}), (m^1_{i,j} - m^2_{i,j}) \rangle = \sqrt{\sum_{i=1}^{27} \sum_{j=1}^{27} |m^1_{ij} - m^2_{ij}|^2}$

- Kalin (1-norm) Distance:
  $D_{L1}(M^1, M^2) = \frac{1}{(27)^2} \sum_{ij} |m^1_{i,j} - m^2_{i,j}|$

With these definitions, we can model a language as a point in a space, and its evolution as a trajectory in that space. We could even measure its rate of movement, in principle, since we have a notion of distance. Certainly we can define language similarity, and use that as a proxy in phylogenetic reconstruction. All this can make sense, however, only if these features are stable: they should be properties of the language, and not of the given document; and they should be able to distinguish between languages. If that can be proven, we can analyze phylogenetic relations between languages in this representation.

1.2. *Suitability of SLS as Features*

Each language has its own statistical signature. In english, digrams such as "th" and "ed" are very frequent, in italian the typical endings in vowels can be seen as high frequencies of digrams "a-", "e-" etc (where we represented the blank symbol by "-"). These differences, that reflect grammatical, phonetic and historical factors, can be readily seen in the feature matrices of the two languages.

To test the stability of these features within a language, as well as their reliability as discriminators between languages, we have used our first corpus: a set of 50 documents (10 each for English, German, Spanish, Italian and French). We computed the average pairwise distance for documents in the same language and for documents in different languages. We than compared their ratio with the same quantity measured for randomly created sets of 10 documents. We repeated this 10,000 times, and each time the resulting ratio was larger: with p-value < 0.0001 this representation is well correlated to the difference between languages. Indeed, this quantity has been used to implement language classification systems for a long time (Beesley, 1988).

1.3. *Language Evolution in* $\Re^{27\times 27}$

If the SLS is a stable property of a language, and it is significantly different in related languages, it must be drifting over time. If this drift resembles a random walk (a hypothesis that should be tested in future work), then its *net* amount of drift should be proportional to the time dividing two languages, though a number of statistical corrections should be applied to the distance measured in feature space to really reconstruct the actual time since divergence. In this project we settle for a simpler test, using the pairwise distance matrix obtained with the expressions above to reconstruct a phylogenetic tree. We used the standard algorithm Neighbor Joining (Saitou & Nei, 1987), that is fairly tolerant to violations of the molecular clock assumption (genetic distance being proportional to time).

The dataset used for this part of the study is a subset of that used in (Benedetto, Caglioti, & Loreto, January 2002), our corpus being formed by 34 translations

of the "Universal Declaration of Human Rights" (UNResol, 1948) into modern languages from Romance, Celtic, German, Slavic, Baltic families, and the Basque language included as an outgroup. Also (Benedetto et al., January 2002) produced phylogenetic trees, using information theoretic tools.

The fact that each document is a translation of the "Universal Declaration of Human Rights" offers the advantage that they all have roughly the same length, which facilitates our statistical analysis. The disadvantage however, is that in very close languages, the translation of the same word can be the same, or have the same root. This means that our estimate distances for adjacent/far languages might be biased.

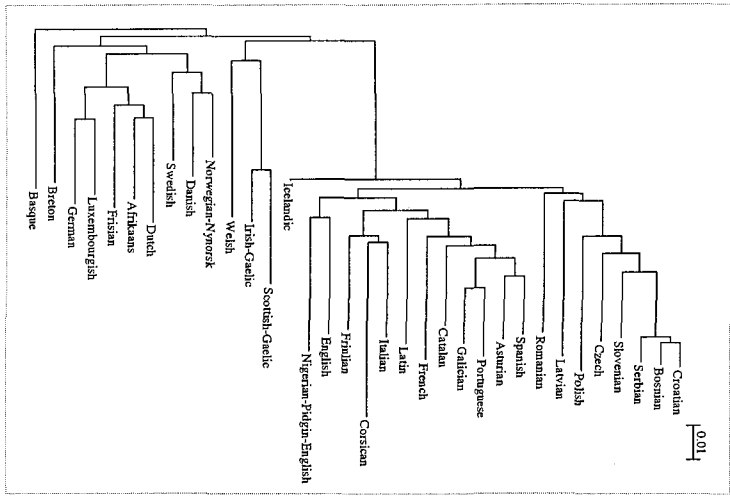


Figure 1. Language Evolution tree using the relative frequency of di-grams as features, and the Frobenius distance

The trees obtained with both SLSs (figures 1 and 2) are mostly compatible with the standard organization of the IE family, with the Karlin odds representation giving better results than the digrams. That means that not only can our SLS characterize a language, but can act as tags to track its evolution over long periods of time. Clearly this quantity seems to be changing slowly, and we can see from the fine organization of the slavic family or from the organization of languages in the iberian peninsula, it seems to also have a fairly steady drift. It is interesting to note that also the violations of the accepted topology of the tree can give us information about language evolution. For example, languages such as Romanian and English clearly are the result of massive borrowing from nearby languages, and an no longer be assigned to their original family (at least not their lexicon, which

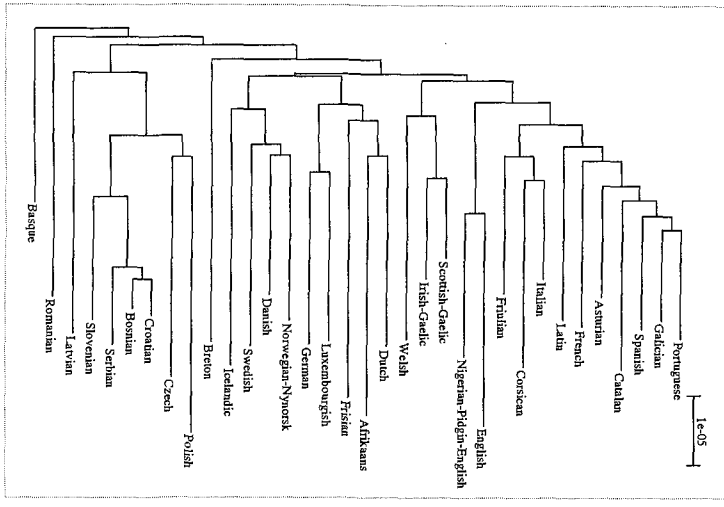


Figure 2. Language Evolution tree using the Odds ratios as features, and the Karlin distance

is what is captured mostly by this representation). In the di-grams representation there are various problems in assigning icelandic (which is instead correctly assigned by Karlin odds) and english in all cases seems to be attracted by french. This is better seen in the Multidimensional Scaling plot of the 34 languages.

Notice that we simplified the text to force it into a 26 letters alphabet, in so doing removing significant information, such as that coming from special letters in various languages. In particular we mapped the letters to their nearest english-alphabet counterpart, without using a linguistic criterion. Our assumption was that given the inherently statistical nature of the approach, we could ignore at a first approximation the effects of this arbitrary step, modeling them as a small perturbation of the signal. This has been the case for most languages, but in some cases, however, this rough simplification has proven to be sufficient to mislead the algorithms (see for example Breton). In the future, we are planning to make use of the phonetic alphabet, to reduce this effect.

1.4. *Time Series Analysis.*

The third experiment focused on time series analysis of documents spanning 22 centuries within the romance family. We constructed a dataset containing 119 different documents, written in Latin, Italian, Spanish, Portuguese and French, start from 200 BC and including the 20th century. Documents are mostly literary works, chosen to cover uniformly every period and every language. The non latin languages start mostly in the XI century, and have about 12 documents per century.

We measure the distance of each document from the oldest one, and we plot

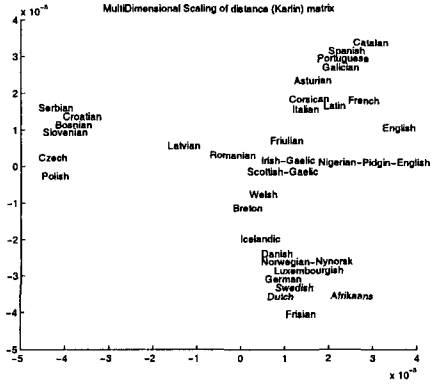


Figure 3. Multi Dimensional Scaling of some IE Languages

this distance as a function of time. The obvious change point observed in the XI century could be an artefact due to the fact that we could not find earlier documents in the non-latin languages, and is clearly a draw-back of using written language as opposite to spoken one. A more careful choice of the data might help us to reduce the gap in that transition. Also we can find written latin documents throughout the entire period, but we have stopped the latin series more or less where the national languages series started. What is more interesting is that the distance from the origin is in all languages more or less comparable: they all seem to have moved of a comparable amount, in the 22 centuries, although not smoothly (see figure 5). We can see the distances between these languages also in figure 4, although this multidimensional scaling representation can be misleading (projects into 2 dimensions a 27^2 dimensional dataset).

1.5. *Conclusions*

Various conclusions can be drawn from the experimental results we obtained: the first one is that some aspects of historical linguistics can indeed be investigated by using statistical tools. This rises hopes of applying the same tools to ancient texts, so as to look further back in time. But at the same time, a number of problems with this approach are visible in the results, directly suggesting various improvements.

First, it is not always the case that this statistical approach is robust enough to ignore the effect of alternative spelling conventions (as seen in the case of Breton and Icelandic). This can be addressed by moving future investigations to documents written using the IPA (international phonetic alphabet). Notice however that it can be argued that even spelling conventions evolve, and are part of the phylogenetic signal we are trying to analyze, as we focus on the evolution of written text. Second, we see the effect of borrowings (as seen in the case of English

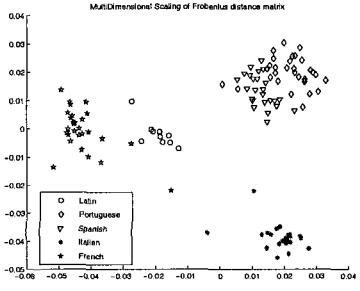


Figure 4. Multi Dimensional Scaling of some Romance Languages

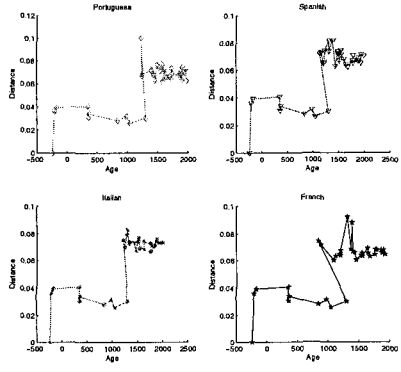


Figure 5. Time Series Analysis of some Romance Languages

and Romanian): in many cases the assumption that the evolutionary history of languages can be represented by a tree is not justified, at least with respect to their lexicon. This can be addressed by using tools from evolutionary biology aimed at reconstructing "phylogenetic networks" rather than trees.

Because of the inherently statistical nature of this approach, however, to a first approximation we believe that all the above effects can be treated as random perturbations, and for most languages they are not sufficient to corrupt the phylogenetic signal. As we refine the method, we expect to find cleaner and more informative patterns in the data.

References

Beesley, K. R. (1988). Language identifier: A computer program for automatic natural-language identification of on-line text. *The 29th Annual Conference of the American Translators Association*, 4754.

Benedetto, D., Caglioti, E., & Loreto, V. (January 2002). Language trees and zipping. *Physical Review Letter*, *88*(4).

Karlin, S., Mrzek, J., & Campbell, A. M. (1997). Compositional biases of bacterial genomes and evolutionary implications. *Journal of Bacteriology*, *179*(12), 38993913. (0021-9193/97/04.0010)

Saitou, N., & Nei, M. (1987). The neighbour-joining method: a new method for constructing phylogenetic trees. *Mol. Biol. Evol.*

Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell Systems Technical Journal*(30), 50-64.

Universal declaration of human rights. (1948, December). (United Nations General Assembly Resolution)

EVOLUTIONARY GAMES AND SEMANTIC UNIVERSALS

ROBERT VAN ROOIJ

ILLC, University of Amsterdam, Nieuwe Doelenstraat 15, Amsterdam, 1012 CP, the Netherlands
R.a.m.vanRooij@uva.nl

An evolutionary perspective on signaling games is adopted to explain some semantic universals concerning truth-conditional connectives; property denoting expressions, and generalized quantifiers. The question to be addressed is: of the many meanings of a particular type that *can* be expressed, why are only some of them expressed in natural languages by 'simple' expressions?

Most work on the evolution of language concentrates on the evolution of syntactic and phonetic rules and/or principles. This is reasonable, because in the generative tradition these disciplines acquired a central place in linguistics. In another sense, however, the under-representation in evolutionary linguistics of work that concentrates on semantics is surprising: how many of us would be interesting in language if it was not the main vehicle used to transmit meanings? Moreover, semantics and pragmatics are by now well-established disciplines within linguistics that study how, across languages, meanings are transmitted by language. In this paper I will concentrate on giving evolutionary motivations for some semantic features shared by all or most languages of the world.

There are in fact many semantic features shared by all languages of the world. For instance, it seems that of all the speech acts that we can express in natural language, only three of them are normally grammaticalized, and distinguished, in mood (i.e., declarative, imperative, and interrogative). In this paper, we will be most interested in similar kinds of universals that make claims about what kinds of meanings are expressed by short and simple terms (e.g. with one word) in natural languages. One of them concerns *indexicals*, short expressions corresponding to the English *I*, *you*, *this*, *that*, *here*, etc., the denotations of which are essentially context-dependent. It seems that all languages have short words that express such meanings (cf. Goddard, 2001), and this fact makes evolutionary sense: it is a useful feature of a language if it can refer to nearby individuals, objects, and places, and we can do so by using short expressions because their denotations can normally be inferred from the shared context between speaker and hearer. In this paper I will be concerned with similar universals involving mainly the connectives, property denoting expressions, and generalized quantifiers.

Signaling games and Connectives In *signaling games* as introduced by David Lewis (1969), signals have an underspecified meaning, and the actual interpretation the signals receive depends on the equilibria of sender and receiver strategy combinations of such games. Recently, these games have been looked upon from an *evolutionary* point of view to study the evolution of language. According to it, a signaling convention can arise in which signal s denotes t if and only if in the *evolutionary stable strategy* (ESS) signal s is only used when the speaker is in situation t. Thinking of meanings as situations, one can show that if there exists a 1-1 mapping between situations and the best actions to be performed there, and there are enough messages, the ESSs, or resulting communication systems, of signaling games always give rise to 1-1 mappings between signals and meanings. It is obvious that in this simple communication system there can be no role for connectives: the existence of a disjunctive or conjunctive message would destroy the 1-1 correspondence between (types of) situations and signals. That gives rise to the question, however, under which circumstances messages with such more complex meanings could arise. In this paper I concentrate only on one particular truth-conditional connective: *disjunction*.

Taking t_i and t_j to be (types of) situations, under which circumstances can a language evolve in which we have a message that means 't_i', one that means 't_j', and yet another with the disjunctive meaning 't_i or t_j'? As indicated above, if there exists a 1-1 function from situations to (optimal) actions to be performed in those situations, a language can evolve with a 1-1 correspondence between signals and situations. The existence of this 1-1 function won't be enough, however, to 'explain' the emergence of messages with a disjunctive meaning. What is required, instead, is a 1-1 function from *sets* of situations to (optimal) actions. We can understand such a function in terms of a payoff table like the following:

	a_1	a_2	a_3	a_4	a_5	a_6	a_7
t_1	4	0	0	3	3	0	2.3
t_2	0	4	0	3	o	3	2.3
t_3	0	0	4	0	3	3	2.3

Notice that according to this payoff table, for each $i \in \{1, 2, 3\}$ action a_i is the unique optimal action to be performed in situation t_i. This table, however, contains more information. Suppose that the speaker (and/or hearer) knows that the actual situation is either t_1 or t_2, and that both situations are equally likely. In that case the best action to perform is neither a_1 nor a_2 – they only have an expected utility of 2 –, but rather a_4, because this action now has the highest expected utility, i.e., 3. Something similar holds for information 't_1 or t_3' and action a_5, and for 't_2 or t_3' and action a_6. Finally, in case of no information, which corresponds with information 't_1 or t_2 or t_3', the unique optimal action to perform is a_7. Thus for all (non-empty) subsets of $\{t_1, t_2, t_3\}$ there exists now a unique best action to be performed. Notice that each such subset may be thought of as an *information*

state, the (complete or incomplete) information an agent might have about the actual situation. Suppose now that we lift the sender-strategy from a function that assigns to each *situation* a unique message to be sent, to one that assigns to each *information state* a unique message to be sent. Now it can be shown that we will end up (after evolution) with a communication system (an ESS) in which there exists a 1-1-1 correspondence between information states (or sets of situations), messages, and actions to be performed.[a] Thus, there will now be messages which have a disjunctive meaning. This by itself doesn't mean yet that we have a separate message that denotes disjunction, but only that we have separate messages with disjunctive meanings in addition to messages with simple meanings. However, as convincingly shown by Kirby and others, a learning bottleneck is a strong force for languages to become compositional. It is reasonable to assume that under such a pressure a complex message will evolve which means $\{t_i, t_j\}$ that consists of three separate signals: one signal denoting $\{t_i\}$, one signal denoting $\{t_j\}$, and one signal that turns these two meanings into the new meaning $\{t_i, t_j\}$ by (set theoretic) *union*. The latter signal might then be called 'disjunction'.

In principle, once we take information states into account, we cannot only state under which circumstances disjunctive messages will evolve, but also when negative and conjunctive messages will evolve.[b] The main difference is that we have to assume more structure of the set of information states. An interesting feature of our evolutionary description of the connectives is that it might answer the question why only humans have communication systems involving (truth-conditional) connectives. In contrast to the signaling games discussed by Lewis, and used to explain the alarm calls of, e,g. vervet monkeys, it was crucial for connectives to evolve to take *information states*, or *belief states* into account, i.e., sender strategies must take *sets* of situations as arguments, and not just situations themselves, and this must be recognized by receivers as well. Perhaps, the existence of such more complicated sender strategies is what that sets us apart from those monkeys.

Why not more connectives? Once we assume that each (declarative) sentence is either true or false, there are *four* potential unary connectives, and as much as *sixteen* potential binary connectives. Although all these potential connectives *can* be expressed in natural language, the question is why only one unary (*negation* and only two (or perhaps three) binary truth-functional connectives (*disjunction* and *conjunction*) are expressed by means of simple words in all (or most) natural languages? That is, can we give natural reasons for why languages don't have the truth-functional connectives that are mathematically possible? For unary connectives this problem is easy to solve. Look at the four possible unary truth-

[a]This is a general result, and not restricted to the particular example discussed above.

[b]More interesting things can be said about why, and of the conditions under which, messages with negative and conjunctive meanings could evolve, but space doesn't allow me to go into this here.

conditional connectives, $c_1, ..., c_4$:

p	$c_1\ p$	$c_2\ p$	$c_3\ p$	$c_4\ p$
1	0	1	0	1
0	1	0	0	1

Connective c_1 is, of course, standard negation. Why we don't see the others in natural language(s) is obvious: they just don't make sense! $c_2\ p$ just has the same truth-value as p itself, and, thus, c_2 is *superfluous*, while the truth values of $c_3\ p$ and $c_4\ p$ are *independent* of the truth value of its argument p, which leaves it unclear why c_3 and c_4 require arguments at all.

For binary connectives the problem is more difficult, but Gazdar & Pullum (1976) show that when we require that all lexicalized binary connectives must be *commutative* and obey the principles of *strict compositionality* and *confessionality*, all potential binary connectives are ruled out except for the following three: conjunction, standard (inclusive) disjunction, and what is known as *exclusive* disjunction. This is an appealing result, because (i) strict compositionality makes perfect sense, (ii) the principle of confessionality – which forbids (binary) connectives which yields the value true when all its arguments are false – can be explained by the psychologically well-established fact that negation is difficult to process, while (iii) the constraint of commutativity is motivated by the not unnatrual idea that the underlying structures of the connected sentences are linearly unordered. The non-existence of a lexicalized exclusive disjunction can be explained, finally, by the standard conversational implicature from *A or B* to *not A and B*, which makes such a connective superfluous.

Properties In extensional terms, any subset of a set of individuals, or objects, can be thought of as a property. Thinking of properties in this way, however, leaves us with many more properties that *can* be expressed, than that there are simple expressions that denote properties in any natural language. This gives rise to the following questions: (i) can we characterize the properties that are denoted by simple expressions in natural language(s), and, if so, (ii) can we give a pragmatic and/or evolutionary explanation of this characterization?

The first idea that comes to mind to limit the use of all possible properties, is that only those properties will be expressed a lot in natural language that are *useful* for sender and receiver. Using our signaling game framework, it is easy enough to show how usefulness can influence the existence of property denoting terms when we either have less messages, or less actions than we have situations.[c] To

[c]These abstract formulations might be used to model other 'real-world' phenomena as well, such as noise in the communication channel which doesn't allow receivers to discriminate enough signals; a limitation of the objects speakers are acquainted with, perhaps due to ever changing contexts; and maybe also non-aligned preferences between sender and receiver.

illustrate the first case, consider a game involving three situations, three actions, but only two messages. Taking the sender and receiver strategies to be functions from situations to messages and messages to situations, respectively, we predict that in equilibrium only two actions will be performed. Which of those actions that will be depends on the utilities and probabilities involved. Consider the following utility tables:

	a_1	a_2	a_3
t_1	8	0	0
t_2	0	4	1
t_3	0	0	2

	a_1	a_2	a_3
t_1	1	0	0
t_2	0	1	0
t_3	0	0	1

In both cases there exists a 1-1 correspondence between situations and messages. If there are three messages, in each situation the sender will send a different message, and the receiver will react appropriately. When there are only two messages, however, expected utility will play a role. In the left-hand table above it is more useful to distinguish t_1 from t_2 and t_3, then to distinguish t_2 from t_3. As a consequence, in equilibrium t_2 and t_3 will not be distinguished from each other and in both situations the same message will be sent. We have implicitly assumed here that the probability of the three situations was equal. Consider now the table on the right-hand side, and suppose that t_1 is much more likely to occur than t_2, which, in turn, is much more likely than t_3. Again, it will be more useful to distinguish t_1 from t_2 and t_3, then to distinguish t_2 from t_3. Thus, also here we find that in equilibrium t_3 will not be distinguished separately, and meshed together with t_2.

A common complaint of Chomskyan linguists (e.g. Bickerton, Jackendoff) against explanations like the one above is that usefulness can't be the only constraint: there are many useful properties, or distinctions 'out there' that are still not really named, or distinguished, in simple natural language terms. Bickerton (1990) mentions *contiguity* (or *convexity*) as an extra constraint, and hypothesizes that the preference for convex properties is an innate property of our brains. Unfortunately, if we think of properties as in standard semantics just as subsets of the universe of discourse, such a constraint cannot even be formulated. For reasons like this, Gärdenfors – following philosophers like van Fraassen and Stalnaker – proposed to use a meaning space to represent meanings in which the notion of convexity makes sense. This meaning space is essentially an n-ary vector space where any subset of this space is (or represents) a property. However, because each point in space can now be characterized in terms of the values of its coordinates, Gärdenfors can make a distinction between 'natural' and 'unnatural' properties: only those subsets can be thought of as natural properties that form *convex regions* of the space.[d] Because only a small minority of all subsets of any struc-

[d]For a set of objects to be a convex region, it has to be closed in the following sense: if x and y are

tured meaning space form convex regions, the hypothesis that (most or all) simple natural language property denoting expressions denote such convex regions is, potentially, a very strong one. Gärdenfors' proposal is quite successful for some categories of property denoting expressions, like colors, and this gives rise to the question what makes convex regions so natural. This question is addressed in Jäger & van Rooij (to appear). It is shown that only those communication systems will be evolutionary stable in a signaling game where the sender strategy is just a function from points in the meaning space to messages, and where the receiver has to guess this point, in case the set of points in the space in which the same signal is sent forms a convex region in this shared meaning space with a prototype.

Gärdenfors (2000) mentions a number of examples (of property-, but also of relation denoting expressions and prepositions) where convexity seems like a natural constraint, and might give rise to semantic universals. We won't go into these examples here, but instead (i) will discuss some examples not discussed by Gärdenfors where convexity can explain some well established semantic universals, and (ii) will speculate a bit on the difference between communication systems of (some) animals, young children and adults humans making use of the above mentioned evolutionary motivation for convexity. I start with the latter.

Basic level properties It is a basic observation that many property denoting expressions used by adults, (e.g. *tool, furniture*) denote objects that are not similar to each other, neither with respect to appearance, nor w.r.t. (basic) function. The psychologist Rosch (1978) made a distinction between *basic level* categories/properties (*chair, dog*), and sub- and superordinate ones (*armchair, furniture*), and proposed that only for the first ones the notion of similarity plays an important role. She also observed that it are the first ones that are learned earlier and easier by children, and – we might speculate –animals never come any further than making basic level category-like distinctions. Now, notice that in terms of meaning spaces, convex sets are defined in terms of a distance measure, where the 'closeness' of two objects to each other depends on their (mutual) resemblance. This gives rise to the hypothesis that in contrast to animals and young children, only 'adult' humans can make use of expressions in their communication systems that denote non-convex properties. Interestingly enough – and in parallel with our above 'explanation' of why only humans make use of connectives –, this contrast might be understood from the complexity of the sender strategies used in signaling games that generate (non-convex) properties. Remember that to explain the emergence of property denoting expressions we assumed that sender strategies were just very simple functions from situations to messages. When we assume that objects exist in structured meaning spaces, all properties that will be expressed in equilibrium form convex regions with obvious prototypes. But this means that

elements of the set, all objects 'between' x and y must also be members of this set.

to explain the existence of those properties that do not denote convex sets (i.e., by hypothesis the sub- and superordinate ones) and/or do not have prototypes, we need either more involved sender strategies (cf. the case of connectives), or utility functions not defined in terms of a very simple measure of similarity. Again, this might explain why only adult humans can make use of non-basic level property denoting expressions. What our analysis also explains is why *conjunction* seems easier to understand and process than *disjunction* and *negation*. Notice that these connectives make sense for properties as well. Now one can show that in contrast to the other connectives the conjunction of two convex properties is guaranteed to be convex as well (this is not true for the connectives of 'quantum logic', though).

Quantifiers and determiners Most work on universals in model-theoretic semantics is concentrated on quantifiers and determiners. This is also very natural, given that the discrepancy between the number of meanings that are predicted to be expressible, and the terms to do so is here much larger than for properties and relations. To get a glimpse of this, in a simple extensional model with only 4 individuals, standard model theoretic semantics predict that there are not less than $2^{2^4} = 65.636$ quantifiers that can be expressed, and even the immense number of 2^{4^4} many determiners! Obviously, constraints are in order to limit the meanings that can be expressed by (simple) noun phrases and determiners.

Because a determiner denotes a relation between properties, or, equivalently, a function from properties to quantifiers, any constraint on quantifiers gives rise to a constraint on determiners as well. So we can safely limit ourselves to constraints on determiners. A simple, and very intuitive constraint is *variety*. A determiner shows variety iff it gives rise to a *contingent* meaning: the sentence of type 'Det Noun VP' in which it occurs is neither always true nor always false. More formally, determiner D is said to show variety iff in every model in which the determiner is defined there are A, B such that $D(A, B)$ is true, and A', B' such that $D(A', B')$ is false. It is clear that we can form complex determiners which do not show variety (like *some or no*), but it is generally assumed that all 'simple' determiners satisfy this constraint. An explanation of this fact is easy to imagine: why would a language end up with a simple determiner the use of which doesn't express an informative, and thus useful, proposition?

In this paper we will only explain one semantic universal, stated in essence already in Barwise & Cooper (1981), which says that all 'simple' determiners satisfy the following *continuity* constraint:

For all A, B, B', B'': if $D(A, B')$, $D(A, B'')$ and $B' \subseteq B \subseteq B''$, then $D(A, B)$.

I claim that the notion of convexity can be used to motivate this universal, at least if we assume that the meaning of natural language determiners are *context-independent* and *conservative*. Assume that E and E' are domains of discourse,

and π a permutation function on E'. The *context-independence* constraint then says that if $A, B \subseteq E \subseteq E'$, then $D(\pi(A), \pi(B))$ is true with respect to E iff $D(A, B)$ is true with respect to E'. Intuitively, this means that the meaning of a sentence of the form $D(A, B)$, where, as before, D is the determiner meaning, A is the noun-denotation, while B is the denotation of the VP, doesn't depend on the domain of discourse, and only on the number of individuals in A, B, and $A \cap B$. The further constraint of *conservativity* then says that the meaning of such a sentence depends only on the number of individuals in $A \cap B$ and $A - B$. Intuitively, a determiner is said to satisfy conservativity iff the truth or falsity of a simple sentence of the form NP VP depends only on the denotation of the noun of the NP. An important observation due to van Benthem (1986) is that all quantifiers that satisfy *context-independence* and *conservativity* can be represented geometrically in the so-called 'tree of numbers'. This tree can be thought of as a binary meaning space with as coordinates the numbers of individuals in $A \cap B$ and $A - B$. Each quantifier satisfying the above two constraints can now be represented as a subset of this meaning space, and only some of these subsets form convex regions. One can now show that the continuous quantifiers all give rise to such convex sets. Thus, if the tree of numbers is a natural representation format of generalized quantifiers, our signaling game analysis can help to motivate one very important semantic universal.

The tree of numbers itself can be argued to be a natural geometrical representation format of (most) generalized quantifiers, by motivating the constraints of *context-independence* and *conservativity*. *Conservativity* will be explained, for instance, by the evolutionary preference of languages to follow a topic-comment structure (as for instance already motivated by linguists with as diverse backgrounds as Givon and Bickerton).

References

Barwise J. & R. Cooper (1981), 'Generalized quantifiers in natural language', *Linguistics and Philosophy*, **4**: 159-219.

Benthem, J. van (1986), *Essays in Logical Semantics*, Kluwer, Boston.

Bickerton, D. (1990), *Language and Species*, Univ. of Chicago Press, Chicago.

Gärdenfors, P. (2000), *Conceptual Spaces*, MIT Press, Cambridge, MA.

Gazdar, G. & G.K. Pullum (1976), 'Truth-functional connectives in natural language', *Chicago Linguistic Society*, pp. 220-234.

Goddard, C. (2001), 'Lexico-semantic universals', *Linguistic Typology*, **5**: 1-65.

Jäger, G. and R. van Rooij (to appear), 'Language structure', *Synthese*.

Lewis, D. (1969), *Convention*, Harvard University Press, Cambridge, MA.

Rosch, E. (1978), 'Principles of categorization', In E. Rosch & B. Lloyd (eds.), *Cognition and categorization*, Hilsdale, NJ: Erlbaum.

OVEREXTENSIONS AND THE EMERGENCE OF COMPOSITIONALITY

PAUL VOGT

Language Evolution and Computation unit, University of Edinburgh
40 George Square, Edinburgh EH8 9LL, U.K.
Computational Linguistics and AI section, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
paulv@ling.ed.ac.uk

This paper investigates the effect overextensions of words may have on the emergence of compositional structures in language. The study is done using a recently developed computer model that integrates the iterated learning model with the language game model. Experiments show that overextensions due to an incremental acquisition of meanings on the one hand attracts languages into compositional structures, but on the other hand introduces ambiguities that may act as an antagonising pressure.

1. Introduction

Over the past decade, a lot of computational models have investigated the emergence of compositional structures in language (for an overview consult, e.g., Briscoe, 2002). Many of these studies have used simulations of multi-agent systems, where the communication system emerges through cultural interactions, individual learning and self-organisation (possibly in combination with the evolution of a LAD). Most of these models have assumed that individual agents (i.e. the individuals of the language community) are 'born' with a predefined semantics (e.g., Kirby, Smith, & Brighton, 2004). Naturally, this assumption is not realistic in our human society. This paper focuses on the overextension of meaning, which can occur when the semantics are not predefined, but are developed during an agent's lifetime.

It is well known that during the process of language acquisition, children go through a phase in which they overextend the meaning of words by using them for inappropriate referents (see, e.g., Clark, 2003). It is unclear what causes this behaviour, but it might be that children cannot yet distinguish among different referents, or that they do not have the proper word for a referent yet. Typically, overextensions occur very early in life and an overextended form can last from one day to several months. Although overextensions are typically considered as a phase relating to the acquisition of word meanings, it is interesting to investigate

if they can have an unexpected (side-)effect.

As a conclusion to a recently studied model on how compositionality can emerge in a simulation in which the semantics develop ontogenetically in individuals, it has been hypothesised that overextensions both can provide a positive attraction towards using compositional structures, and an antagonising pressure against the emergence of compositional structures (Vogt, 2005c, 2005a). This simulation is based on a model that integrates the *iterated learning model* (Kirby et al., 2004) and the *language game* (or guessing game) model of the Talking Heads experiment (Steels, Kaplan, McIntyre, & Van Looveren, 2002).

This paper further explores the hypothesised effect of overextensions on the emergence of compositionality in language. The next section briefly introduces the model. Section 3 then presents some experimental results that test the hypothesis. Finally, Section 4 concludes.

2. Grounded iterated learning

Earlier work on the ILM has shown how initially holistic languages can evolve into compositional ones when the language is iteratively transmitted from one generation of individuals to the next, provided the individuals have the appropriate learning mechanisms to discover compositional structures, and the language is transmitted through a *bottleneck*, such that children only learn from a subset of the language (Kirby et al., 2004). One limitation of this earlier work is that all agents start their lives with a predefined semantics.

Similar results were achieved in Vogt (2005c), where the individuals acquire their meanings incrementally as they engage in language games to develop their language. In this model compositionality emerges in the first generation. However, compositionality only remains stable over time when the language is transmitted through a bottleneck, like in the earlier ILMs. When the language is not transmitted through a bottleneck, compositional languages tend to collapse into holistic languages, provided the language is transmitted purely in a vertical direction (i.e. all speakers are adults and all hearers are children).

The model used in these recent studies is implemented in a simulation toolkit of the Talking Heads experiment (Steels et al., 2002), called THSim.[a] In this model, a population of agents tries to evolve a simple language by which they can describe geometrical coloured objects presented to them. The agents achieve this by engaging in a series of guessing games, which are played by two agents – a speaker and a hearer – selected from the population. In the model discussed here, all speakers are selected from the adult population and all hearers from the child population; thus the language is transmitted vertically like in most ILMs. The remainder of this section summarises the model very briefly; for a detailed explanation, the reader is referred to Vogt (2005c).

[a]THSim is available at http://www.ling.ed.ac.uk/~paulv/thsim.html.

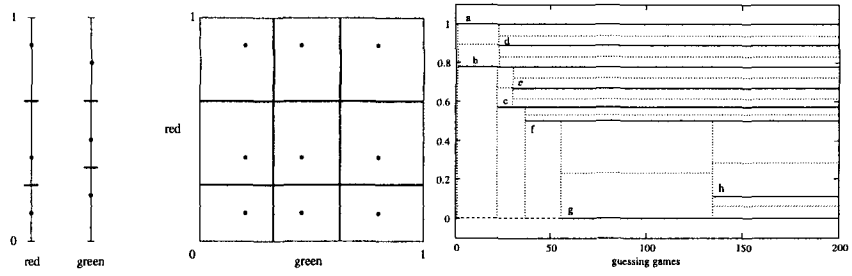


Figure 1. The left figure illustrates how categorical features (the dots on the two far left lines) can be combined to form categories in a 2-dimensional space. The right graph shows the development of categorical features in one quality dimension of an agent's meaning space during the agent's childhood. The x-axis shows the time in guessing games and the y-axis shows which values are occupied by a categorical feature. The solid lines show the CFs present at a certain time step and the dotted lines indicate the sensitive range of the CFs. Initially there are a few CFs, which are sensitive to a wide area in the feature space. Later on, as more CFs are constructed, the range of the CFs becomes more narrow.

Both agents in a guessing game are presented with a context containing a given number of objects. From these objects, both agents extract perceptual features concerning colour (represented by the red, green and blue components of the RGB colour space) and shape (based on the ratio between the object's area and the area of its smallest bounding box). The thus resulting four features are then categorised. First, for each object, each feature is categorised using a *categorical feature* (CF), which is a region in one quality dimension represented by a prototypical value. Then all the CFs of an object are combined to form a category (Fig. 1, left), which thus represents a region in a 4-dimensional *conceptual space* (Gärdenfors, 2000).

At the start of each agent's lifetime, the agent has no CFs in its repertoire. In order to communicate about an object, the agent is forced to distinguish the category of one (or more) object(s) from the other objects' categories in the context by playing a *discrimination game* (Steels et al., 2002). If categorising an object does not yield a distinctive category, the agent expands its repertoire of CFs by constructing new CFs for which it takes the object's features as exemplars. This way, the agents gradually constructs a repertoire of categorical features. Initially, these CFs are general and sensitive to a wide area of a quality dimension. Over time, when more CFs are constructed, these CFs become more specific and narrow down their sensitivity (see Fig. 1, right). As a result, when a category is used in a naming event, the reference of the used expression can be overextended. It is this property that is subject of the current investigation. (Note that, even though an agent may have only a few CFs in one dimension, in combination with the CFs acquired in the other dimensions, there are many more possible categories. Since the discrimination game only considers whether different objects in a context are

distinctive or not, these few CFs may be employed successfully for a period of time.)

Once the objects are categorised, the guessing game proceeds. The speaker, who selects one object as the topic, searches its grammar for rules with which it can encode an expression which conveys this topic. The grammar contains simple rewrite rules that are either holistic (e.g., `S->aword/ameaning`) or compositional (e.g., `S->A/m1 B/m2`). Holistic rules take meanings as categories formed by all 4 dimensions as though they are a single atomic concept. The meanings of compositional rules are formed through a combination of categories from conceptual spaces of lower dimension. If the rule is compositional, the agent will have other rules that rewrite the non-terminals (A and B) to words (e.g., `A->word/meaning`). Each rule is given a score, which indicates the effectiveness of the rule based on previous guessing games, and which is adapted according to the outcome of a game. When the speaker finds more than one way to encode an expression, it will select the composition that has the highest combined score. If the speaker fails to encode an expression, it invents a new form either holistically or – in the case that it can encode a part of an expression – in relation to one existing non-terminal.

The encoded expression is then uttered to the hearer, who in turn searches its grammar for ways to decode the expression. Each possible parse results in a possible meaning for the expression. All resulting possible meanings are then filtered such that only those meanings that are consistent with the current context remain. If more than one meaning is left, the hearer selects the one with the highest combined score and the object that belongs to that meaning is then guessed as the speaker's topic. This information is then conveyed back to the speaker (similar to pointing), who verifies whether or not the hearer guessed the right topic. If this is the case, the speaker acknowledges success, otherwise, it will inform the hearer which object was the topic (again similar to pointing).

If the game was successful, the agents increase the scores of the rules that were used, while the scores of competing rules are inhibited (a rule is competing if it could have been used in the same situation). If the hearer guessed the wrong topic, the scores of the rules it used are decreased. In addition, the hearer then adopts the expression with the meaning of the correct topic. If the hearer could not decode the expression, it also adopts the expression with the meaning of the topic.

Adopting an expression is done in one of three ways. First, if the hearer could parse a part of the expression with the intended topic (i.e. a part of the expression maps onto one constituent of an already existing rule of which the meaning matches a part of the topic's meaning), the remaining part of the expression is associated with the remaining part of the meaning. Second, if the first method fails, the hearer will try to chunk (or break up) the expression in two. To achieve this, the hearer searches an instance-base that contains all previously used expression-

meaning pairs, and finds those instances that fit a part of the expression-meaning pair to be learnt. If there are such instances, the heard expression is chunked such that it best fits the data acquired so far. Third, if the expression cannot be chunked, the hearer incorporates the expression holistically and adds its association with the topic's meaning to its grammar unanalysed.

It is important to stress that at the start of each agent's lifetime, its grammar is empty. All linguistic knowledge is thus acquired by playing these guessing games. In the simulations, the guessing game model is integrated with the ILM, such that the population of each iteration consists of a number of adults and a number of children. During an iteration, the population plays a given number of guessing games, after which all adults are removed, the children become adults and new children are introduced. This process repeats for a given number of iterations.

3. Overextensions and compositionality

In order to test the hypothesis presented in the introduction, two things need to be shown: (1) overextensions increase the tendency for compositionality to emerge, and (2) overextensions provide an antagonising pressure against compositionality. As mentioned in the previous section, overextensions in this model tend to emerge due to the gradual development of categorical features, as a result of which categories are initially sensitive to a wider range of objects (cf. Fig. 1, right). So, for example, if the agent in Figure 1 learns the word for triangle – for which the proper CF has value 0 – very early in its life, say around guessing game 20, then this word would wrongly be associated with the CF labelled b, which corresponds to a hexagon. Likewise, when this agent needs to produce a word associated with the CF labelled b during the same period, this word could be overextended to nearly all shapes. (Note that in the current study, children only start producing utterances once they are adults.)

To avoid the emergence of overextensions, it is possible to equip each agent with a predefined set of CFs that have a one-to-one correspondence to the features of all objects. In a previous study where this was done, it was shown that compositional structures tend to emerge more rapidly when the agents go through a period of overextensions than when they do not (Vogt, 2005a).[b] As the only difference in these two conditions was the presence or absence of overextensions, this result supports part (1) of the hypothesis.

In the same study, it was shown that when the language is not transmitted through a bottleneck (i.e. all children observe the entire language during their learning period), the compositional structures that arise when there are no overextensions remain stable over subsequent generations. In the case where there were overextensions, the compositional structures tended to collapse in favour of holis-

[b]Note that the focus in Vogt (2005a) was not on overextensions, but on statistical properties of the input to language learners.

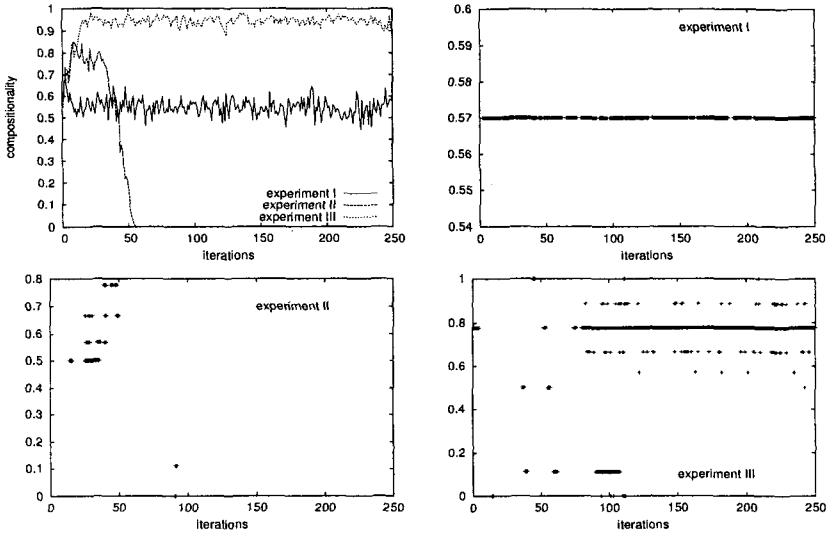


Figure 2. Top left: Compositionality of a typical run of the three experiments measured at the end of each iteration. The other graphs show the typical dynamics of preferred meanings in the shape dimension of a word that was used successfully in experiments I (top right), II (bottom left) and III (bottom right).

tic languages when the language was not transmitted through a bottleneck. This, thus, proves part (2) of the hypothesis.

Now let us take a closer look at some of the dynamics of these findings. In the simulations presented here, the population size was set to 6 (3 adults and 3 learners), the world contained 120 objects (10 colours × 12 shapes), the population played a total of 3,000 guessing games per iteration, and the simulations were run for 250 iterations.[c] At the end of each iteration, the population was tested in 200 situations, where each agent produced expressions about the same objects and each other agent tried to interpret each produced expression without learning.

From these test phases, the proportion of expressions that were made using compositional rules was measured. Figure 2 (top left) shows the evolution of this *compositionality* measure for typical runs in three different experimental settings (see Vogt, 2005a, for a statistical analysis). In experiment I, the CFs were predefined and the language was transmitted without a bottleneck. As the graph shows, in this experiment, compositionality rapidly emerged to a level around 0.55 and remained stable at this level. In experiment II, the CFs were not predefined and the language was again transmitted with no bottleneck. Here compositionality rapidly increased to a level near 0.8, but collapsed a little later within a few itera-

[c]These are the same parameter settings used in Vogt (2005a, 2005c).

tions to a level of 0. In experiment III, again the CFs were not predefined, but this time the language was transmitted through a bottleneck in which the agents only communicated about 50% of all possible objects. Clearly, when the language was transmitted through a bottleneck, compositionality kept rising until a stable system emerged, with a level of compositionality above 0.9. A similar result is achieved if the CFs are predefined and the language is transmitted through a bottleneck (not shown here, but see, Vogt, 2005a).

The three other graphs in Figure 2 show the typical dynamics of a word that is used to express a certain shape. These graphs were generated by plotting at each 50th guessing game, for each agent, the preferred meaning (i.e. prototypical value of the CF) of a hand-selected word for shape that the population used with a relatively high degree of success. When the CFs are predefined (experiment I), all agents use the same meaning invariably over time. The two other graphs, in which the CFs are not predefined so that these languages are subject to overextensions, show that different agents quite frequently tend to use the same word to express different meanings. For experiment II, the words for expressing shape (or colour for that matter) tend to die when the language becomes holistic, from generation 50 onward. For experiment III, however, there is a clear tendency for the majority of agents to prefer the same meaning for a word.

4. Discussion

In this paper the effect of overextensions on the emergence of compositionality in language is investigated. The simulations show that when agents are subject to overextending words as a result of the incremental construction of categories, compositionality tends to emerge to a higher degree but the population have more difficulty in arriving at a shared system. The latter is due to the fact that when children hear a particular word early in their development, this word may be associated with a category that has a wide scope. Once the children have acquired all categorical features in a given dimension, this word may then be associated with the wrong meaning. Nevertheless, due to the limited number of interactions among the agents, such associations may survive. This may also happen because other agents may be able to understand such a word in a particular context. As a result, the meaning of a word can drift from one position in the conceptual space to another over time (cf. Fig. 2 bottom left).

When there is no bottleneck on linguistic transmission, the compositional systems that tend to emerge initially all collapse after a period of time. As argued elsewhere, this has to do with the lack of pressure to form compositional structures when the need for them is not there, which is the case if children can learn from the entire language of their predecessors (Kirby et al., 2004; Vogt, 2005c). However, the current study – as well as the one presented in Vogt (2005a) – shows that when there are no overextensions during development, this collapse does not occur. So it is not only the lack of pressure due to the absence of a bottleneck, but

also the additional difficulties in arriving at a shared system due to overextensions that make compositional structures less stable than holistic ones. Intuitively, this can be understood by realising that a meaning drift in one dimension (i.e. linguistic category) of a compositional system affects a larger part of the language than a meaning drift in one dimension of a holistic system (see also, Vogt, 2005c).

Concluding, overextensions arising from the ontogenetic development of meanings in this model do indeed cause both a positive and a negative effect on the emergence of compositional structures. Now, is it possible to extrapolate from this study to the case of human language evolution (and its acquisition)? This question is hard to answer, because the current model is a highly simplified model of human language evolution. Perhaps the easiest part is to find evidence for the positive effect. If the hypothesis is extensible to natural language, the results would predict, e.g., that children – while going through the phase of overextensions – become increasingly proficient at forming compositional structures. Currently, research is underway investigating to what extent such a tendency can be detected in child language acquisition. The negative effect that language becomes more unstable is harder to assess, because this effect occurs only in the absence of a bottleneck, which seems very unlikely for young children learning language (Vogt, 2005b). However, on the other hand, the model would predict that due to overextensions, differences in preferences on language use would emerge, though, again, this will be very hard to assess empirically.

References

Briscoe, E. J. (Ed.). (2002). *Linguistic evolution through language acquisition: formal and computational models.* Cambridge: Cambridge University Press.

Clark, E. V. (2003). *First language acquisition.* Cambridge: Cambridge University Press.

Gärdenfors, P. (2000). *Conceptual spaces.* Bradford Books, MIT Press.

Kirby, S., Smith, K., & Brighton, H. (2004). From UG to universals: linguistic adaptation through iterated learning. *Studies in Language, 28*(3), 587-607.

Steels, L., Kaplan, F., McIntyre, A., & Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language.* Oxford, UK: Oxford University Press.

Vogt, P. (2005a). Meaning development versus predefined meanings in language evolution models. In L. Kaelbling & A. Saffiotti (Eds.), *Proceedings of IJCAI-05* (pp. 1154–1159). IJCAI.

Vogt, P. (2005b). On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. *Adaptive Behavior, 13(4)*, 325–346.

Vogt, P. (2005c). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence, 167(1–2)*, 206–242.

GRAMMATICALISATION AND EVOLUTION

HENK ZEEVAT
ILLC
University of Amsterdam
henk.zeevat@uva.nl

Grammaticalisation is relevant for language evolution in two ways. First, it is possible to model grammaticalisation processes by evolutionary simulations (iterated learning). This paper provides two such models of a central step in the grammaticalisation process: the recruitment of lexical and functional words for a new functional role. These models help in better understanding the processes involved. Second, it is possible to reason backwards to earlier stages of human language. The paper argues that all that is necessary for the genesis of natural languages is the conventionality of the form-meaning association and the possibility of introducing new lexical words. Once there is a communication system of this kind, all the additional complexities of human languages follow.

1. Grammaticalisation

Functional items in natural languages comprise prepositions, particles, auxiliaries, determiners, pronouns of different kinds and inflectional morphology. To the extent that their etymology is clear, they are -often phonologically reduced- versions of lexical nouns and verbs, one of the reasons why it is generally believed that all functional items come from lexical words. It is also hard to see in what way one could introduce a word for the meanings of functional items, since it is impossible to establish joint attention to abstract concepts like negation, past, possibility or uniqueness without linguistic means for expressing these concepts. The process by which lexical words change into functional items is called grammaticalisation and examples of it have been extensively studied by historical linguists. The following general characteristics (Pagliuca, 1994), (Traugott, 1993)) are standardly assumed:

1. Bleaching of the meaning of the word towards a weaker, vaguer and more pragmatic meaning.
2. Rise in frequency and obligatoriness
3. Phonological and syntactic reduction

Let me try to illustrate these properties by a simple example. The article $a(n)$ transparently derives from the cardinal *one*. *One* is more optional in the sense that it never appears just for syntactic reasons as $a(n)$ does. In consequence, the frequency of $a(n)$ is also much increased with respect to that of *one*. The meaning

of *one* can be characterised as saying that the intersection of the denotations of the noun and the predicate has precisely one member. The meaning of $a(n)$ is often described as: the referent of the complex phrase is unfamiliar to the hearer. This is weaker, vaguer and more pragmatic. Finally, it is clear that there is a phonetic reduction both in the loss of a vowel feature and in the optionality of the final nasal.

The targets of grammaticalisation are not arbitrary. The typology of human languages includes aspect and tense marking, modality, particles, case systems, pronouns and prepositions and while there may be vast differences in the inventories of different languages both in the concepts for which a functional item is present or in the category in which it is realised, there are very substantial overlaps in the functions that get marked. These overlaps are brought out by the semantic map methodology (Croft, 2003), (Haspelmath, 2003), (Auwera & Vladimir A. Plungian, 1998), (Malchukov, 2004). The concepts expressed are central and the conclusion that the functional items are needed because otherwise the expressivity of our languages would be insufficient for the purposes that we pursue with our linguistic communication is unavoidable. I will here not model phonetic reduction.

There can be no proof that the models presented here are correct, but only that something analogous to grammaticalisation happens under the described conditions. On the other hand, an informal concept that cannot be underpinned by an evolutionary reconstruction is flawed. There is at the same time ample space for other models of grammaticalisation, both within the same framework ("Gricean evolution") or in other concepts of evolution, but I am not aware of any other work.

2. Basic Concepts

Meanings are linked to forms by a convention. A corpus is —in the contexyt of this paper— a collection of such conventions that has one record for every time a certain meaning is used with a certain form. A corpus can be represented by an assignment of probabilities to form-meaning pairs. $p(Form, Meaning)$ is the number of times that $Form$ was used meaning $Meaning$ divided by the total number of times anything was used with any meaning. A corpus can then be represented by a function $f : Forms \times Meanings \rightarrow [0,1]$ such that $\Sigma_{Form \in Forms, Meaning \in Meanings} f(Form, Meaning) = 1$

The corpus is taken to determine both how a speaker would express a meaning and how a hearer would interpret a form. The speaker selects a form for a meaning according to the probability that that form is used for that meaning. I.e. if the speaker wants to express M the probability that she will select F to express it is $\frac{p(F,M)}{\Sigma_{G \in Forms} p(G,M)}$.

Similarly the hearer will select the meaning M for the form F with the probability $\frac{p(F,M)}{\Sigma_{N \in Meanings} p(F,N)}$.

A communication act starts with the speaker selecting a meaning for communication. The speaker selects this meaning as speakers do, i.e. with a probability that can also be determined from the corpus as the probability $\Sigma_{F \in Forms} p(F, Meaning)$. This reflects the natural frequency of the meaning and reflects the propensity of speakers to select the meaning $Meaning$. We identify the natural frequency with its value in the first corpus. Natural frequency could in principle be determined by looking at a set of corpora for different languages, under the assumption that the natural frequency of meaning is an invariant over languages.

A communication act is successful iff the hearer will correctly interpret the form as having the meaning the speaker intended to communicate with her expression. The corpus representing the next generation will consist of only the successful communications. This reproduces $p(F, M)$ as $natural_frequency(M) * \frac{p(F,M)}{\Sigma_{G \in Forms} p(G,M)} * \frac{p(F,M)}{\Sigma_{N \in Meanings} p(F,N)}$. Normalisation to 1 gives the next corpus. Evolution is modeled by iterating this process thus following the paradigm of iterated learning (Hurford, 2002).

This can be called Gricean evolution (because it employs the Gricean criterion of success in communication from (Grice, 1957)) or bidirectional evolution (because it is related to optimality theoretic bidirectionality (Blutner & Zeevat, 1994)). The next two notions are corrections on the notion of success. The first is $Importance$. A semantic feature is **important** if not recognising it when it is intended is worse than wrongly assuming it is there when it is not. (Though strictly speaking neither is successful.) Let M and M' be such that M is M' without the important semantic feature. In that case if M is chosen when it should have been M' is just failure whereas choosing M' when it should have been M is still somewhat OK, perhaps half of full success.

A good example of an important feature is the speech act of correction. Corrections need to be processed differently from straight assertions because the corrected material needs to be removed (or be made harmless in other ways), so it is important to recognise it. Wrongly assuming that one is dealing with a correction is not problematic: there is just nothing to remove. But not recognising a correction would lead to inconsistent information.

Ambiguities are the causes of lack of communicative success. But ambiguities come in flavours. Some ambiguities are protected by pairs of presuppositions that —in case the presuppositions are part of the given information as they should be— guarantee that the hearer gets the right reading. We can call this an **protected ambiguity** and correct the success rates as follows. Let F be an isolated ambiguity between M and M'. Then the chance that the hearer gets it right for either M or M' is $\frac{p(F,M)+p(F,M')}{\Sigma_{N \in Meaning} p(F,N)}$ (or close to that).

The final notion to be introduced is weak entailment. This is a probabilistic logical notion that is defined by: M weakly entails M' iff $p(M'|M) >$

$p(\neg M'|M)$. It is just a property of the initial probability assignment: $\Sigma_{F \in Forms} p(F, M \wedge M') > \Sigma_{F \in Forms} p(F, M \wedge \neg M')$. Weak entailment can be due to many different relations, such as generalised conversational implicature, default inferences (ravens are black), causal reasoning (glass breaks if it falls on hard floors) and others. The negation must sometimes be interpreted as the absence of the feature, e.g. the negation of correction is a proper non-correcting assertion.

3. The Weakening Model

Suppose:
F means M and M weakly entails M'
M is less frequent than $\neg M$
M' is less frequent than $\neg M'$
M' is important.

Then *ceteris paribus* and eventually, F will start meaning M'. If moreover $\neg M \wedge M'$ is more frequent than M it will take over F entirely (**usurpation**), otherwise F will be ambiguous between M' and M (**spread**).

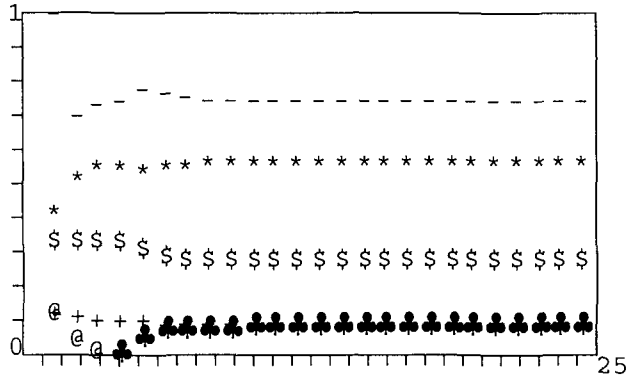
Ceteris paribus forbids the presence of other elements that could express M', *eventually* indicates that it happens after a number of generations when the model reaches stability.

The main reason why the change occurs is because the meaning $\neg M \wedge M'$ is dominated by $\neg M \wedge \neg M'$ as a meaning for zero expression. It is bad to interpret something as its non-dominant meaning and it becomes worse. As it goes on, it negatively affects the choice of zero-marking as a means of expression of $\neg M \wedge M'$ in favour of its competitor F. Since F is more successful (M' is important) F as a means of expression of $\neg M \wedge M'$ grows and will start meaning it more and more often. The growth is limited by the natural frequency of $\neg M \wedge M'$ and this determines whether usurpation will happen or not.

The following is a picture produced by a simulation. The original corpus frequencies are:
zero, $\neg M \wedge \neg M'$, 200
zero, $\neg M \wedge M'$, 100
zero, $M \wedge \neg M'$, 1
zero, $M \wedge M'$, 1
F, $\neg M \wedge \neg M'$,1
F, $\neg M \wedge M'$, 1
F, $M \wedge \neg M'$,20
F, $M \wedge M'$, 50

M''s importance makes it worse not to recognise M' than to overrecognise it. This favours means of expression which are more biased to recognising M'. The value is here set to 0.5. E.g. if one tries to express $X \wedge \neg M'$ and the hearer

recognises $X \wedge M'$, it is still half right.



Spreading grammaticalisation of F to start meaning M without M (♣♣♣♣♣). M without M starts out by being zero-expressed (). The zero-expression is eventually monopolised by the absence of M and M' (). The uses of F for M and M' ($$$$$) and for M without M' (+++++) are reduced but preserved.

The model explains the rise of frequence of the grammaticalised item, both on spread and on usurpation. Weak entailment takes care of the weaker, vaguer and more pragmatic meaning, with spread responsible for the extra vagueness.

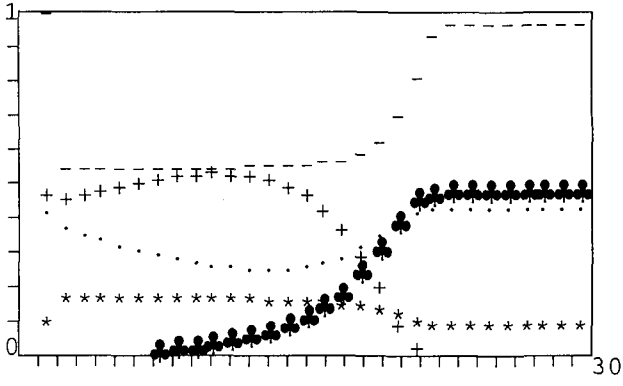
Spread in the recruitment of functional items is responsible for the emergence of the lexicographical nightmares like prepositions, cases, certain aspect classes and certain particles. Usurpation of functional items in its turn leaves behind an expressive gap which will be filled in by new recruitments.

The major conflict with what is known about grammaticalisation processes is the assumption that there is nothing available for expressing the important new meaning. If one adds a good expressive possibility to the model, nothing will happen. But this situation seems to occur with a reasonable frequency (Pagliuca, 1994). It is probably necessary to see the alternative expressive possibilities as bad, at least for weakening.

Metaphor is different because it does not involve weak entailment of the new meaning but especially because it gives very good expression alternatives in the form of a protected ambiguity. Metaphorical expression works only in a context where it is clear that the literal interpretation cannot apply. In this situation the intended interpretation is the most strongly suggested alternative. The notion of suggestion based on similarity and analogy cannot be modelled inside a statistical model. Both the old meaning and the new meaning are fully protected from each

other in this case. If the context allows the old meaning, that meaning will be chosen, if the context doewas not allow it, the new meaning will be chosen.

In the metaphor model, there is an ambiguous way of expressing the target meaning, a form shared by the target meaning and a distractor meaning. Initially, the carrier of the metaphor has its old meaning, with the metaphorical meaning being a rare event. Since these two meanings are protected from each other, the metaphorical use of the carrier is more successful than the old ambiguous expression for the target meaning. Protection can be modelled by twisting the success rates: the source and target meanings of the carriers are just added.



Grammaticalisation by metaphor. Initially the meaning M shares a form G with a distractor meaning D (+++++ and .....) and M is also a rare metaphoric interpretation of form F meaning M (♣♣♣♣♣♣ and *****). The success of the metaphoric expression of M leads to its becoming the standard way of expressing M and the monopolisation of G by the distractor meaning D.

4. Language Evolution

The grammaticalisation events modelled in the last two sections happen under circumstances that are not rare at all. It seems safe to say that a human language without functional inventory is inherently unstable: there are lots of important distinctions (in the sense of section 2) that go unexpressed and will attract weakening and metaphorical grammaticalisation. Adding phonological decay and syntactic evolution, such a language will evolve into something like the human languages we know: with verbal and nominal morphology, discourse particles, conjunctions, prepositions and clitics. Also with grammatical meanings like modality, tense, evidentiality, mood, case and thematic roles. The study of word order freezing

((Jakobson, 1984) (Lee, 2001) and (Zeevat, to appear) indicates that the conditions on word order arise naturally under functional pressure and can explain the arisal of permanently frozen constructions as one finds in e.g. English or Chinese from the weaker word order tendencies that one finds in Sanskrit, Korean or Russian. While many of the processes are only partially understood and formal analyses are almost completely lacking, it seems that the application of the iterated learning method for modeling these processes has serious potential. I hope to have made a case for that in the preceding sections.

One can also reason backwards to the minimal conditions on languages for grammaticalisation to start. If it is possible to adopt new words with lexical meanings and if the words can be combined into complex messages, one obtains the inherently unstable language in which grammaticalisation will start. So those are the only two things that biology needs to account for.

References

Auwera, J. van der, & Plungian, V. A. (1998). Modality's semantic map. *Linguistic typology, 2*, 79-124.

Blutner, R., & Zeevat, H. (1994). *Optimality theory and pragmatics.* Palgrave MacMillan.

Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar. tense, aspect, modality in the languages of the world.* University of Chicago Press.

Croft, W. (2003). *Typology and universals.* Cambridge University Press.

Grice, H. (1957). Meaning. *Philosophical Review, 67*, 377-388.

Haspelmath, M. (2003). The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In M. Tomasello (Ed.), *The new psychology of language* (p. 211-243). New York.

Hopper, P., & Traugott, E. (1993). *Grammaticalization.* Cambridge University Press.

Jakobson, R. (1984). Morphological observations on slavic declension (the structure of russian case forms. In L. R. W. . M. Halle (Ed.), *Roman Jakobson. Russian and Slavic grammar: Studies 1931-1981* (p. 105-133). Mouton de Gruyter.

Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: an overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148). Springer.

Lee, H. (2001). Markedness and Word Order Freezing. In P. Sells (Ed.), *Formal and empirical issues in optimality-theoretic syntax.* CSLI Publications.

Malchukov, A. L. (2004). Towards a semantic typology of adversative and contrast marking. *Journal of Semantics, 21*, 177-198.

Zeevat, H. (to appear). Freezing and marking. *Linguistics.*

STAGES IN THE EVOLUTION AND DEVELOPMENT OF SIGN USE (SEDSU)

JORDAN ZLATEV

Lund University, Department of Languages and Literature, Box 201, 221 00 Lund, Sweden

THE SEDSU PROJECT*

Centre for Cognition, Computation and Culture, Department of Psychology, Goldsmiths, University of London London, SE14 6NW, UK

We present the rationale and ongoing research of an interdisciplinary international project aiming at developing a novel theory of semiotic development, on the basis of broad developmental, cross-species and cross-cultural research. We focus on five social-cognitive domains: (i) perception and categorization, (ii) iconcity and pictures, (iii) space and metaphor, (iv) imitation and mimesis and (v) intersubjectivity and conventions, each of which is briefly described. Our main hypothesis is that what distinguishes human beings from other animals is an advanced capacity to engage in sign use, which on its part allowed for the evolution of language.

1. Introduction

There is no consensus about what makes humans intellectually and culturally different from other species, and even less so concerning the underlying sources of these differences. The main hypothesis of the project *Stages in the Evolution and Development of Sign Use* (SEDSU) is that it is not language per se, but an advanced ability to engage in sign use that constitutes the characteristic feature of human beings. In particular, this implies the ability to differentiate between the sign itself, be it gesture, picture, word or abstract symbol, and what it represents, i.e. the *sign function* (Piaget, 1945), and thus to use (the same) sign systems for both communication and cognition. The SEDSU project is highly interdisciplinary, involving developmental and cognitive psychologists, linguists, philosophers, primatologists, and semioticians from five European countries and Brazil, and fieldwork in Europe, South America, Africa and Asia. This single research effort affords new possibilities for methodological innovation, and the collection and analysis of developmental, cross-cultural and cross-species data in a joint theoretical framework.

---

* Ingar Brinck (Lund University), Josep Call (MPI-EVA Lepizig, Partner Leader), Jules Davidoff (Goldsmiths, Project Coordinator), Christine Deruelle (INCM-CNRS Marseille), Joël Fagot (INCM-CNRS Marseille, Partner Leader), Peter Gärdenfors (Lund University), Pam Heaton (Goldsmiths), Stephen Nugent (Goldsmiths), Patrizia Poti (ISTC-CNR Rome), Vasu Reddy (Univesity of Portsmouth), Wany Sampaio (Federal University of Rondônia) Chris Sinha (University of Portsmouth, Partner Leader), Göran Sonesson (Lund University), Giovanna Spinozzi (ISTC-CNR Rome, Partner Leader), Elisabetta Visalberghi (ISTC-CNR Rome), Jörg Zinken (University of Portsmouth)

Our central research objective is to investigate the developmental and comparative distribution of semiotic processes and their effect on cognition. For this purpose we have singled out *five social-cognitive domains* and study their interrelations and role in the development of sign use (see Section 2). These domains are all characterised by stage-like developmental profiles that correlate with differences in sign use. The investigations in the different domains are being carried out in parallel, with extensive sharing of methodologies and results. Our ultimate goal is to integrate all the results of the SEDSU project in a coherent new theory of semiotic development, placing the question of the evolution of language in a broader perspective. In this article, we outline our general theoretical orientation, describe some of our ongoing work in each of the five social-cognitive domains, and outline how it contributes to an integrated theory of semiotic evolution and development.

2. Sign use and the five social-cognitive domains

Research in the last decades has established significant continuities between humans and non-human species, particularly primates. Nevertheless, when it comes to determining what makes humans unique, it is often claimed that there is one ability – language – that makes human beings special (Christiansen & Kirby, 2003). However, it could be argued that there are more basic differences between our species and others; for example, representational activity (Piaget, 1945), mimesis (Donald, 1991), and understanding (communicative) intentions (Tomasello, 1999). We would suggest that all these proposals crucially involve *differential abilities in sign use.* Taking a semiotic perspective and distinguishing between different types of sign systems on the basis of factors such as expression-meaning relation (icon/index/symbol), intentionality, conventionality and complexity permits a gradient approach. This enables us to characterise their emergence in terms of *stages*, allowing us to situate discontinuities between human and non-human cognition and communication within a broadly continuous evolutionary-developmental framework. Furthermore, studying sign use allows us to scrutinise the semiotic capacities of other species, pre-linguistic and impaired children.

In the SEDSU project we investigate a number of social-cognitive domains characterised by stage-like profiles, where some transitions are more quantitative, while others appear to be qualitative. The domains are: *perception and categorisation*, *iconicity and pictures*, *space and metaphor*, *imitation and mimesis* and *intersubjectivity and conventions.* While these may be studied separately, we would argue that they interact so closely in both evolution and ontogeny, that an integrative approach is required. In order to provide an account of the link from individual attention to joint linguistic reference we must inquire into the differences between perceptual and linguistic discrimination, the role of pictures as signs, the conceptualisation of space, the relation between imperative and declarative pointing and the role of bodily mimesis.

2.1 Perception and categorization

In studying this domain, we consider the possible reorganization of information around a focus of attention as a function of sign use. In order to visually identify objects and segregate them from the background, organisms must be able to group their component parts into perceptual wholes. Comparative studies, however, point to important differences between humans and non-human primates. For example, faced with hierarchical stimuli, several primate species, such as tufted capuchins (Spinozzi, De Lillo & Truppa, 2003) and chimpanzees (Fagot & Tomonaga, 1999) process the local details better than the global structure. These findings contrast sharply with the well-known phenomenon of "global advantage" showed by humans. Our hypothesis is that this difference relates to sign use in general, and linguistic performance in particular.

Recent cross-linguistic and phylogenetic investigations (Davidoff, Davies & Roberson, 1999; Fagot, Goldstein, Davidoff & Pickering, in press) have also shown a linguistic basis to performance on what again might appear to be solely perceptually based tasks. These studies have indicated that cultural and linguistic training "distorts" perception by stretching perceptual distances at category boundaries. Such effects that depend on both discrimination between categories and identification within category boundaries allow objects to be recruited for sign use by labelling (Brinck, 2003). To further scrutinise the interaction between perceptual processing and sign use we are exploring phylogenetic and developmental trends in perceptual categorisation tasks. These studies were designed so that they could be comparatively conducted in non-human primates and in different groups of children (normal, autistic and deaf).

The question remains whether global categorization has been selected for in primate and hominid evolution and can account for some of the difficulties that children with autism encounter with language acquisition. Our preliminary results show a complicated pattern with respect to our target populations. The Marseille group focussing on visual stimuli, have shown that children with autism show a local, as opposed to global, processing bias, which is also the case for baboons. Chimpanzees, in contrast show some intermediary performance. The Goldsmiths group have collected new evidence for enhanced local colour memory in cognitively impaired children with autism. However, they have shown that, while autistic children exhibit a local bias, this does not prevent normal global processing within the musical domain (Heaton, in press). To complicate matters further, there is tentative evidence that the Himba from Namibia also have a local processing bias in the visual domain. So, it remains to be shown how categorization might vary under these processing differences.

2.2 Iconicity and pictures

According to classical semiotic theory (Peirce, 1931-58) icons are signs that resemble the thing for which they stand, indices are signs that are connected to

their referent by means of some independently known or perceived relationship; symbols, on the other hand, are conventional. It has therefore often been argued that icons and indices are elementary phenomena, common to most animals, while symbols are unique to the human species. In order to grasp the similarities and differences in the sign use of human beings, other species, children and individuals suffering from disorders of the semiotic capacity, we separate the properties of iconicity, indexicality, and symbolicity *per se* from the sign function, defined by Piaget (1945) in terms of differentiation between expression and content. Iconicity and indexicality could conceivably be simple properties accessible to many animals, giving rise to the perception of sameness and/or category membership, and S-R relations, respectively. In contrast, the use of iconic signs such as pictures appears to be a highly sophisticated capacity only found in humans and perhaps some higher primates.

A picture is a surface equipped with markings giving rise to a vicarious perception of objects and actions of the perceptual world (Gibson, 1982). In order to see a picture as a picture, i.e., as a sign, it is necessary to perceive at the same time the similarity and the difference between the surface and that which it depicts; this, according to Gibson, is a capacity only found in human beings. In order to investigate Gibson's surmise, we distinguish *primary iconical signs* in which the perception of similarity *precedes* the knowledge of a sign relationship between picture and depicted, and *secondary iconical signs*, in which the opposite is the case. Primary iconical signs such as pictures seem to presuppose a distinction between two-dimensionality and three-dimensionality (Sonesson, 2000), which has independently been shown to be difficult to grasp for at least some non-human primates (Barbet & Fagot, 2002). Donald (1991) has suggested that picture use follows language and requires the ability to handle organism-independent representations, which originate with pictures but at later stages render possible writing and theoretical thinking. If so, language may conceivably be a necessary, but not a sufficient, condition for the development of organism-independent representations such as pictures. However, this view is contradicted by experimental investigation of picture use in non-human primates, suggesting that differentiation is possible at least in enculturated chimpanzees. We are currently conducting experiments attempting to show picture-as-sign understanding in (non-enculturated) baboons and chimpanzees.

2.3 Space and metaphor

The spatial domain has been central to recent research into the origins of symbolization, the cognitive foundations of language, and the motivation of linguistic conceptualisation by both universal and culturally specific cognitive processes. Landmarks are perceptible environmental elements or objects that can be used to locate hidden goals. It has been suggested that appreciating the spatial-designation function of landmarks indicates achieving a "symbolic" understanding and that practical achievements in the domain of spatial cognition

such as using landmarks could be a pre-requisite for identifying spatial relations in language. Since nonhuman primates use landmarks to locate objects in space (e.g., Potì, Bartolommei, Saporiti, 2005), we are assessing to what extent this use is based on different cognitive processes or on different levels of the same process as in humans, which would also have implications for the relations between spatial language and spatial cognition in humans.

It has been proposed that properties of the primate spatial cognitive system directly motivate properties of spatial language, giving rise to strong universals (such as the closed class/open class distinction) and constraints on typological variation. Clearly, such claims need to be evaluated against comprehensive linguistic data. The semantic and cognitive domain of space has been paradigmatic in cognitive typology. One aspect of language variation that has been subject to extensive cross-linguistic study from a cognitive perspective recently is *motion-event typology*, i.e. the way different languages frame events of translocation. Our research will deepen our existing analyses focussing on Amondawa (Sampaio et al, in press) and Thai. The spatial domain has also been adduced in support of strong claims for linguistic and cognitive universals. There has been much research on such hypothesised universals in metaphorical mapping from the conceptual domain of space onto conceptual domains that are less accessible to experience; however, details of that mapping vary considerably. Specifically, recent research suggests that the cultural conventions entrenched in a particular language might be more important than previously thought. Our research extends the database to allow a comprehensive understanding of sign use in spatial conceptualisation and metaphor.

2.4 Imitation and mimesis

Within the chain of the usually recognised stages from ritualised movements, imperative pointing to declarative pointing, the relationship between expression and content becomes sufficiently distinct to allow the emergence of the sign function. However, imperative pointing can be shown to arise from ritualisation, while (human) declarative pointing emerges by imitation (Brinck, 2003). It has also not been sufficiently well explained how the ability to imitate gestures and use them in intentional communication relates to action understanding and co-operation (Brinck & Gärdenfors, 2003). We hold that the concept of *bodily mimesis* (Donald 1991) can help us reach a better understanding of these stages in the use of gesture. We distinguish between a dyadic form of mimesis, the clearest form of which is imitation, and triadic mimesis, where someone mimes something for someone else, e.g. pantomime (Zlatev, Persson & Gärdenfors, 2005). Research has shown that apes, especially those raised and trained by humans, are capable of mimesis in its dyadic form (Call, 2001). In contrast, it does not seem that apes are also capable of triadic mimesis in the form of iconic gestures or declarative pointing (Tomasello et al, 1997) but there is some evidence to the contrary. We are currently investigating the basis for the

differences in the mimetic skills of apes and humans. In particular, we are focusing on the ability to use imitation to acquire novel communicative signs.

Furthermore, we are investigating whether other mechanisms than imitation could be involved in the rise of the first communicative gestures of pre-linguistic children. One possibility is that children could create novel representational acts on the basis of the similarity of the observed objects or events, i.e. on the basis of primary iconicity (see 2.2 above). Evidence for this would be if children from (widely) different linguistic and cultural environments have similar gestures. To study the role of cultural transmission for the emergence of children's gestures we are comparing longitudinal data consisting of spontaneous videotaped interactions between caregivers and children from Thailand and Sweden.

2.5 Intersubjectivity and conventions

The goal in this domain is to define the progressive emergence of intersubjectivity in evolution and ontogeny as well as to study the role of culture-specific patterns for the formation of conventions. The two are intimately related since intersubjectivity involves the ability to share the mentality of others and conventions exist as a form of shared, common knowledge. A basic form of intersubjectivity involves the awareness of others' feelings and attention to oneself; this requires both a species-general capacity for empathy (Preston & de Waal, 2002), but also engagement in acts of mutual attention, displayed in phenomena such as eye-contact, intense smiling, coyness, calling vocalizations and showing-off (Reddy, 1991). Careful comparisons of videotaped episodes of mother-infant interactions in humans and non-human apes will show to what extent such behaviours are specific for our species. A second developmental and possibly evolutionary stage of intersubjectivity involves the ability to understand the intentions of others. Children master this second stage around the age of one, and newer evidence and analyses show that chimpanzees too achieve this level (Hare, Call & Tomasello, 2001), at least in competitive contexts. A third stage involves understanding others' attention to one's own attention and communicative intentions. It has been suggested that apes cannot master this in cooperative settings, but this has not been explored in the context of mother-infant interaction. Experiments with food sharing between ape mothers and infants, in various contexts, are being conducted in order to test their potentials for collaboration and gestural communication.

Understanding the relationship between sign use and intersubjectivity is further enhanced by a cross-cultural investigation of the framing of compliance in early parent-infant interactions in two different cultural environments (Portsmouth, UK and Hyderabad, India). Compliance, considered a sign of developmental and interpersonal maturity by Western psychology, is in fact an intrinsically relational and culturally variable achievement. For infants to become aware that they may need to amend their own actions in relation to others' intentions, they not only need a certain level of developmental maturity,

but an environment where others are in fact communicating such intentions. This requires not only a belief in the desirability of compliance but also in its possibility, beliefs which vary between different situations and cultures. The "Western" focus on consistency in parental actions and on the positive correlates of child compliance neglects the complexity of communication in such engagements, particularly in Asian cultures where negotiation tends to predominate over rules even in childhood (Reddy, 1983). We use parental recognition, emphasis and negotiation of different situations as a frame for the understanding of intentions and the emergence of sign use.

3. Conclusions

The investigations in the different social-cognitive domains described in this article are being conducted in parallel, with extensive sharing of methodologies and results. Since we hold that each domain plays a key role in providing cognitive prerequisites for the development of sign use, and at the same time is transformed by the acquisition of the latter, we expect to find considerable similarities and interactions between developments in the domains. Finally, we plan to integrate all the results in a *coherent theory of semiotic development* in which we (a) identify stage-like transitions within each one of the five social-cognitive domains, (b) investigate interactions, dependencies and synergies between such transitions across the different cognitive domains and (c) relate such transitions to sign use, both in terms of precursors and prerequisites and in terms of the transformations wrought in the domains by the acquisition and development of semiotic skills. Our contention is that such a theory is hitherto lacking. Even though the SEDSU project is only 9 months old, we are confident that due to its interdisciplinary, integrative character it will at least contribute to such a theory, and hence, to explaining the evolution of language.

References

Barbet, I. & Fagot, J. (2002). Perception of the corridor illusion by baboons. *Behavioural Brain Research, 132*, 111-115.

Brinck, I. (2003). The pragmatics of imperative and declarative pointing, *Cognitive Science Quarterly, 3(4),* 2003, 429-446.

Brinck, I. & Gärdenfors, P. (2003). Co-operation and communication in apes and humans", *Mind and Language, 18(5)*, 484-501.

Call, J. (2001). Body imitation in an enculturated orangutan (*Pongo pygmaeus*). *Cybernetics and Systems, 32(1-2)*, 97-119.

Christiansen, M. H. & Kirby S., (2003). (Eds.) *Language evolution.* Oxford: Oxford University Press.

Davidoff, J., Davies, I. & Roberson, D. (1999). Colour categories of a stone-age tribe. *Nature, 398*, 203-204.

Donald, M. (1991). *Origins of the modern mind. Three stages in the evolution of culture and cognition,* Cambridge, Mass.: Harvard University Press.

Fagot, J. & Tomonaga, M. (1999). Comparative assessment of global-local processing in humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*): Use of a visual search task with compound stimuli. *Journal of Comp. Psychology, 113*, 3-12.

Fagot, J., Goldstein, J., Davidoff, J. & Pickering, A. (in press). Cross species differences in colour categorisation. *Psychonomic bulletin and review.*

Gibson, J. (1982). *Reasons for realism. Selected essays of James J. Gibson,* E. Reed & R. Jones (Eds.). Hillsdale, NJ: Lawrence Erlbaum.

Hare, B., Call, J., & Tomasello, M., (2001). Do chimpanzees know what conspecifics know and do not know? *Animal Behaviour, 61*, 139-151.

Heaton (in press). Interval and contour processing in autism. *Journal of Autism and Developmental Disorders.*

Peirce, C. S. (1931-58). *Collected Papers I-VIII.* Hartshorne, C, Weiss, P, & Burks, A, (Eds.). Cambridge, MA: Harvard University Press.

Piaget, J. (1945). *La formation du symbole chez l'enfant.* Neuchatel: Delachaux & Niestlé. Third edition 1967.

Potì, P., Bartolommei, P. & Saporiti, M. (2005). Landmark use by *Cebus apella. International Journal of Primatology, 26 (4)*, 921-948.

Preston, S. D. and de Waal, F. B. M. (2002). Empathy: its ultimate and proximal causes. *Behavioral and Brain Sciences, 25,* 1-20.

Reddy, V. (1983). Responsiveness and rules: Parent-child interaction in Scotland and India. Unpublished PhD Thesis, University of Edinburgh.

Reddy, V. (1991). Teasing, joking and mucking about in the first year. In A. Whiten (Ed.) *Natural theories of mind* (pp. 143-158). Oxford: Blackwell.

Sampaio, W., Sinha, C. and da Silva Sinha, W. (in press) Mixing and mapping: motion and manner in Amondawa. In E. Lieven (Ed.) *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Slobin.* Mahwah, NJ: Lawrence Earlbaum Associates.

Sonesson, G. (2000). Iconicity in the ecology of semiosis. In T. D. Johansson, M. Skov & B. Brogaard (Eds.) *Iconicity - a fundamental problem in semiotics,* (pp 59-80). Aarhus: NSU Press.

Spinozzi G., De Lillo C., & Truppa V. (2003). Global and local processing of hierarchical visual stimuli in tufted capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology, 117*, 15-23.

Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard University Press.

Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The ontogeny of chimpanzee gestural signals: A comparison across groups and generations. *Evolution of Communication, 1*, 223-259.

Zlatev, J., Persson, T. & Gärdenfors, P. (2005). *Bodily mimesis as the "missing link" in human cognitive evolution. LUCS 121.* Lund: Lund University.

Abstracts

This page is intentionally left blank

ALARM CALLS AND ORGANISED IMPERATIVES IN MALE PUTTY-NOSED MONKEYS

KATE ARNOLD AND KLAUS ZUBERBÜHLER
School of Psychology, University of St Andrews, St Andrews, KY16 8PL, UK

1. Functional reference in primate alarm calling systems

Functionally referential alarm calling systems have been documented in a number of primate species. Vervet monkeys, Diana monkeys, Campbell's monkeys and ringtailed lemurs all produce at least two acoustically distinct alarm call types in response to different types of predators, usually ground and aerial predators (Seyfarth et al., 1980; Zuberbühler, 2000, 2001; Pereira & Macedonia, 1990). Redfronted lemurs and white sifakas also have a specific alarm call for raptors but produce a more general call associated with high arousal in the face of ground predators and to other forms of disturbance (Fichtel & Kappeler, 2002). Functionally referential systems exhibit a high degree of production specificity, discrete structure and context independence and have the potential to designate external objects or events. A functionally referential alarm calling system provides conspecific listeners with sufficient information about the eliciting stimulus to enable them to respond to alarm calls as though they had direct evidence of the presence of the predator and do not require additional contextual information for selection of the appropriate anti-predator response.

2. Alarm calling in male putty-nosed monkeys

We investigated the alarm calling system of wild putty-nosed monkeys in Gashaka Gumti National Park, Nigeria. We used playback methods to simulate the presence of two of their natural predators, crowned eagles and leopards. Male putty-nosed monkeys have two loud call types, 'pyows' and 'hacks'. Hacks were strongly associated with playbacks of eagle shrieks while pyows were commonly associated with playbacks of leopard growls. However, both call types occurred within alarm calling series to both predators. In addition, hacks were given to a wide range of disturbing stimuli including falling trees, unfamiliar loud noises, baboon fights and harmless birds. Pyows were given in

an even wider range of contexts and appear to have multiple functions including intergroup communication. Unlike alarm calling in other guenon species, the calls of male putty-nosed monkeys are not functionally referential and are, at best, only probabilistically associated with predators of different categories.

3. Alarm call sequences

When we examined the call series in detail, we found a number of regularities in calling patterns. The two call types formed part of three basic sequences: (a) hack sequences, consisting only of hacks, (b) pyow sequences, consisting only of pyows and (c) pyow-hack (P-H) sequences, consisting of between one and four pyows followed by between one and four hacks. These three basic sequences could be combined to form more complex call series. Transitional series consisted of a hack sequence followed by a pyow sequence while hack, pyow or transitional series could be interrupted by a P-H sequence at different locations. The insertion of P-H sequences appeared to follow certain rules. P-H sequences were inserted after around five hacks in an otherwise pure series of hacks, at the transition point in a transitional series, at the beginning of an otherwise pure series of pyows or they were given alone. In addition, the stereotypical ordering of the calls and temporal markers made them particularly conspicuous within call series. Furthermore, we have demonstrated, both experimentally and observationally, that P-H sequences function to instigate group movement in predatory and nonpredatory contexts. Whereas 'meaning' clearly does not reside in individual calls in this species, the P-H sequence offers the possibility that functional reference can evolve at higher levels of signal organisation.

References

Fichtel, C. & Kappeler, P. M. (2002). Anti-predator behavior of group-living Malagasy primates: mixed evidence for a referential alarm calling system. *Behavioral Ecology and Sociobiology, 51,* 262-275.

Pereira, M. E. & Macedonia, J. M. (1990). Ringtailed lemur antipredator calls denote predators, not response urgency, *Animal Behaviour, 41,* 543-544.

Seyfarth, R. M., Cheney, D. L. & Marler, P. (1980). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Animal Behaviour, 28,* 1070-1094.

Zuberbühler, K. (2000). Referential labeling in wild Diana monkeys. *Animal Behaviour, 59,* 917-927.

Zuberbühler, K. (2000). Predator specific alarm calls in Campbell's guenons. *Behavioral Ecology and Sociobiology, 50,* 414-422.

PERCEPTION ACQUISITION AS THE CAUSES FOR TRANSITION PATTERNS IN PHONOLOGICAL EVOLUTION

AU, CHING-PONG

Dynamique Du Langage, UMR 5596 CNRS, Université Lyon 2, ISH
14 avenue Berthelot 69363 Lyon Cedex 07, France

A computational model linking up developmental properties and sound changes was built, in order to seek for possible solutions for some controversial issues about the implementation of sound changes (Au, 2005). In the model, there is a population of agents. Each of them has a cognitive structure with four internal subsystems (perception, decoding, coding and production). The subsystems of the agents develop individually during development. The formation of perceptual categories is driven by statistical distributions of sounds that the new-born agents have listened to (e.g. Maye et al, 2002). A self-organizing map is used to simulate the category formation (Guenther et al, 1996).

In the simulation results of the model, two seemingly contradicting hypotheses on sound change transitions Neogrammarian regularity (lexically regular; Osthoff et al, 1878) and lexical diffusion (lexically irregular; Wang, 1969), can both be observed under different conditions. During a shift, the pronunciations of the lexical items change regularly as described in the Neogrammarian hypothesis; during a merger, the spoken forms display a regular pattern at the beginning, and then become irregular lexically as described in lexical diffusion. These conditions are primarily matched with the empirical data supporting the two opposing hypotheses. With further investigation on the subsystems of the agents, the **consistency of perceptual responses** among agents was found to be the causes of different transition patterns. At the later stage of a merger of two sounds, when two groups of words become acoustically close, the perceptual responses of individual agents become **inconsistent** throughout the population due to the statistically determined nature of perceptual development. The locations and the sharpness of boundaries between two categories vary and some agents may even have only one category across the acoustic range of the two original sounds. As the word pronunciations are learnt through self-listening, the spoken forms of various words are scattered along the acoustic range of the two original sounds. This is the basis of the irregularity; but when a perceptual category is still far enough from the neighboring categories, the category formed of each agent is similar and stable as in the shifts or the beginning stages of mergers. All spoken forms of the words in the same group are picked within the same small phonetic range bounded by the perceptual category. The category location in the acoustic domain may differ slightly from

generation to generation. When the acoustic differences accumulate, it appears that the spoken forms under the same perceptual category change simultaneously and gradually in the same direction as described in the Neogrammarian hypothesis.

In conclusion, the model here provides a more precise description on how the phonological systems evolve from time to time. If the present model is able to describe the reality appropriately, it can be potentially extended into a model that provides insights for the emergence of phonological systems.

References

Au, Ching-Pong (2005), *Acquisition and Evolution of Phonological Systems*. PhD Dissertation. City University of Hong Kong.

Guenther F. H. and Gjaja M. N. (1996), The Perceptual Magnet Effect as an Emergent Property of Neural Map Formation. *Journal of the Acoustical Society of America*, **100**, pp. 1111-1121.

Maye, J., Werker, J.F., & Gerken, L. (2002), Infant Sensitivity to Distributional Information Can Effect Phonetic Discrimination. *Cognition*, **82**(3), B101-B111.

Osthoff, H. and Brugmann, K. (1878), Morphologische Untersuchungen auf dem Gebiete der indo-germanischen Sprachen, Vorwort I. iii-xx. (English Translation in Lehmann 1967)

Wang, W. S-Y. (1969), Competing Changes as a Cause of Residue. *Language.* **45**:9-25.

THE EVOLUTION OF SYNTACTIC CAPACITY FROM NAVIGATIONAL ABILITY

MARK BARTLETT & DIMITAR KAZAKOV
Department of Computer Science, University of York, Heslington, York, YO10 5DD, UK

1. Syntax And Navigation

Many recent computational models (most notably those of Kirby (2002)) have shown how syntax may naturally emerge in language in order to exploit structural properties of a semantic space. However, while such models can explain why early human protolanguages may have gained in structural complexity to become full languages, they do not explain how the ability of individuals to handle compositionality of linguistic fragments evolved: while existing models explain the emergence of syntax in language, this is predicated on an existing syntax handling capability. We present one possible explanation for the evolution of this neurological under-pinning of syntax, and outline results from a computational model which has been developed to assess its feasibility.

We believe a link exists between motor and verbal sequence processing that may hold the key to the origins of syntax. We have previously discussed a model of navigation which demonstrates this link (Kazakov & Bartlett 2004), using landmarks as beacons and describing the path between two points by the list of landmarks one has to pass by on a journey from one position to another. One can devise a impoverished formalisation which represents such a map as a regular grammar, in which landmarks correspond to terminals, crossroads to non-terminals, and rules describe paths between two positions, e.g. the rule $Y \rightarrow X\, l_1\, l_2\, l_3$ states that to reach Y it is sufficient to be at X and then to pass by the three landmarks listed in order. With this representation, planning or following a path is equivalent to generating or parsing, respectively, a sentence of a regular language (RL). Should the navigational needs of individuals necessitate return along the same path as the outward journey, the navigational task requires a more complex formulation equivalent to a context-free language (CFL). The equivalence between the processor needed to understand these routes and a RL or CFL parser is important: if a parser was needed for navigation, it may have first evolved for this purpose. Once this parser was developed, only a relatively small change in the neural connections may have been required to make this parser available to the human brain speech circuitry.

This theory draws support from existing neurological research. Ullman (2004) pinpoints several memory circuits in the brain, the procedural memory, which are

associated with syntactic processing, and are distinct from declarative memory which stores information about facts and events, including the mental lexicon. The model suggests a common basis for the processing of verbal and non-verbal sequences which is supported by others, such as Hoen *et al.*(2003) who report that using non-verbal symbols to exercise the ability to reorder sequences, helps patients with speech difficulties to understand sentence that need to have their constituents rearranged in the same way (such as to form a passive sentence).

2. Evidence From Artificial Life

In order to test the evolutionary plausibility of this theory to explain the origins of linguistic syntactic ability, a second, supplemental theory, that one of the original purposes of language may have been for use in navigation, has been developed. From this, a multi-agent simulation has been created in which the relative performance of populations with differing behaviours are tested for their abilities to survive and reproduce. The behaviours in the model incorporate varying degrees of planning/parsing competence and those linguistic and navigational activities possible at each level. Experimental results indicate clear advantages, as manifested by greater population sizes, in those populations in which communication is permitted, especially when 'syntactic' navigation is used. In addition to using the model to essay the relative successes of these behaviours, the role of the environment structure in determining the benefit of a behaviour has also been examined. It has been established that populations able to communicate grow faster and are more resilient to volatility of resources than those unable to do so. Such results point towards a possible source of evolutionary pressure for the ability to use language. This, combined with the biological plausibility of adapting navigational abilities into syntactic handling skills for language, suggests that this theory be further considered as one possible mechanism to explain the origins of syntactic ability in humans.

References

Hoen, M., Golembiowski, M., Guyot, E., Deprez, V., Caplan, D., & Dominey, P. F. (2003). Training with cognitive sequences improves syntactic comprehension in agrammatic aphasics. *NeuroReport, 14*, 495-499.

Kazakov, D., & Bartlett, M. (2004) Co-operative navigation and the faculty of language. *Applied Artificial Intelligence, 18*, 885-901.

Kirby, S. (2002) Learning, Bottlenecks and the Evolution of Recursive Syntax. In: T. Briscoe (Ed), *Linguistic Evolution through Language Acquisition: Formal and Computational Models* (pp. 173-203). Cambridge: Cambridge University Press.

Ullman, M. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition, 92*, 231-270.

THE SUBTLE INTERPLAY BETWEEN LANGUAGE AND CATEGORY ACQUISITION AND HOW IT EXPLAINS THE UNIVERSALITY OF COLOUR CATEGORIES

TONY BELPAEME

School of Computing, Communication and Electronics, University of Plymouth
A318 Portland Square, Plymouth, PL4 8AA, United Kingdom
tony.belpaeme@plymouth.ac.uk

JORIS BLEYS

Artificial Intelligence Lab, Vrije Universiteit Brussel

When studying natural language, one inevitably needs to explain how linguistic signs and constructions map onto semantic concepts. Among concepts, perceptual categories form a special class in the sense that an insight into how they are acquired will have an important impact on theories of linguistic relativism. Linguistic relativism, also known as the Sapir-Whorf hypothesis, suggests an interplay between language and cognition, whereby language and concept acquisition influence each other. Among perceptual categories, colour categories are without doubt the best studied and still their origins and nature are controversial. Stakes are high; as Deacon writing on colour categories puts it "... this may at first appear to be a comparatively trivial example of some minor aspect of language, but the implications for other aspects of language evolution are truly staggering." (p. 120, 1997)

Berlin and Kay (1969) first reported the *universality of colour categories*: the fact that the foci of colour categories show a high degree of similarity across cultures. Their findings have recently been reconfirmed in a large-scale World Color Survey (Kay & Regier, 2003). Although the universal character of colour categories has been disputed (for a recent view see Roberson, 2005), many have accepted it and have put forward hypotheses about the processes underlying it. The most prominent hypothesis holds that colour categories are the result of the expression of innate constraints on colour perception and cognition. A second hypothesis puts forward that colour categories reflect the structure of human ecology. And a third hypothesis suggests that colour categories are culturally learned and puts somewhat less stress on their universal character. Combinations of these three views have received interest as well (for an overview see Steels & Belpaeme, 2005). However, most theories accounting for universalism are rhetorical and

therefore never quiet satisfactory. We on the contrary aim to explain the universal character using a computational simulation which draws on a psychological model of colour perception and a model of lexicon acquisition.

In our simulations we study populations of individuals which autonomously learn and adapt categories and linguistic labels for those categories. This enables the individuals to (a) distinguish between perceptual stimuli and (b) communicate with each other about perceptual stimuli. The essential ingredient of our model is an interaction between two agents, whereby one agent tries to linguistically convey the meaning of a colour to a second agent. In order for this to succeed both agents need to know the same colour terms, but more importantly, the colour categories of both agents need to be coordinated. Our simulations differ from previously presented work in that we now present data of a large-scale experiment.

As a yardstick to compare the model to, we use the data from the World Color Survey (Kay & Regier, 2003). The WCS contains data and an analysis of colour terms and their referents of 110 remote and non-industrialised societies. Our simulations contain 110 populations, which can be seen as 110 isolated societies. An analysis of the categories of the artificial societies reveals a structure showing the same typology as observed in the WCS. However, comparing two populations leaves the impression that colour categories are arbitrary, the universal structure only reveals itself when analysing the categories of a larger number of populations. This suggests that even if the genetic and ecological constraints are rather weak, on a macroscopic scale a certain structure will be observed: the universal structure of colour categories. We argue that the universality of colour categories can be explained through a linguistic acquisition process on top of genetic and ecological constraints. These constraints are formed by the nature of human colour perception and to a lesser extent by the chromatic environment.

References

Belpaeme, T., & Bleys, J. (2005). Explaining universal colour categories through a constrained acquisition process. *Adaptive Behavior, 13*(4), 293-310.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley, CA: University of California Press.

Deacon, T. W. (1997). *The symbolic species: the co-evolution of language and the brain.* New York: W.W. Norton.

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences, 100*(15), 9085-9089.

Roberson, D. (2005). Color categories are culturally diverse in cognition as well as in language. *Cross-Cultural Research, 39*, 56-71.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language. A case study for colour. *Behavioral and Brain Sciences, 24*(8), 469-529.

THE EVOLUTION OF MEANINGFUL COMBINATORIALITY

JILL BOWIE

Department of Applied Linguistics, University of Reading, Whiteknights, PO Box 218, Reading, RG6 6AA, England

This paper shows how the experimental study of artificial reduced language systems can shed light on evolutionary questions, when placed alongside evidence from other simpler language systems such as early child language.

The combination of meaningful elements into larger structures is recognized as fundamental to human language and to its use as an open-ended communicative resource. How this combinatoriality emerged is therefore a major issue in the field of language evolution. Two opposing kinds of account have been proposed: synthetic and holistic. In the synthetic account (e.g. Bickerton, 1998; Jackendoff, 2002), there emerged first of all single words and then simple combinations of words, with complex syntax developing later. In the holistic account (e.g. Arbib, 2005; Wray, 1998), there were first of all longer utterances which functioned holistically as complete messages, only gradually over time being broken down into words.

The holistic account appeals to many who wish to emphasize continuity between animal communication and human language. It has also been argued that a simple synthetic protolanguage would have lacked communicative effectiveness (e.g. Wray, 1998). However, the holistic account is problematic in a number of ways (Tallerman, 2006), while strong support for the synthetic account comes from the known range of simpler combinatorial systems (such as early child language, pidgin, home sign, and enculturated ape productions). These provide evidence that complexity is largely built up synthetically, and that symbols used singly or in simple combinations can have some degree of communicative effectiveness, although their interpretation is more heavily context-dependent than is the case for full grammatical language.

The present research explores the potential of an additional source of evidence: the experimental study of artificial reduced language systems as used by adults in communication tasks. The aim is to investigate the communicative effectiveness of simple synthetic systems, and also to examine their use in discourse — an issue rarely considered in evolutionary accounts.

The experiments required pairs of adults to use a restricted vocabulary of approximately fifty English words in a communicative task. Their productions

were recorded on videotape, while comprehension was determined by having each participant record in full English his or her understanding of the messages conveyed by the other. The communicative task was designed to produce a short discourse, relating to a specific context and including different kinds of speech act (e.g. statements offering information, questions soliciting information, requests for action, responses to requests for action). Various kinds of propositional content were also included (e.g. locational predication, attributional predication, event predication involving one or more participants).

Preliminary results indicate a considerable degree of communicative effectiveness for the simple synthetic system. Participants were able to understand the basic content of the messages at relatively consistent levels, despite variation in the way these messages were expressed (e.g. choice of word combinations, use of prosody and paralinguistic features). An important factor was the exploitation of the discourse context to fill out elements of meaning missing from the productions.

Acknowledgements

This research is funded by a doctoral award from the Arts & Humanities Research Council. I gratefully acknowledge their support and that of my supervisors, Professor Michael Garman and Professor Steven Mithen.

References

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, forthcoming.*

Bickerton, D. (1998). Catastrophic evolution: the case for a single step from protolanguage to full human language. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), *Approaches to the evolution of language: social and cognitive bases* (pp. 341-358). Cambridge: Cambridge University Press.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

Tallerman, M. (2006). Did our ancestors speak a holistic protolanguage? *Lingua, forthcoming.*

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language & Communication, 18*, 47-67.

THE ADAPTIVE ADVANTAGES OF KNOWLEDGE TRANSMISSION

JOANNA J. BRYSON

Artificial models of natural Intelligence (AmonI) Group, University of Bath
Bath, BA2 7AY, United Kingdom
jjb@cs.bath.ac.uk

Language is normally seen as a mechanism of communication. However, Dessalles (2000) has argued that language must have evolved as a form of costly signalling, because giving up knowledge is not an adaptive trait. Knowledge transmission disadvantages the transmitting agent, because the transmitter gives up knowledge to its neighbours / competitors. However, it has long been known that altruistic behaviour *can* evolve in conditions where a population is viscous — that is, when children tend to stay near their parents (Hamilton, 1964; Queller, 1994; Griffin et al., 2004). The genes that are being benefited are to some extent the same as those being disadvantaged; and so long as this benefit times the relatedness does not exceed the cost, altruism can be adaptive (Hamilton, 1964).

However, it has also been shown mathematically that in such cases, one's kin become one's competitors (e.g. Marshall and Rowe, 2003). In this case, the costs and benefits of altruism should equalise and altruism is selected neither before or against. This argument has been sustained in a bacteria-based live simulation, where the altruistic act is digesting food external to the cell for the benefit of all surrounding cells (Griffin et al., 2004). Griffin et al. show that in the case of low relatedness (for food competition) and local competition (for reproduction) altruism dies out, in cases of high relatedness and global competition altruism is selected for, and in the other two cases (including the viscous one — high relatedness / local competition) altruism is a neutral trait.

Čače and Bryson (2005) demonstrate in an agent-based ALife simulation that altruism can be selected for when the altruistic act is communicating about accessing food. We simulate two 'species', Talkers (altruists) and Silents (free riders). At each iteration of the simulation, a Talker tells any agent nearby one piece of its knowledge about how to eat complicated / special foods. In all other respects the two species are identical. Both profit by hearing knowledge equally, both have lifespans determined by either a fixed upper bound or starvation, both reproduce asexually at a rate dependent on their success in foraging, and always give birth to another individual of the same species. New knowledge enters the system during

an infant's first cycle, when five percent of agents discover new ways to eat.

There is a clear cost to transmitting this information, yet Talkers always outcompete Silents into extinction, provided only that there is anything to learn and that they have a large enough initial population to survive random fluctuations. In a classic Simpson's paradox, any Talker who knows about k types of food will have a lower average energy level (and thus a lower probability of reproduction) than a Silent who also knows about k types, however the average Talker has more energy than the average Silent. This is because Talkers tend to know more things, because in a viscous population they tend to live near more other Talkers.

Why does competition from 'kin' (both memetic and one-bit genetic) not neutralise the advantage of communication? I believe this is another instance of ABM finding a gaff in abstract mathematical modelling. The more information present in the environment, the higher its realized carrying capacity. This effect is not large in our simulation, but it is enough to tip the equilibrium.

The salience of this work to the evolution of language should be evident. We have shown that any transmission of knowledge (about food at least) is adaptive. Further, our simulations show that the higher the rate of transmission, the faster the Talkers outcompete the Silents. Thus assuming hominids communicate knowledge about food (Steele, 2004), we now know incremental increases in communicative efficacy *could* be sustained by selective pressure.

References

Dessalles, J.-L. (2000). Language and hominid politics. In Knight, C., Studdert-Kennedy, M., and Hurford, J., editors, *The Evolutionary Emergence of Language*, pages 62–79. Cambridge University Press, Cambridge, UK.

Griffin, A. S., West, S. A., and Buckling, A. (2004). Cooperation and competition in pathogenic bacteria. *Nature*, 430:1024–1027.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, 7:1–52.

Marshall, J. A. R. and Rowe, J. E. (2003). Viscous populations and their support for reciprocal cooperation. *Artificial Life*, 9(3):327–334.

Queller, D. C. (1994). Genetic relatedness in viscous populations. *Evolutionary Ecology*, 8:70–73.

Steele, J. (2004). What can archaeology contribute to solving the puzzle of language evolution? plenary talk at *The Evolution of Language*.

Čače, I. and Bryson, J. J. (2005). Why information can be free. In Cangelosi, A. and Nehaniv, C. L., editors, 2^{nd} *International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05)*, pages 17–22, Hatfield, UK.

DETERMINING SIGNALER INTENTIONS; USE OF MULTIPLE GESTURES IN CAPTIVE BORNEAN ORANGUTANS (*PONGO PYGMAEUS*)

ERICA CARTMILL AND RICHARD BYRNE

School of Psychology, University of St Andrews, St Andrews, Fife, KY16 9JP, UK

Many researchers have studied primate call systems for clues to the antecedents of human language. Several species of primates use calls that have functionally referential meanings (Zuberbühler, 2003). It has yet to be demonstrated, however, that the meanings encoded in these calls are intended by the senders. It is difficult to see how a vocal system of unintentional signals, rigid in structure and situation-specific, could have transitioned into a flexible language-like system with recombinative power.

Recent studies of various primate species have shown that non-human primates use gesture as a flexible medium of communication, altering the nature of the signals in different social contexts and to achieve different goals (Liebal et al., 2004a; Liebal et al., 2004b; Pika et al., 2003; Tanner & Byrne, 1999; Maestripieri, 1996). The study of natural gestural communication in non-human primates provides us with a unique opportunity to address questions about primates' social understanding and personal expectations.

Gesture is not constrained by the same "bounded syllables" and constrictive physiology that limit the vocal communication systems of most non-human primates. Gestural studies, however, are hindered by the difficulty of determining when an animal is signaling. Unlike most vocalizations, gestures lack clear boundaries and overlap with other movements of daily living, so it is hard to tell when the function of a movement is mainly communicative. Nonhuman gestures can be identified as such by recipients' responses, however this approach fails to capture the complexities of the use of multiple gestures in a single signaling event - and, more importantly, includes no measure of the signaler's intentions.

To address the aforementioned problems of multiple-gesture combinations and signaler intentions, we studied the gestural bouts of 9 captive Bornean orangutans housed at Apenheul Primate Park, NDL. Gestural strings were defined as temporally-linked movement sequences made by signalers to conspecific recipients that failed to respond in any way. Our study focused on what alternative behaviors signalers exhibited in cases where initial communicative attempts failed. When a recipient does not respond, gestures can

be recognized by subsequent repetition or patterned modifications of the signal. Such modifications provide information about the signaler's goal and awareness of the recipient's attentive state.

In our sample, string lengths ranged from 2-9 gestural elements. Multi-gesture strings were most often produced by juveniles to initiate play and by adults in food-sharing or displacement situations. The probability of a signaler performing another gesture in a string increased from the 1st gesture until the 4th gesture. Beyond that, the probability of giving up and ending the sequence increased. The time between gestures decreased as the number of elements in the sequence increased. When recipients did not respond to gesturing, signalers often touched recipients and/or moved closer to them or into their visual fields. These findings are noteworthy because they show both persistence and goal-directed behavior on the part of the signaling orangutan. Persistence and goal-directed behavior within the communicative system, coupled with the non-formulaic nature of the gestural sequences, demonstrate that orangutans may have specific intended results for some of their gestures.

References

Liebal, K., Pika, S., & Tomasello, M. (2004a). Social communication in siamangs (*Symphalangus syndactylus*): use of gestures and facial expressions. *Primates, 45*, 41-57.

Liebal, K., Call, J., & Tomasello M. (2004b). Use of gesture sequences in chimpanzees. *American Journal of Primatology, 64*, 377-396.

Maestripieri, D. (1996). Gestural communication and its cognitive implications in pigtail macaques (*Macaca nemestrina*). *Behaviour, 133* (13-14), 997-1022.

Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (*Gorilla gorilla*): gestural repertoire, learning, and use. *American Journal of Primatology 60*, 95-111.

Tanner, J. and Byrne, R. (1999) Spontaneous gestural communication in captive lowland gorillas. In S. Parker, R. Mitchell & H. Miles (Eds.) *The mentalities of gorillas and orang-utans in comparative perspective*, (pp.211-239). Cambridge University Press.

Zuberbühler, K. (2003). Referential signalling in non-human primates: cognitive precursors and limitations for the evolution of language. *Advances in the Study of Behavior,* 33, 265-307.

NUCLEAR SCHIZOPHRENIC SYMPTOMS AS THE KEY TO THE ORIGINS OF LANGUAGE

TIMOTHY J CROW

Prince of Wales International Centre for SANE Research into Schizophrenia, Warneford Hospital, Warneford Lane, Oxford, OX3 7JD, United Kingdom

From at least de Saussure onwards it has appeared that some sort of compartmentation (eg Signifier versus the signified, thought versus speech, syntax versus the lexicon) is what is characteristic of human language. Such dichotomies might correspond to distinctions between neural systems, but how could new neural boundaries have arisen, relatively rapidly, in the course of hominid evolution? Here it is argued that Broca's (1877) concept that "Man is, of all the animals, the one whose brain in the normal state is the most asymmetrical...It is this that distinguishes us most clearly from the animals" is of central importance.

Asymmetry in the hominid lineage took the form not of a simple left-right distinction but of a 'torque' or bias from right frontal to left occipital.. This innovation, assumed to have depended upon a single improbable event, had the effect that human association cortex is constituted as four separate chambers – right and left anterior, and left and right posterior – by contrast with the two chambers – anterior motor and posterior sensory – of association cortex of other primates. The torque has the additional consequence that the difference between the sides is in an opposite direction in the anterior (motor) and posterior (sensory) domains. Thus according to the *"quadri-cameral"* concept the perceptual and productive elements are in parallel with each other but orientated *in opposite directions*.

Ignoring the interface with the external world, there are three and only three interfaces: 1) from the perceptual to conceptual (from primary phonological engrams to 'meanings'), 2) from concepts or meanings to plans or intentions that are within the individual's control and 3) the transition from thought to speech that occurs from right to left dorso-lateral prefrontal cortex.

Individuals suffering from what are described as 'schizophrenic' symptoms have two core (nuclear) subjective experiences – 1) They experience thoughts as outside their own control. Thus in the phenomena of *thought insertion* the individual experiences thoughts, which he identifies as not his own,

as inserted into his mind, and in the case of *thought withdrawal*, he experiences his own thoughts as removed from his mind by an outside force. 2) They experience neural activity which is manifestly self-generated (thoughts or plans for action) as spoken aloud by persons or other agents in the external world. These symptoms can be conceived as leaks between compartments.

Conclusions

The phenomena of psychosis exemplify the role of the self in language. Karl Buehler (1934) regarded language as constructed around a deictic origin in the first person, the present moment in time and the location of the speaker. Nuclear symptoms reflect a breakdown of the barrier between what is self- and what is other generated in language. These symptoms, interpreted in terms of the cerebral torque, cast light on the functions of the compartments. They tell us for example that thought is real and distinct from speech production, and located in right dorsolateral prefrontal cortex. They tell us that speech production and perception are separate but parallel processes with opposite polarity. They indicate that the engrams in the left hemisphere in Broca's area must be distinct from those in Wernicke's area, although closely related to them.

References

Buehler, K. (1934) *Sprachtheorie.* Translated (1990) by D.W.Goodwin as J Benjamins: Amsterdam.

Crow, T. J. (1998). Nuclear schizophrenic symptoms as a window on the relationship between thought and speech. *British Journal of Psychiatry 173,* 303-309.

Crow, T. J. (2004a). Auditory hallucinations as primary disorders of syntax: An evolutionary theory of the origins of language. *Cognitive Neuropsychiatry, 9,* 125-145.

Crow, T.J. (2004b) Cerebral asymmetry and the lateralisation of language: core deficits in schizophrenia as pointers to the gene. *Current Opinion in Psychiatry, 17,* 97-106.

Crow, T. J. (2005). Who forgot Paul Broca? The origins of language as test case for speciation theory. *Journal of Linguistics, 41,* 133-156

Mitchell, R. L. C. & Crow, T. J. (2005) Right hemisphere language functions and schizophrenia: the forgotten hemisphere? *Brain, 128,* 963-978.

ARTICULATOR CONSTRAINTS AND THE DESCENDED LARYNX

BART DE BOER
Artificial Intelligence, Rijksuniversiteit Groningen,
Grote Kruisstraat 2/1,
9712 TS Groningen , the Netherlands

1. Introduction

The descent of the larynx is a hotly debated topic in the evolution of language. Some argue that it can be explained as an adaptation to producing more and more distinctive speech sounds while others argue that a descended larynx is not necessary for distinctive speech, and that it has descended for other reasons.

Recently, computer modelers have joined the debate by building models of the vocal tract and investigating what sounds can be produced (Boë, Heim, Honda, & Maeda, 2002; Carré, Lindblom, & MacNeilage, 1995). However, the different groups draw diametrically opposite conclusions, even though they use very similar methods. While Carré *et al.* find that a pharyngeal cavity is essential for producing distinctive speech sounds and that therefore the descended larynx is adaptive, Boë *et al.* find that their model of the Neanderthal vocal tract (with a smaller pharyngeal cavity) can produce as distinctive vowel sounds as a modern human vocal tract. They therefore conclude that a descended larynx is not adaptive for speech.

The two studies find the same thing, but interpret it differently. They both find that two cavities of controllable size are essential for producing the range of sounds in modern speech. Carré *et al.* see this as proof that a pharyngeal cavity and thus a descended larynx are necessary, while Boë *et al.* claim that a back cavity can also be made without a descended larynx. Both conclusions are debatable, however, as neither model has realistic constraints on what configurations can be made by movement of the tongue, jaw and lips. In Carré *et al.*'s model, motion is unconstrained, while in Boë *et al.*'s model it is constrained by deformations that have been statistically derived from observed human vocal tract motion. In order to investigate the difference in acoustic range between human-like and ape-like vocal tracts, one must use models that have realistic constraints on articulator motion.

2. The Model

We propose to use an articulatory synthesizer that is based on the actual geometry of the vocal tract and on physical control of the articulators. The Mermelstein (Mermelstein, 1973) model fulfills these criteria. It will be used for investigating the potential vowel space of modern humans. It is straightforward to modify this model so that it conforms more to an ape-like vocal tract with a higher larynx (figure 1). It will then be investigated how this influences the

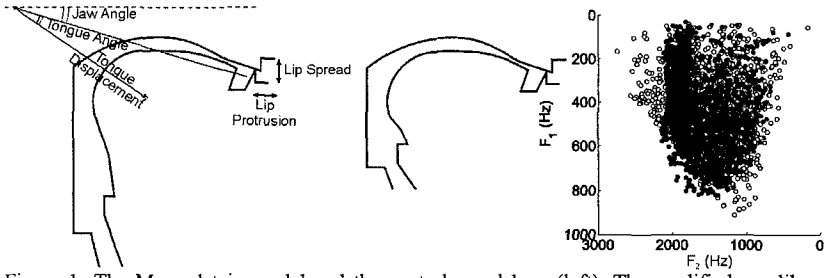


Figure 1: The Mermelstein model and the controls used here (left). The modified ape-like model (middle). The possible vowels (right). Open circles indicate the human tract, filled circles the ape-like tract.

range of sounds that can be produced with the same constraints on movement of the articulators.

3. Preliminary Results

In figure 1 it is shown which vowel positions can be reached by the two models (assuming equal length of the vocal tracts). It is clear that the human-like vocal tract is able to produce more distinctive vowels than the ape-like tract. These results are preliminary, however. The articulatory model needs to be refined, using more realistic data about ape- and Neaderthal vocal tracts, it must be made continuously variable and the results must be analyzed more carefully.

The results do seem to indicate, however that a lowered larynx allows for more distinctive vowel sounds, because it allows more different configurations of the front and back cavity, given constraints of articulator movement. A tract with a higher larynx is more articulatorily constrained. It can therefore tentatively be concluded that a descended larynx has adaptive value for speech.

References

Boë, L.-J., Heim, J.-L., Honda, K., & Maeda, S. (2002). The potential neandertal vowel space was as large as that of modern humans. *Journal of Phonetics, 30*(3), 465–484.

Carré, R., Lindblom, B., & MacNeilage, P. (1995). Rôle de l'acoustique dans l'évolution du conduit vocal humain. *Comptes Rendus de l'Académie des Sciences, Paris, 320*(série IIb), 471–476.

Mermelstein, P. (1973). Articulatory model for the study of speech production. *The Journal of the Acoustical Society of America, 53*(4), 1070–1082.

EVOLUTIONARY SUPPORT FOR A PROCEDURAL SEMANTICS FOR GENERALISED QUANTIFIERS

SAMSON TIKITU DE JAGER

Institute for Logic Language and Computation, Universiteit van Amsterdam, Nieuwe Doelenstraat 15,
Amsterdam, 1012 CP, The Netherlands
S.T.deJager@uva.nl

1. Setting the scene

An extensional semantics gives the denotation of expressions as sets of objects, relations between objects, relations between sets of objects, and so on. A predicate is true of an object iff that object appears in the set that is the denotation of the predicate. In a procedural semantics, on the other hand, a predicate denotes a procedure which when given an object determines whether the predicate applies to the object (Benthem & Eijck, 1982).

Extensional semantics has given a coherent and compositional account of the meanings of many determiners ("all", "some", "few") as relations between sets of objects. Semantically speaking, the theory of generalised quantifiers (see for instance Keenan and Westerståhl (1997)) gives a very tidy account meanings of Det+NP expressions (such as "all men", "some tidy bedrooms") as well as many others that occur in the same syntactic environment (some syntactically quite complex, for instance "half the schoolchildren, all the teachers except Bob, and the neighbour's cat"). However the extensional semantics for generalised quantifiers, while capable of describing most of the determiners we see in natural language, also allows the possibility of a vast number of determiners that are not attested. (A determiner is analysed as a relation between sets of objects, so if only three objects exist in the domain then there are $2^{2^4} = 65536$ possible determiner denotations.) Many properties are known that restrict this space (e.g., permutation-invariance, which makes the truth value of a determiner dependant only on the sizes of the sets involved, not the identities of their elements.).

Other properties have been identified as "trends" or "weak universals" (Keenan & Westerståhl, 1997); the vast majority of determiners expressed as simple lexical items are upwards monotonic ("All Englishmen are dirty scoundrels" implies "All Englishmen are scoundrels") a small number are downwards monotonic ("Few Englishmen are knaves" implies "Few Englishmen are cowardly knaves") and very few indeed are not monotonic at all.

2. Evolutionary contribution

Neither extensional nor procedural semantics on their own can explain these trends. However using evolutionary reasoning we can approximate these results using a particular model of procedural semantics based on deterministic finite automata (DFAs; see Benthem, 1987). Monotonicity of quantifiers corresponds to a natural simplicity bias on automata (the denotations of quantifiers). An iterated learning model incorporating this learning bias[a] then explains both a preference for monotone quantifiers and the presence of non-monotone ones (since certain non-monotone quantifiers *can* be learned, given sufficient examples).

Furthermore, an extensional semantics is totally unable to account for the bias towards *upward* monotonicity, while the same learning bias within a procedural semantics can do so. Indeed, the simplicity bias also predicts some of the non-monotonic and downward monotonic quantifiers that are indeed attested ("no", some small exact numbers).[b]

Finally, the iterated learning perspective provides an explanation for a gap between the semantics and the pragmatics of "some" (taken pragmatically to mean "some but not all"). The question is not how such a pragmatic meaning arises, but why it is not fossilised into semantic meaning by the learning process. In this model the upward monotonic DFA corresponding to the semantics is easier to learn than the DFA representing the pragmatic meaning; acquisition of one or the other meaning depends both on how frequently infelicitous examples are provided ("some" used when "all" would also be appropriate) and on the number of examples given (the 'learning bottleneck' of the iterated learning paradigm).

References

Benthem, J. van. (1987). Towards a computational semantics. In P. Gärdenfors (Ed.), *Generalized quantifiers: Linguistic and logical approaches* (pp. 31–71). Dordrecht: Reidel.

Benthem, J. van, & Eijck, J. van. (1982). The dynamics of interpretation. *Journal of Semantics*, *1*, 3–20.

Grünwald, P. (2005). A tutorial introduction to the minimum description length principle. In P. Grünwald, I. J. Myung, & M. Pitt (Eds.), *Advances in minimum description length: Theory and applications* (pp. 3–80). MIT Press.

Keenan, E. L., & Westerståhl, D. (1997). Generalized quantifiers in linguistics and logic. In J. F. A. K. van Benthem & G. B. A. ter Muelen (Eds.), *The handbook of logic and language* (pp. 837–893). MIT Press.

[a]Formally speaking, I use the Minimum Description Length principle (see for example Grünwald, 2005) to drive a DFA learning algorithm using greedy state-merging.

[b]The precise prediction depends on parameter settings of the model, which is unfortunately too crude for a match to real usage to have much independant meaning.

THE EVOLUTION OF SPOKEN LANGUAGE: A COMPARATIVE APPROACH

W. TECUMSEH FITCH
School of Psychology,
University of St Andrews
St Andrews, Fife KY16 9AJ (UK)
wtsf@st-andrews.ac.uk

The study of the evolution is entering an exciting new period of interdisciplinary collaboration. Biologists, linguists, psychologists and many others are combining theoretical perspectives with an ever-increasing influx of data in exciting and innovative ways. Old barriers to interdisciplinary communication are being broken down, and diverse sources of data are being used to place increasingly exacting constraints on models of language evolution. One important new source of data results from applying the comparative method to living organisms. Comparative data from many levels, including molecular genetics, development, neuroscience, ecology, and behavioural studies of animal cognition and communication, are all playing an increasingly important role in biolinguistics. In this talk I will illustrate the power of the comparative approach to language evolution with a detailed discussion of the evolution of speech. New comparative data on mammalian vocal production show that many mammals lower the larynx and tongue during vocalization, dynamically attaining a vocal tract morphology comparable to that of adult humans. Various other mammal species have recently been discovered to have permanently descended larynges like that of humans, but none of these species produce complex vocalizations comparable to speech. Therefore, the importance of the rearrangement of human vocal anatomy to the evolution of speech appears to have been overemphasized in the past. In particular, fossil cues to vocal anatomy cannot conclusively demonstrate the presence or absence of speech in extinct hominids. In contrast, the comparative data on vocal imitation (vocal learning of complex signals) shows that the evolution of novel neural mechanisms for vocal control represented a crucial hurdle in the evolution of speech. New molecular data concerning the genetic basis of vocal control offer tantalizing insights into the evolution of this capacity. I end with a briefer discussion of the evolution of language per se, describing methods for examining some of the neural mechanisms underlying syntax, and concluding with a discussion of the selective forces that could have driven our

species' unusual propensity to cooperatively share meaning. For all of these topics, a rich but often neglected store of comparative data is available. I conclude that there is a rich future for integrating comparative studies into the field of language evolution.

ALLEE EFFECT ON LANGUAGE EVOLUTION

JOSÉ F. FONTANARI

Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, São Carlos, SP 13560-970, Brazil

LEONID I. PERLOVSKY

Air Force Research Laboratory, 80 Scott Rd., Hanscom Air Force Base, MA 01731, USA

The case for the study of the evolution of communication[a] within a multi-agent framework was probably best made by Ferdinand de Saussure in a famous statement made in his lectures at the University of Geneva (1906-1911) "language is not complete in any speaker; it exists only within a collectivity... only by virtue of a sort of contract signed by members of a community" (Saussure, 1966). More than one decade ago, seminal computer simulations were carried out to demonstrate that natural selection (MacLennan, 1991) or, alternatively, learning (Hurford, 1989) could lead to the emergence of ideal communication codes (i.e., one-to-one correspondences between objects or meanings and signals) in a population of interacting agents. Typically, the behavior pattern of the agents was modeled by (probabilistic) finite state machines. The work by Hurford, in particular, set the basis of the celebrated Iterated Learning Model (ILM) for the cultural evolution of language (Smith et al, 2003). In those studies, language is viewed as a mapping between meanings and signals. The abovementioned ideal codes that emerge from the agents interactions are examples of non-compositional or holistic communication, in which a signal stands for the meaning as whole. In contrast, a compositional language is a mapping that preserves neighborhood relationships – similar signals are mapped into similar meanings. The emergence of compositional languages in the ILM framework beginning from holistic ones in the presence of bottlenecks on cultural transmission was considered a major breakthrough in the computational language evolution field. Our aim in this contribution is twofold. First, we show that in practice, though contrasting at first sight, the cultural evolution approach in which the offspring *learn* their language from their

[a] Here we take the more conservative viewpoint that language evolved from animal communication rather than from animal cognition.

parents (or from other members of the community) differs very little from the genetic approach, in which the offspring *inherit* their communication ability from their parents. For instance, errors in the learning stage or the inventiveness associated to bottleneck transmission have the same effect of mutations in the genetic approach. Second, we show, through extensive simulations of language evolutionary games, that once an ideal communication code, say a holistic one, is established in the population, i.e., all individuals use the same code, it is impossible for a mutant to invade, even if the mutant uses a better code, say, a compositional one. This is essentially the Allee effect (Allee, 1931) of population dynamics which, for instance, prevents a population of asexual individuals of being invaded by a sexual mutant. The ILM circumvents this difficulty by assuming that the population is composed of two individuals only, the teacher and the pupil, and that the latter always replaces the former. However, according to Saussure (see quotation above), this is not an acceptable framework for language. The solution of the conundrum - how a compositional code can evolve in a population of agents that communicate through a holistic code - may give a clue on the interplay between cultural and genetic mechanisms in the evolution of language as well as support the viewpoint that language can in principle emerge from animal communication.

References

Allee, W. C. (1931). *Animal Aggregations. A Study in General Sociology.* Chicago: University of Chicago Press.

de Saussure, F. (1966). *Course in General Linguistics.* Translated by Wade Baskin. New York: McGraw-Hill Book Company.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua, 77,* 187-222.

MacLennan, B. J. (1991). Synthetic ethology: an approach to the study of communication. In Artificial Life II, SFI Studies in the Sciences of Complexity, vol. X (pp. 631-658). Redwood City: Addison-Wesley.

Smith, K., Kirby, S., Brighton, H. (2003). Iterated Learning: a framework for the emergence of language. *Artificial Life,* 9, 371-386.

RAPIDITY OF FADING AND THE EMERGENCE OF DUALITY OF PATTERNING

BRUNO GALANTUCCI, THEO RHODES & CHRISTIAN KROOS
Haskins Laboratories, 300 George St.,
New Haven, CT - 06511, USA

Hockett (1960) identified *duality of patterning*, that is, the fact that few meaningless units generate a large number of meaningful elements, as one of the critical design-features of human languages. Another design-feature identified by Hockett (1960) is *rapidity of fading*, that is, the fact that linguistic messages are transmitted in a medium over which signals quickly fade. We propose a link between the two design-features. In particular, we hypothesize that the more rapidly signals fade in a medium, the more likely it is that human communication systems emerging over that medium develop duality of patterning. To test this hypothesis, we ran an experiment using the method developed by one of us (Galantucci, 2005) for studying the emergence of human communication systems in the laboratory.

Pairs of participants played a videogame with interconnected computers. The videogame required players to communicate, but players played from different locations and could not see or hear one another. Instead, they could reach one another by using a magnetic stylus on a small digitizing pad. The resultant tracings were relayed to the computer screens of both players. However, players controlled only the horizontal component of the tracings on the screen via the horizontal component of their stylus' movements. The vertical component of the stylus' movement did not affect the tracings. Rather, the tracings either (a) moved with a constant downward drift (slow fading signal, henceforth SF condition) or (b) had no vertical movement, appearing as a dot moving horizontally at a fixed height on the screen (fast fading signal, henceforth FF condition). In both conditions, the use of standard graphic forms (e.g., letters) was practically impossible. This constraint forces players to develop communication systems from scratch (Galantucci, 2005).

Ten pairs of participants participated in the experiment: five pairs in the SF condition and five pairs in the FF condition. In both conditions, pairs played a videogame in which each player controlled one agent. The game was organized in rounds. In each round the agents started in two different rooms at random in a four room virtual environment (2x2 grid) and had to find one another without

making more than one room change each. The scoring mechanism of the game was such that, in the absence of effective communication, the score would stably fluctuate around its initial value. If the pair reached a score that indicated successful communication, players were invited to play the game at a new stage: The game environment was enlarged (6 rooms, 2x3 grid) and an additional room change per round was allowed. For successful pairs, the size of the environment (and the number of room changes allowed) could grow three more times until the environment, at the fifth and final stage, was composed of 16 rooms (4x4 grid). Pairs were invited to play for three sessions of 2 hours each and were told that their goal in the game was to achieve as high a score as possible. For the entire duration of the game, the movements of the agents and the activity on the pad were recorded at approximately 30 Hz. On termination of the third session, participants were asked to describe in detail the communication systems they developed for playing.

The game performance of the pairs did not differ significantly in the two conditions. The mean maximum stage reached was 3.2±1.5 in the SF condition and 3±2.1 in the FF condition ($F<1$). To measure the degree of duality of patterning of the pairs' sign systems, first we determined the number of signs (*S*) used by the pairs to identify the game's rooms. Second, the total number of unique separable units (*U*) in the sign system was determined. (Separable units in a sign were defined as portions of stylus' activity made with uninterrupted contact with the pad.) Finally, an *index of combinatoriality* (*C*) was computed as $C = 1-(U/S)$. *C* equals 0 for systems with no duality of patterning and approaches 1 for systems with maximal duality of patterning. The mean *C* was .34±.34 for the pairs in the SF condition and .79±.14 for the pairs in the FF condition. The difference is statistically significant, $F(1,8) = 7.1$, $p = .03$, $\eta 2 = .47$.

The results will be discussed in the context of recent hypotheses about the origins of duality patterning (Nowak, Krakauer, & Dress, 1999; Studdert-Kennedy & Goldstein, 2003).

References

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science, 29*, 737-767.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*(3), 88-96.

Nowak, M. A., Krakauer, D. C., & Dress, A. (1999). An error limit for the evolution of language. *Proceedings of the Royal Society of London Series B, 266*, 2131-2136.

Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution*. New York: Oxford University Press.

RECONSIDERING KIRBY'S COMPOSITIONALITY MODEL TOWARD MODELLING GRAMMATICALISATION

TAKASHI HASHIMOTO & MASAYA NAKATSUKA

School of Knowledge Science,
Japan Advanced Institute of Science and Technology (JAIST)
1-1, Nomi, Ishikawa, Japan, 923-1292
{hash,m-naka}@jaist.ac.jp

Grammaticalisation is a potent candidate to structuralise and complexify human languages in the evolution of language. It is a phenomenon of language change, in which content words such as nouns and verbs change into functional words such as auxiliaries and prepositions. New functional categories, tense, mood, and so forth, can emerge in a language structure through grammaticalisation, then structure and lexicon of a language can become complex and fruitful. It is important to understand the process of and the cognitive ability for grammaticalisation in the context of the origin and the evolution of language. We discuss constructing a computational model for grammaticalisation to achieve this end.

It is assumed that reanalysis and analogy are underlying mechanisms of grammaticalisation (Hopper & Traugott, 2003). Reanalysis is structural change without observable change in forms. This occurs when a hearer understands a form to have a structure differently from that of a speaker. Analogy is to apply a grammatical rule to forms in which the rule was not applied formerly. These mechanisms postulate a coginitive ability to find analogy among situations and among forms. We call the former "linguistic analogy" and the latter "cognitive analogy" .

We thoroughly analysed Kirby's compositionality model (Kirby, 2002), especially the relationship between learning mechanisms in the model and the underlying mechanisms for grammaticalisation from the cognitive viewpoint in order to develop a model of grammaticalisation based on reanalysis and analogy. In this model, a language learner acquiring his own grammar performs three operations to generalise his grammar: chunk, merge and replace (the third one is not named in (Kirby, 2002)). Cognitive analogy is premised in chunking and merging. Reanalysis is realised partly in chunking, since a learner can analyse utterances in different way from a speaker's by chunking operation. The important feature of linguistic analogy is expressed in merging and replacing, for a learner extensively applies a grammatical rule, which was used for only an instance, to all members in a category to which the instance belongs. It was also recognised that these two

operations were so strong that one instance triggers complete integration of different categories. Consequently, reanalysis and analogy are thought of as being modelled in part in Kirby's model.

Accordingly, it is expected that a phenomenon superficially comparable to grammaticalisation is observed in simulations of the model. The meanings in the model, however, consist of verbs and nouns, no function meaning. Thus, we investigated meaning change in which syntactic category of a word varies with time. Grammaticalisation is a subset of this type of meaning change, since syntactic category of a word changes over time such as from verb to auxiliary and from noun to preposition. In search of such meaning change, we slightly modified the model in order not to converge but to keep changing.

We actually observed phenomena in which a form for a noun was to be used commonly for various verbs in simulations of Kirby's model. They occur through the following process: 1) There are two forms for one noun meaning. 2) Both two happen to appear in an utterance of a speaker. 3) A learner analyses one of them as representing the noun and the other as a part of a form for another meaning. 4) The latter form is to acquire another meaning later. Our scrutiny revealed that a meaning change in which syntactic category of a word was transformed was caused by the deviation of intention between speaker and learner, and the differentiation of word meaning brought by the existence of synonyms. We also found that the replacing operation played an important role in this change process.

We introduce function meanings as an additional argument in predicate logical expressions, which are employed as meaning representations, since the Kirby's original model was not able to express a functional meaning. In this study, we used tense, that is, past, present and future. The change of word meaning over content and function categories, such as from nouns or verbs to tense, was also observed. Accordingly, we confirmed that a slight modification of the Kirby's compositionality model can work as a basic model of grammaticalisation.

Further, in order to equip a meaning space with particular structure, two modifications was brought in. One is to change the criterion to apply the chunking operation. The other is to change the appearance frequency of meanings. Both modifications are concerned with a verb "go" and a tense "future". This premises that the agent has a cognitive disposition to consider, or the world has a physical structure, that actions of going often cause something in future. The effect of these modifications on the phenomena of grammaticalisation will be discussed.

References

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization.* Cambridge: Cambridge University Press.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition* (pp. 173–203). Cambridge: Cambridge University Press.

THE INTERRELATED EVOLUTIONS OF COLOUR VISION, COLOUR AND COLOUR TERMS

DAVID JC HAWKEY

Language Evolution and Computation Research Unit, University of Edinburgh, George Square,
Edinburgh, EH8 9LL, Scotland
dave@ling.ed.ac.uk

The World Colour Survey (WCS) identifies cross linguistic commonalities in colour-terms (Kay & Regier, 2003). Steels and Belpaeme (2005) present computer simulations designed to test the abilities of competing theories to produce a consistent set of colour terms, both within and across communities. The three theories tested are Nativism (colour categories are innate) Empiricism (categories are individually learned from the colours in the environment) and Culturalism (categories are coordinated by language). All three models presume a common model of colour terms in which a colour term is a label for a mental category. This view is problematic in light of the facts that newborn infants react to light in categorical manner (Bornstein, 1997), but take a long time to learn to use colour-terms for colour, and an even longer time to learn to use them for appropriately (Sandhofer & Smith, 2001). Whatever mechanism ensures infants' innate responses to colours appears not to form the basis of their acquisition and use of colour terms.

The notion that colour terms are names for mental colour categories derives from a commonly held view that language is essentially a vehicle for encoding and decoding mental entities. This view is problematic and neither necessary nor well founded (Wittgenstein, 1958; Harris, 1981). Integrationist linguistics is an alternative to this "language myth" in which "signs are not *prerequisites* of communication, but its *products*" (Harris, 2005, p. 110).

In this paper I present a reanalysis of the WCS data which avoids the problems associated with standard "colour spaces" highlighted by Saunders and van Brackel (1997). Cross linguistic universal properties of colour terms are identified and related to data on infant innate colour responses. The universal properties of colour terms are examined within an integrational model. Colour terms emerge in a language through human interactions to which colour is relevant. In order for colour terms to emerge from such interactions (both through interpretation and creative use of language) there must exist correlations between colours and what is being communicated. Innate responses to colour are an integral part of the mechanism

which drives languages to divide colours along the same fault-lines, though they do not provide "mental representations" underpinning colour terms.

Several evolutionary mechanisms are identified which conspire to make colour a semi-reliable signal in the environment: the evolution of innate responses to naturally occurring colour signals; (e.g., discriminating objects from a background of leaves); the evolution of colour signals on organisms in response to the evolutionary pressures set up by other animals' colour responses (e.g., the evolution of a colour–signal of ripeness of some fruits, Regan et al., 2001); and niche selection by animals with hardwired colour responses. These mechanisms tend to correlate colours in the human environment with properties of coloured objects: colour tends to becomes a signal, the "meaning" of which is correlated with innate responses to colour. These colour signals form the basis of the correlations between human communicational acts and colours from which colour-terms can arise.

Brill (1997) suggests that colour science underpins the technologies that colour the modern world and so shapes modern colour responses. The model presented here parallels this idea with the notion that innate colour response tunes (and is tuned to) colouring of the human-relevant environment, and this relationship underpins the universal tendencies of colour terms. This model is simultaneously Nativist, Empiricist and Culturalist, though with a non-mentalist flavour.

References

Bornstein, M. H. (1997). Selective vision. *Behavioral and Brain Sciences, 20,* 180-181.

Brill, M. H. (1997). When science fails, can technology enforce color categories? *Behavioral and Brain Sciences, 20,* 182-183.

Harris, R. (1981). *The language myth.* London: Duckworth.

Harris, R. (2005). *The semantics of science.* London: Continuum.

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *PNAS, 100*(15), 9085-9089.

Regan, B. C., Julliot, C., Simmen, B., Viénot, F., Charles-Dominique, P., & Mollon, J. D. (2001). Fruits, foliage and the evolution of primate colour vision. *Philosophical Transactions of the Royal Society of London B, 356*(1407), 229-283.

Sandhofer, C. M., & Smith, L. B. (2001). Why children learn color and size words so differently: Evidence from adults' learning of artificial terms. *Journal of Experimental Psychology: General, 130*(4), 600-620.

Saunders, B. A. C., & van Brackel, J. (1997). Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences, 20,* 167-228.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences, 28,* 469-529.

Wittgenstein, L. (1958). *The blue and the brown books.* Oxford: Blackwells.

A LITTLE BIT MORE, A LOT BETTER - LANGUAGE EMERGENCE FROM QUANTITATIVE TO QUALITATIVE CHANGE

JINYUN KE

English Language Institute, University of Michigan, 401 E. Liberty St. Ann Arbor, MI 48104, USA

CHRISTOPHE COUPÉ

Laboratoire Dynamique du Langage, ISH, 14 Ave Berthelot 69363 Lyon, Cedex 07 France

TAO GONG

Department of Electronic Engineering, Chinese University of Hong Kong, Shatin, NT Hong Kong, CHINA

The draft of chimpanzee genome was published recently (Nature September 2005). It has been known that chimpanzees share more than 98% of our DNA and almost all of our genes. In addition to the striking genetic closeness, the studies on chimpanzees in both laboratories and natural habitats have revealed that they share with us many cognitive abilities (Tomasello & Call 1997; Hauser 2005), and exhibit complex social behaviors (de Waal 2005) and rich cultural traditions which are transmitted through social learning (Whiten 2005). In particular, chimpanzees have demonstrated cognitive abilities which are considered crucial for learning and using language, including manipulation of symbols, understanding of abstract concepts, intention reading and attention sharing, the ability of imitation, and so on.

While chimpanzees share a strikingly high degree of similarity with humans, the question about language origin become more intriguing: if chimpanzees are so close to humans in cognitive abilities and social behaviors, why can't they invent a complex communication system with compositionality, hierarchy, and recursion similar to humans?

Elman (2005) points out that "language sits at the crossroads of a number of small phenotypic changes in our species that interact uniquely to yield language as the outcome" (p114). It is these small phenotypic differences between human and chimpanzees that result in a communication means of a totally different nature. The study of complex nonlinear systems has shown abundant examples of such small quantitative differences leading to phase transitions, i.e. qualitative

differences, in the system dynamics. One classic example is the bifurcation observed in the logistic map (May 1976), in which the system changes from a stable end state to an oscillation end state, when the parameter changes from 2.999 to 3.001.

We use a computer agent-based model to show how the small changes of a few parameters of cognitive abilities would result in such a phase transition in the outcome of the communication system. The model simulates a group of agents interacting with each other with increasing communication ability. The agents possess a set of pre-linguistic abilities which have been shown to be shared by chimpanzees and humans, i.e. they have simple semantic distinctions between entity and action, and are able to sequence items, learn and use symbols, detect the interlocutor's intentions, and detect recurrent patterns (Gong et al 2005). The last three abilities are taken as parameters and varied as probabilities in the model.

The simulations show that when these parameters are all of low values, the group of agents can only develop a limited number of holistic signals. However, when these parameters cross some thresholds, a compositional language could emerge with a set of words and a certain dominant word order shared by the agents, which dramatically increases the communication efficacy of the group. The model thus suggests that even the chimpanzees share a great deal with humans, some small differences could divide the two species apart definitely.

References

de Waal, F. B. M. (2005). A century of getting to know the chimpanzee. Nature, 437/7055: 56-59.

Elman, J.L. (2005). Connectionist models of cognitive development: Where next? Trends in Cognitive Science. 9/3:111-117.

Gong, T., Minett, J. A. Ke, J-Y., Holland, J. H. & Wang, W. S-Y. 2005. Coevolution of lexicon and syntax from a simulation perspective, Complexity, 10(6):1-13.

Hauser, M. (2005). Our chimpanzee mind. Nature, 437/7055:60-63.

May, R. (1976). Simple Mathematics Models with very complicated dynamics. Nature, 261(5560):459-467.

Tomasello, M. and J. Call (1997). Primate cognition. New York: Oxford University Press.

Whiten, A. (2005). The second inheritance system of chimpanzees and humans. Nature, 437/7055:52-55.

MAJOR TRANSITIONS IN THE EVOLUTION OF LANGUAGE

SIMON KIRBY

Language Evolution and Computation Research Unit, University of Edinburgh, 40 George Square, Edinburgh EH8 9LL

Maynard Smith & Szathmary (1997) set out a number of major evolutionary transitions in the history of life. Their goal was not merely to enumerate these significant moments of change, but to highlight commonalities between a range of different transitions. If it is possible to identify features shared by a number of different transitions in evolutionary history then we may be able to transfer our understanding of one particular transition to the study of others.

Significantly for our field, Maynard Smith & Szathmary include the origins of language as the last of their major transitions. This is justified because one of the shared features of transitions that they propose is *change in the system of information transmission*. Human language provides us with a framework for the transfer of semantic information, ultimately enabling the reliable persistence of complex socio-cultural systems.

Whilst this is a relevant and interesting feature of language, there is another, perhaps more important, property that we must take into account. Language not only transmits semantic information, it also encodes information about its own construction. In other words, the linguistic system itself is, at least in part, transmitted culturally. The language learner uses utterances received to reconstruct the language of the previous generation. I have argued elsewhere that this means that language is an evolutionary system in its own right (Kirby 2000).

In this paper, I propose that we can extend to the linguistic domain Maynard Smith & Szathmary's view of evolutionary transitions in biology. If language itself is an evolutionary system, then we may expect to find major transitions in the evolution of language. Furthermore, some of the commonalities Maynard Smith & Szathmary find across biological transitions may also be seen in language. To flesh out this proposal, I will hypothesise three major transitions in language evolution (Figure 1).

Mathematical and computational models of the first (e.g. Oudeyer 2005) and second (e.g. Kirby 2000) of these transitions suggest that self-organisation and adaptive processes arising from linguistic transmission can account for their evolution. In other words, although biological changes may accompany these

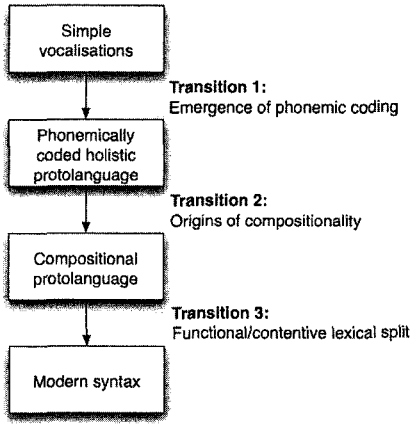


Figure 1. Possible transitions in the evolution of language. Computational and mathematical models suggest that each of these transitions could be driven by self-organising/adaptive mechanisms arising from the cultural transmission of language (although they may be supported by biological changes arising from gene/culture coevolution). Note that this diagram is only a partial picture (for example, it ignores the origin of symbol use, the development of semantic structure etc.).

transitions, we should understand them in the light of language as an evolutionary system in its own right.

Viewing these transitions in this way demonstrates that they share features that Maynard Smith & Szathmary highlight in their work: **division of labour** – different parts of the replicating system have distinct and differentiated functions; **contingent irreversibility** – elements of replicating entities lose their capacity for independent replicability; **new ways of transmitting information** – the range of possible states of a system that can be reliably transmitted increases.

Given these parallels, I will argue that the final transition can likewise be viewed as the inevitable result of language being a system that transmits information about its own construction. This suggests there may be a unified cultural evolutionary mechanism that can take us from unstructured signaling all the way to a syntactic system underpinned by a lexicon divided into functional and contentive elements.

References

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight (Ed.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form* (pp. 303–323). Cambridge: Cambridge University Press.

Maynard Smith, J. and Szathmary, E. (1997). *The Major Transitions in Evolution*. New York: Oxford University Press.

Oudeyer, P.-Y. (2005). From holistic to discrete speech sounds: The blind snow-flake maker hypothesis. In M. Tallerman (Ed.), *Language Origins: Perspectives on Evolution* (pp. 68–99). Oxford: Oxford University Press.

MODELLING UNIDIRECTIONALITY IN SEMANTIC CHANGE

FRANK LANDSBERGEN
Leiden University Centre for Linguistics, Leiden University
PO Box 9515, 2300 RA Leiden, The Netherlands

1. The semantic change of Dutch *krijgen*

Krijgen in Present Day Dutch (PDD) has the prototypical meanings 'to receive' and 'to get' in the inchoative sense of 'to get a headache'. Both senses developed in the 12th/13th century out of the older meaning 'to obtain by effort, to seize'. This meaning has become extinct in PDD, although its relics can still be found in uses such as *te pakken krijgen* 'to get to hold', *in handen krijgen* 'to get in hands' and compounds like *krijgsgevangene* 'prisoner of war'.

This change shows some characteristics of grammaticalization, in that there is semantic bleaching, generalization and the fact that PDD *krijgen* can be used as an auxiliary in restricted contexts. Furthermore, the unrelated English *get* has followed a similar path in its development (Gronemeyer 1999), which could suggest a unidirectional cline.

The aim of this paper is to get a better insight in the relationship between mechanisms of change and unidirectionality in the semantic change of *krijgen*, using computer models of cultural evolution.

2. A computer simulation of semantic change

Unidirectionality in change is evolutionarily interesting for two reasons. First, although changes do not necessarily take place, *if* they take place, the change seems directional in that it follows a specific path. Second, it is very hard to determine the necessary conditions for a language to initiate such a change. In other words, it is very difficult to explain why in language A a change took place in 1300, in language B in 1500 and in language C not at all.

These phenomena are studied for PDD *krijgen* with a computer model. The model is based on the usage-based views that (adult) users continuously construct their linguistic knowledge on the basis of the input they receive, and that change comes about by innovations made by speakers (Traugott & Dasher 2002).

In the model, the meaning of *krijgen* is represented by the direct objects it can be used with. These objects have certain semantic properties, which are represented on a one-dimensional scale. The linguistic knowledge of each individual in the model is a set of objects. In communication, a speaker exchanges an object of his set with a hearer. Hearers construct their knowledge of *krijgen* with this input. Innovation is the use of a new, unlearned direct object.

The role of (partial) synonyms such as *veroveren* 'to conquer' and *pakken* 'to take', which entered the semantic field of *krijgen* at different times in history, is considered by additional selectional pressures.

3. Results

Preliminary findings seem to indicate that the unidirectional tendency in the semantic change of *krijgen* can be explained by an asymmetry in the semantic properties of the set of direct objects. This asymmetry leads to innovations being made on one side of the set more frequently than on the other side. This effect can occur by random drift, without the selection pressures caused by synonyms.

References

Gronemeyer, Claire (1999). On deriving complex polysemy: the grammaticalization of *get*. *English Language and Linguistics, 3.1*, 1-39.

Traugott, Elizabeth Closs & Richard B. Dasher (2002). Regularity in semantic change. Cambridge: Cambridge University Press.

THE ORIGIN OF MUSIC AND ITS LINGUISTIC SIGNIFICANCE FOR MODERN HUMANS

STEVEN MITHEN
School of Human & Environmental Sciences, University of Reading, Whiteknights, PO Box 217, Reading, RG6 6AH, UK
s.j.mithen@reading.ac.uk

While there has been considerable discussion and debate within palaeoanthropology regarding the origin and evolution of language and art, that of music and dance have been neglected. This is as surprising as it is unfortunate as these behaviours are universal amongst human communities today and in the historically documented past. We cannot understand the origin and nature of Homo sapiens and language without also addressing why and how we are a musical species. I argue that while both language and art are most likely restricted to H.sapiens, music – by which I mean singing and dance rather than the use of instruments – has a significantly earlier appearance in human evolution and was utilised by a wide range of hominin ancestors and relatives. Indeed, without appreciating this, we are left with a very restrictive understanding of past communication methods and lifestyles in general.

At present, there are two key approaches to the evolution of language with regard to the nature of 'proto-language. One of these can be called 'compositional' and is especially associated with the work of Derek Bickerton and Ray Jackendoff. In essence, this argues that words came before grammar, and it is the evolution of syntax that differentiates the vocal communication system of H.sapiens from all of those that went before. An alternative approach to proto-language is that developed by Alison Wray and Michael Arbib. This suggests pre-modern communication was constituted by 'holistic' phrases, each of which had a unique meaning and which could not be broken down into constituent words. As such, discrete words that can be combined to make new and unique utterances were a relatively late development in the evolutionary process that led to language. I favour the holistic approach and envisage such phrases as also making extensive use of variation in pitch, rhythm and melody to communicate information, express emotion and induce emotion in other individuals. As such, both language and music have a common origin in a communication system that I refer to as 'Hmmmmm' because it had the

following characteristics: it was Holistic, manipulative, multi-modal, musical and mimetic (see Figure).

Appreciating that human ancestors and relatives had a sophisticated vocal communication system of this type helps to explain numerous features of the archaeological and fossil record. The long-running debate about the linguistic capabilities of the Neanderthals, for instance, arises from apparently contradictory lines of evidence that can now be resolved. That from their skeletal remains suggests the capabilities for vocal communication similar to that of modern humans (and which has, therefore, been assumed to be language) while the archaeological evidence provides few, if any, traces for linguistically mediated behaviour. This seeming paradox is resolved by appreciating that the Neanderthals did indeed have a complex vocal communication system, but it was a type of Hmmmmm rather than language. Another type of Hmmmmm was used by the immediate ancestors of Homo sapiens in Africa, both having origined from a 'proto-Hmmmmm' used by a common ancestor.

While the fossil and archaeological and records provide substantial evidence for the co-evolution of music and language prior to their separation into two largely distinct communication systems in Africa c. 200,000 years ago, further evidence can be found from modern human themselves. Studies of how music and language are constituted in the brain from studies of lesions and brain scans have shown neither total separation nor that one system is entirely dependent on the other. Also, studies of communication by and to infants have stressed the significance of musicality for pre-linguistic humans, suggesting its likely significance for pre-linguistic hominins. In addition, the last decade has seen a recognition that emotion is of central importance to rational decision n making which implies that music – the key means by which emotions are expressed and induced – is likely to have been of central importance to any large-brained hominin.

The separation of a Hmmmmm into the two systems of communication that we now refer to as language and music most likely occurred as part of the process by which modern H. sapiens originated in Africa. The appearance of compositional language would have had a profound cognitive impact, leading to the capacity for metaphor that underlies art, science and religion. Music has continued to deliver the adaptive benefits previously gained from the musicality of Hmmmmm, notably group bonding the expressing of emotional states, and the manipulation of behaviour by inducing emotional states in others.

References

Blacking, J. (1973). *How Musical is Man?* Seattle: University of Washington Press.

Carruthers, P. (2002). The cognitive functions of language. *Brain & Behavioral Sciences, 25,* 657-726.

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind & Body*. London: Weidenfeld & Nicolson.

Wallin, N.L. et al. (eds) (2000). *The Origins of Music*. Cambridge, MA: MIT Press.

CO-EVOLUTION OF LANGUAGE AND BEHAVIOUR IN AUTONOMOUS ROBOTS

SARA MITRI

Ecole Polytechnique Fédérale de Lausanne
EPFL-STI-I2S-LIS, Station 11
CH-1015 Lausanne, Switzerland
sara.mitri@epfl.ch

PAUL VOGT

Language Evolution and Computation Research Unit, University of Edinburgh
40 George Square, Edinburgh, EH8 9LL, UK
paulv@ling.ed.ac.uk

Computational studies on the evolution of language have often been criticised for the large amount of assumptions and simplifications they make. One particular criticism concerns the meaning of words (Ziemke & Sharkey, 2000), which are often predefined (e.g., Kirby & Hurford, 2002), or, in the case where they do develop ontogenetically, are typically unrelated to the agents' behavioural survival task (e.g., Vogt, 2003).

In an attempt to address these problems, this work explores the co-evolution and correlation between language use and behavioural learning in a realistic simulated environment of robotic agents, where a task must be solved to ensure survival. Experiments involving different environmental setups, population sizes and learning schemes are used to study the conditions under which language can emerge and stabilise and how the language affects the collective behaviour of the agents using it. The aim of the study is to investigate whether learning language together with simple survival skills can lead to an overhead in complexity, or can work as a tool for a more rapid emergence of increasingly intelligent behaviour, as well as a flexible, yet robust language.

The simulated Nomad 150 robots in this study are given a "survival task" of collecting red and blue balls and depositing them in a red or blue bin in return for energy. After a ball has been deposited, the agent must decide – using a reinforcement learner – which ball to collect next and where to take it, receiving a reward, depending on the amount of energy gained. If a ball is deposited in the bin of the opposite colour, no energy gain is received, otherwise the increase in energy is regulated by the environmental setup. Three environmental setups are

used: a "cooperation" environment, in which two agents must deposit the same colour ball in the correct bin at the same time; a "division of labour" environment, where two agents must simultaneously deposit opposite colours in the correct bin; and a simple environment, where there is no need for collective action and energy is gained if an agent deposits a ball in the right bin. The robots must learn to coordinate their actions in order to achieve higher performance, which is an incentive for developing and using language. The evolved vocabulary was restricted to 8 wholistic utterances. The implications of using a horizontal model based on the language game model (Steels, 1997) as opposed to a vertical one based on the Iterated Learning Model (Kirby & Hurford, 2002) are compared.

The results and their significance can be summarised in the following four points: (1) A perfect language with a fixed meaning space is not useful in every environment. (2) Where language is useful and a horizontal learning mechanism is used, a stable language evolves and leads to higher performance levels and faster behavioural learning. (3) A larger population size leads to an increase in language coherence, suggesting that language might evolve faster in large populations. (4) Even when a partially-stabilised language is evolved, the minimal performance is still sufficient for survival and is higher than that of non-communicating agents.

These results stress the difficulty of language development and stabilisation, but also show how in an environment where cooperation is highly beneficial, language can stabilise over time to help coordinate the behaviours of individual agents and improve the overall efficiency of a population. The interdependence of behavioural learning and language learning therefore helps to bootstrap both processes, leading to a higher performance in solving a survival task. The outcome of this study contributes to the field of language evolution by showing that language and behaviour can co-evolve as interdependent learning processes in a model where language has a function for survival, but also highlights the benefits of a bottom-up design for intelligent, autonomous and flexible robots that can survive in a dynamically changing environment through the use of a language that is developed during their lifetime according to a survival task.

References

Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148).

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication*, *1*(1), 1-34.

Vogt, P. (2003). Anchoring of Semiotic Symbols. *Robotics and Autonomous Systems*, *43*(2-3), 109-120.

Ziemke, T., & Sharkey, N. (2000). *A Stroll through the Worlds of Robots and Animals: Applying Jakob von Uexküll's Theory of Meaning to Adaptive Robots and Artificial Life.*

ICONIC VERSUS ARBITRARY MAPPINGS AND THE CULTURAL TRANSMISSION OF LANGUAGE

PADRAIC MONAGHAN
Department of Psychology, University of York, York, YO10 5DD, UK

MORTEN H. CHRISTIANSEN
Department of Psychology, Cornell University, Uris Hall, Ithaca, NY 14853, USA

Most theories of language evolution assume that the ability to use symbols was a crucial step towards modern language (for a review see, e.g., Christiansen & Kirby, 2003). Following de Saussure, symbol use is typically construed as the capacity for establishing arbitrary mappings from sounds or gestures to specific concepts and/or percepts for the purpose of communication. Although intuition suggest that iconic relationships between form and meaning should make the learning of such mappings easier (e.g., sound symbolism), recent simulations by Gasser (2004) have demonstrated that, for large vocabularies, the learning advantage is for arbitrary relationships. Because systematic iconic mappings between forms and meanings require strong constraints on the space of possible pairings (e.g., a particular onset phoneme is restricted to only co-occur with a particular facet of meaning) it is only possible to encode efficiently a relatively small number of words. In contrast, arbitrary mappings between form and meaning impose fewer constraints and therefore permit the learning of a large and extendable vocabulary, which is the hallmark of human language[a].

However, the cost of arbitrariness is that generalities about the language structure, such as the lexical category of a word, are not readily learnable from the sounds of the language. Such systematicity has been seen as advantageous, perhaps even necessary, for learning categories (Braine, 1987). In this paper, we hypothesize that cultural transmission has shaped language so as to incorporate certain systematic properties of iconic mappings in order to facilitate the learning of lexical categories. Importantly, the iconic mapping is not between form and meaning but between *form* and *lexical category*.

[a] Though some degree of iconicity may be useful in localized cases, such as expressives in Japanese and Tamil (Gasser, Sethuraman, & Hockema, 2005).

Table 1. Number of significant cues and successful classification for each language.

	Open/Closed		Noun/Verb	
	Cues	Classification	Cues	Classification
English	17	62.1%	7	61.4%
Dutch	14	61.4%	16	71.0%
French	16	62.4%	16	64.9%
Japanese	8	61.8%	17	74.5%

A crucial prediction from the form-category mapping hypothesis is that current languages ought to reveal systematic relations at the lexical category level even though they are absent in sound-meaning mappings. We tested this prediction by analyzing the 1000 most frequent words from large corpora of child-directed speech in English, Dutch, French, and Japanese. For each language, we assessed approximately 50 cues that measured phonological features across each word. Table 1 shows the number of cues that significantly distinguished function from content words and nouns from verbs in each language (corrected for multiple comparisons). Classifications using discriminant analysis tested that the cues were able to correctly identify the category of a significant proportion of the words (all $p < .001$).

The presence of significant effects across four distinct languages supported our hypothesis that form-category systematicity is a property of natural languages. Because the number of lexical categories in any language is minimal and restricted, the strict constraints imposed on form-meaning mappings do not apply. Consequently, cultural transmission is likely to have favored languages that incorporate such form-category systematicity as it facilitates initial learning of grammatical structure without sacrificing vocabulary size. Thus, as indicated by our analyses, current languages may have evolved to incorporate an optimal compromise between arbitrary and iconic mappings in language learning.

References

Braine, M.D.S. (1987). What is learned in acquiring word classes: A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp.65-87). Hillsdale, NJ: LEA.

Christiansen, M. H. & Kirby, S. (2003). *Language evolution.* Oxford: OUP.

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the Cognitive Science Society Conference* (pp.434-439). Hillsdale, NJ: LEA.

Gasser, M., Sethuraman, N., & Hockema, S. (2005). Iconicity in expressives: An empirical investigation. In S. Rice and J. Newman (Eds.), *Experimental and empirical methods.* Stanford, CA: CSLI Publications.

MOTHER TONGUE: CONCOMINANT REPLACEMENT OF LANGUAGE AND MtDNA IN SOUTH CASPIAN POPULATIONS OF IRAN

IVAN NASIDZE & MARK STONEKING

Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, Deutscher Platz 6, D-04103, Leipzig, Germany

Comparative analysis of mtDNA and Y chromosome variation in the same groups reveals their maternal and paternal histories. Often these are the same, but sometimes there are differences in the patterns of mtDNA and Y chromosome variation, which then provide novel insights into the history of such groups. We describe here an instance in which patterns of mtDNA and Y chromosome variation differ, for the Gilaki and Mazandarani groups from the South Caspian region of Iran.

The Gilaki and Mazandarani occupy the South Caspian region of Iran and speak closely-related languages belonging to the North-Western branch of Iranian languages (Ethnologue, 2000), as do other groups in this region. Little is known about their history; it has been suggested that their ancestors came from the Caucasus region, perhaps displacing an earlier group in the south Caspian (Negahban, 2001). Linguistic evidence supports this scenario, in that the Gilaki and Mazandarani languages (but not other Iranian languages) share certain typological features with Caucasian languages (Stilo, 1981, 2005).

Here, we report the results of mtDNA and Y-chromosome analyses of the Mazandarani and Gilaki, in comparison with their geographic and linguistic neighbors (i.e., other Iranian groups) and with South Caucasian groups.

Based on mtDNA HV1 sequences, the Gilaki and Mazandarani most closely resemble their geographic and linguistic neighbors, namely other Iranian groups. However, their Y chromosome types most closely resemble those found in groups from the South Caucasus. A scenario that explains these differences is a south Caucasian origin for the ancestors of the Gilaki and Mazandarani, followed by introgression of women (but not men) from local Iranian groups, possibly because of patrilocality. Given that both mtDNA and language are maternally-transmitted, the incorporation of local Iranian women would have resulted in the concomitant replacement of the ancestral Caucasian language and mtDNA types of the Gilaki and Mazandarani with their current Iranian language and mtDNA types. Concomitant replacement of language and mtDNA may be a more general phenomenon than previously recognized.

References

Ethnologue (2000). (www.ethnologue.com).

Negahban, E.O. (2001). Gilân. In E. Yarshater (Ed.), *Encyclopedia Iranica* (pp. 618-634). New York: Bibliotheca Persica Press.

Stilo, D. (1981). The Tati language group in the sociolinguistic context of Northwestern Iran and Transcaucasia. *Iranian Studies, 14,* 137-185.

Stilo, D. (2005). Iranian as buffer zone between the universal typologies of Turkic and Semitic. In E.A. Csato, B. Isaksson & C. Jahani (Eds.), *Linguistic Convergence and Areal Diffusion. Case Studies from Iranian, Semitic and Turkic.* (pp. 35-63). London: Routledge Curzon.

WHAT CAN GRAMMATICALIZATION TELL US ABOUT THE ORIGINS OF LANGUAGE?

FREDERICK J. NEWMEYER
University of Washington
fjn@u.washington.edu

Grammaticalization is the historical process whereby grammatical elements lose some of their 'independence'. Nouns and verbs become pronouns and auxiliary elements respectively, pronouns and auxiliaries become affixes, and so on. This change in structure is often (but not always) accompanied by 'bleaching' (loss of semantic specificity) and phonetic reduction. Interestingly, grammaticalization is largely unidirectional. It is quite rare, for example, for an affix to change historically into an auxiliary or a pronoun or for a pronoun or auxiliary to become a noun or verb.

The unidirectionality of grammaticalization has led some scholars to speculate that this process provides a key to what the grammar of the earliest human language might have looked like (see Heine and Kuteva 2002; Hurford 2003; Burling 2005). Since the process starts with nouns and verbs, the argument goes, the earliest stages of language might have possessed these elements, but not auxiliaries, pronouns, affixes, or other elements that play a principally 'grammatical' role. For simplicity, I refer to the position that grammaticalization leads us back to the categorial inventory of the earliest human language as the 'Grammaticalization→Origins' theory or 'G→O'.

For the following reasons I am skeptical that the unidirectionality of grammaticalization invites the conclusion that the only grammatical categories at the dawn of human language were nouns and verbs:

- Grammaticalization is a cycling process in which existing lexical items are worn down, but at the same time new ones are created. G→O demands picking one point on the cycle as the starting point, namely the point where lexical items are in place, but which for some reason have never undergone grammaticalization. Why should one assume that?

- Not all elements that arise from grammaticalization play a largely grammatical role. Elements with real semantic content, such as prepositions and tense/aspect morphemes, can also be the product of grammaticalization. Yet there is no reason to assume that the earliest humans could not express concepts like 'in' and 'past time'. Perhaps these concepts were indeed expressed by nouns and verbs, or perhaps prepositions and tense morphemes existed at the outset of human language as independent categories, or perhaps they were *already*

grammaticalized (say in Proto-Language). Both possibilities diminish the conclusions that can be drawn from grammaticalization about *human* language.

- Languages spoken today differ enormously from each other in terms of the degree to which they manifest the effects of grammaticalization. For example, Riau Indonesian manifests very little (Gil 2001). But if a language spoken today can manifest grammaticalization as poorly as a language spoken 100,000+ years ago putatively did, then it follows that grammaticalization per se cannot tell us very much about the origin and evolution of language.

- G→O depends on a degree of uniformitarianism in language history that might not be warranted. If what is frequently expressed has changed over time, or if the balance of functional and 'counter-functional' (Haspelmath 1999) factors has not remained constant over time, then the process of grammaticalization might lack sufficient unidirectionality (or at least consistency) to support G→O.

To summarize, observations about the process of grammaticalization are not likely to lead to insights about the origins and evolution of human language. While it is possible that the first true human language possessed only two categories, namely nouns and verbs, grammaticalization does not provide much evidence for that conclusion.

References

Burling, Robbins. 2005. *The talking ape: How language evolved.* Oxford: Oxford University Press.

Gil, David. 2001. Creoles, complexity, and Riau Indonesian. *Linguistic Typology* 5:325-371.

Haspelmath, Martin. 1999. Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18:180-205.

Heine, Bernd, and Kuteva, Tania. 2002. On the evolution of grammatical forms. In *The transition to language*, ed. Alison Wray, 376-397. Oxford: Oxford University Press.

Hurford, James R. 2003. The language mosaic and its evolution. In *Language evolution,* eds. Morten H. Christiansen and Simon Kirby, 38-57. Oxford: Oxford University Press.

BOOTSRAPPING SHARED COMBINATORIAL SPEECH CODES FROM BASIC IMITATION: THE ROLE OF SELF-ORGANIZATION

PIERRE-YVES OUDEYER

Sony CSL Paris,
75005 Paris, France

Human vocalizations have a complex organization. They are discrete and combinatorial: vocalizations are built through the combination of units, and these units are systematically re-used from one vocalization to the other. These units appear at multiple levels (e.g.the gestures, the coordination of gestures, the phonemes, the morphemes). While for example the articulatory space that defines the physically possible gestures is continuous, each language only uses a discrete set of gestures. While there is a wide diversity of the repertoires of these units in the world languages, there are also very strong regularities (for example, the high frequency of the 5 vowel system /e,i,o,a,u/). Moreover, in each language there are "rules" which determine what combinations of phonemes can or cannot be produced: this is what is called phonotactics.

It is then obvious to ask where this organization comes from. There are two complementary kinds of answers that must be given (Oudeyer, 2006). The first kind is a functional answer stating what is the function of systems of speech sounds, and then showing that systems having the organization that we described are efficient for achieving this function. This has for example been proposed by (Lindblom, 1992) who showed that discreteness and statistical regularities can be predicted by searching for the most efficient vocalization systems. This kind of answer is necessary, but not sufficient : it does not say how evolution (genetic or cultural) might have found this optimal structure. In particular, naive darwinian search with random mutations (i.e. plain natural selection) might not be sufficient to explain the formation of this kind of complex structure : the search space is just too large (Ball, 2003). This is why there needs a second kind of answer stating how evolution might have found these structures. In particular, this amounts to show how self-organization might have constrained the search space and helped natural selection. This can be done by showing that a much simpler system spontaneously self-organizes into the more complex structure that we want to explain.

In this talk, I will present a computational model which is a generalization of the model developped in (Oudeyer, 2005a,b, 2006), in which only one type of neuron is used. This model involves a population of agents endowed with operational models of the ear, of the vocal tract, and of the neural structures that

connect them. It shows how the generic coupling of evolutionarily simple neural structures can produce spontaneously, thanks to self-organization, a primitive combinatorial vocalization system with phonotactics shared by a population of agents whose vocalizations were initially holistic and unorganized. What is original is that: 1) there is no explicit pressure for building a system of distinctive sounds (and there are no repulsive forces whatsoever in the system); 2) agents do not possess capabilities of coordinated interactions, in particular they do not play language games; 3) agents possess no specific linguistic capacities; 4) initially there exist no convention that agents can use.

I will also propose a new interpretation of this model. The neural structures which are used look very much like what is needed for basic vocal imitation, defined as the capacity to reproduce a sound which has been perceived. As a consequence, they might have biologically evolved under a pressure for imitation. What is interesting, is that thanks to self-organization, a combinatorial speech code with phonotactics is formed as a side effect, even if such a speech code is not necessary for imitation (indeed, basic vocal imitation does not even need a system of distinctive sound categories nor a repertoire of discrete vocalizations). This shows that the evolutionary step from vocal imitation to shared combinatorial human-like speech codes might have been rather small. I will also discuss how this is confirmed by the observation that many species of birds and whales capable of vocal imitation do indeed possess such a shared primitive combinatorial "vocal" code.

References

Ball P. (2001) The self-made tapestry, Pattern formation in nature, Oxford University Press.

Lindblom, B. (1992) Phonological Units as Adaptive Emergents of Lexical Development, in Ferguson, Menn, Stoel-Gammon (eds.) *Phonological development: Models, Research, Implications*, York Press, Timonnium, MD, pp. 565-604.

Oudeyer, P-Y. (2005a) The Self-Organization of Speech Sounds, *Journal of Theoretical Biology*, Volume 233, Issue 3, pp.435—449

Oudeyer, P-Y. (2005b) The self-organisation of combinatoriality and phonotactics in vocalization systems, *Connection Science*, vol. 17, No. 3, pp. 1-17

Oudeyer, P-Y. (2006) Self-Organization in the Evolution of Speech, *Studies in the Evolution of Language*, Oxford University Press.

HOW LANGUAGE CAN GUIDE INTELLIGENCE

LEONID I. PERLOVSKY
Air Force Research Laboratory, 80 Scott Rd., Hanscom Air Force Base, MA 01731, USA

JOSÉ F. FONTANARI
Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, São Carlos, SP 13560-970, Brazil

Today the favored explanation for the evolution of language seems to lie in the field of social intelligence. According to this view, language developed as a social glue: the primary selective pressure being the binding together of the early hominids in large groups, with gossip substituting costly grooming as the main mechanism of social interaction and cohesion (Dunbar, 1998). Nevertheless, advancing the argument that, taking language away, human social life may not be more complex than those of chimpanzees and bonobos, Calvin & Bickerton (2000) have championed the viewpoint that the selective pressures for language must have come from the brute exigencies of survival, e.g., hunting, food gathering and predator detection, rather than from human social life. Here we build on this proposal by considering these elementary survival needs as problems to be solved by the (artificial, in our case) organisms and ask how and whether communication can improve the performance of the individual organisms to solve a specific problem. This approach is in line with the seditious view of language as the cause of our species becoming more intelligent rather than that language being an inevitable consequence of greater intelligence.

The specific task we consider in this contribution is the differentiation problem, i.e., how organisms develop a more detailed knowledge of their surroundings. In particular, we address the problem of the "true" number of objects in the world, which is described as follows. We assume that the world contains a certain number of objects, e.g., points on a single axis or sets of points drawn from a Gaussian distribution, and that the organisms are endowed with a categorization system inspired in the modeling field theory (MFT) approach (Perlovsky, 2001) that, in principle, enables them to distinguish, through the creation of internal representations or concepts, those objects. At the beginning each organism starts with a single concept-model – a modeling

neuronal field chosen randomly - which then becomes associated to a specific object or group of objects. The organisms then exchange information – the values of their models or, alternatively, signs (words) associated to those models – which prompt them to create new concept models and finally to identify unambiguously all objects. We discuss the trade-off between the number of objects and the number of organisms needed to achieve perfect categorization. In doing so we demonstrate that categorization is better (in the sense that all objects are identified) and faster when communication is allowed.

This formulation allows us to go beyond the simplistic view of language as a mapping between objects in the real world and words (or, alternatively, between conceptual representations – meanings - and words) that underlies most of the simulation models on the evolution of language. In fact, since de Saussure it is known that there are at least two mapping operations between the real world and language: first our sense perceptions are mapped onto a conceptual representation, and then this conceptual representation is mapped onto a linguistic representation (Bickerton, 1990). The importance of the incorporation of this second hierarchy level in models for language evolution is the fact that linguistic representations can help creating conceptual categories , which may aid in coping with the external world. Another approach, that also shows the benefit of language to solve tasks that require the coordinated action of distinct agents, is the Predator-Prey Pursuit Problem (see, e.g., Jim & Giles, 2000). However, rather than provide additional support to this hardly surprising finding, our aim here is to verify the emergence of improved structure in combined categorization and communication abilities when the more realistic two-steps mapping between objects and words is implemented through the MFT formalism.

References

Dunbar, R. (1998). *Grooming, Gossip, and the Evolution of Language*. Cambridge: Harvard University Press.

Bickerton, D. (1990). *Language & Species*. Chicago: University of Chicago Press.

Calvin, W. H., & Bickerton, D. (2000). *Lingua ex Machina*. Cambridge: MIT Press.

Jim, K.-C., & Giles, C. L. (2000). Talking Helps: Evolving Communication Agents for the Predator-Prey Pursuit Problem. *Artificial Life, 6*, 237-254.

Perlovsky, L. I. (2001). *Neural Networks and Intellect: Using Model-Based Concepts*. Oxford: Oxford University Press.

THE ROLES OF SEGMENTATION ABILITY IN LANGUAGE EVOLUTION

KAZUTOSHI SASAHARA
Laboratory for Biolinguistics, Brain Science Institute, RIKEN, Japan
sasahara@brain.riken.jp

BJORN MERKER
Department of Psychology, Uppsala University, Sweden
gyr694c@tninet.se

KAZUO OKANOYA
Laboratory for Biolinguistics, Brain Science Institute, RIKEN, Japan
okanoya@brain.riken.jp

We focus on segmentation ability as a prerequisite of language, studying what part it plays in language evolution with a simple computational model. Language is mediated by distinct sounds and has a characteristics of 'duality'. For such a structure, it is necessary to have segmentation ability, which is an ability to find out discrete units in continuous sound sequences.

To model segmentation ability, we review some experimental findings. In songbirds, it has been found that male Bengalese finches have songs with duality; a chunk consists of phonemes and a song consists of chunks (Okanoya, 2002). A male juvenile of the Bengalese finch learns a song from his father within a certain period. To do so, he must detect the discrete parts (e.g. song elements and chunks) in the flow of the song sample. In infants, a number of experiments have shown that infants are able to find discrete patterns in the flow of adults' utterances by detecting word frequency, the transition rate of sounds, accent patterns, and so on (Tomasello, 2003). Both cases have two common features: (i) the statistical cues in strings contribute to segmentation; and (ii) the dyadic interaction (i.e. father and juvenile, mother and infant) is a 'leader-follower' kind in which one of the two is a well-versed agent and the other is not; hence there is an asymmetry of information flows.

In light of above consideration, we model an evolution of discourse where agents utter strings by turns. Let's suppose a society of N agents, each of which can produce long sound strings and has simple statistical ability. Each agent is modeled by a recurrent network (RNN) that studies the transition rate of sound

elements in the sound strings it hears. In initial state, all network weights of every agent are randomly initialized. Then two agents are randomly chosen to engage in conversation. The utterances of the agents consist of the outputs of their RNN and they are translated into letters (here, *A, B, ..., J*), considered sound elements. When one agent utters a sequence of sounds, the other agent hears it one-by-one and predicts the next sound element in the utterance. After that, the agent's RNN is trained with supervised learning in such a way that it can predict the transition of sound elements better. Then the agents take turns uttering and hearing. The procedures are repeated over within a certain number of discourses.

With this model, we demonstrate how common shared words (i.e. frequently used sound patterns) emerge and how the distribution of sound elements changes from random initial state as common words increase. In the early stages of the evolution, common words were rare in the artificial society because the patterns of sounds were almost random. However, once common patterns emerged in sound strings, some of them came to stay in the discourses of the agents. Furthermore, we consider how the leader-follower interaction of agents contribute to the emergence of words. Self-organization the leader-follower interaction in our agents was difficult just because of statistical cues in discourse. Our results show that if agents have simple statistical ability, the frequently appearing patterns in sound strings may become established as words through the interaction of the agents and that emerging words may affect succeeding discourses in evolution.

So far the certain patterns of sounds have been described as words; however these are not exactly the same as words due to the lack of meaning. If we take the following 'mutual segmentation' hypothesis into account, our model may deal with syntax and semantics within a single framework (Merker & Okanoya, 2005). Suppose a society without language. When agents with segmentation ability collaborate, the common parts of behavioral, environmental and social context they face and the common parts of sound strings they utter could be mutually segmenting, and the segmented small parts of sound strings could link to ever more specific contexts; a word and a meaning could emerge into co-existence. Our model at present doesn't have any context. We plan to extend it by introducing behavioral context of agents (e.g. sensory-motor experience) to explore that hypothesis.

References

Merker, B., & Okanoya, K. (2005). Contextual semanticization of songstring syntax: a possible path to human language. *Proceedings of Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC 2005)*, 72–76.

Okanoya, K. (2002). Sexual display as a syntactical vehicle. In *The transition to language* (pp. 46–63). Oxford University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press.

PRIMATE SOCIAL COGNITION AND THE COGNITIVE PRECURSORS OF LANGUAGE

ROBERT SEYFARTH & DOROTHY CHENEY
University of Pennsylvania
Philadelphia, PA 19104 USA
seyfarth@psych.upenn.edu, cheney@sas.upenn.edu

If we accept the view that language first evolved from the conceptual structure of our pre-linguistic ancestors, several questions arise, including: What kind of structure? Concepts about what? In this talk, we focus on some recent field experiments which suggest that nonhuman primates have a sophisticated knowledge of other animals' social relationships. This knowledge is based on discrete-valued traits (identity, rank, kinship) that are combined to create a representation of social relations that is hierarchically structured, open-ended, rule governed, and independent of sensory modality. We propose that in the earliest stages of language evolution communication had a formal structure that grew out of its speakers' knowledge of social relations.

AGONISTIC SCREAMS IN WILD CHIMPANZEES: CANDIDATES FOR FUNCTIONALLY REFERENTIAL SIGNALS

KATIE SLOCOMBE & KLAUS ZUBERBÜHLER

School of Psychology, University of St Andrews, St Mary's Quad, St Andrews, KY16 9JP, U.K.

The comparative perspective examines the abilities of non-human primates in order to identify which cognitive capacities involved in language processing are phylogenetically old, with their evolutionary roots deep in the primate lineage, and which cognitive capacities are unique to humans. Some non-human primates have demonstrated the capacity to communicate about external objects or events, suggesting primate vocalizations can function as referential signals. From a comparative perspective, functional reference can be considered a precursor to the semantic capacities evident in modern human listeners. However, despite evidence for functionally referential communication in a variety of animal species, and particularly monkeys (Seyfarth et al., 1980; Macedonia, 1990; Zuberbühler, 1999) there is no comparable evidence available for any of the great ape species. This is problematic both because apes are more closely related to humans and they are widely considered cognitively more advanced than monkeys (Byrne, 1995). We attempt to address this problem by examining the agonistic screams of the chimpanzees for evidence of functional reference.

We studied screams produced during agonistic encounters by the wild chimpanzees of Budongo Forest, Uganda. Vocalisations were recorded and the behaviour and context accompanying each call was noted in detail. Acoustic analysis of the vocalizations allowed us to provide quantitative descriptions of the fine acoustic structure of the calls. We were then able to determine if the chimpanzees were producing context-specific calls by examining the relationship between the acoustic structure of the calls and the eliciting context.

The chimpanzees of Budongo forest, Uganda, give acoustically distinct screams during agonistic interactions depending on the role they play in a conflict. We determined the role the chimpanzees played in a conflict (victim or aggressor) by noting the presence of specific behaviours. We analysed the acoustic structure of screams of 14 individuals, both in the role of aggressor and victim. We found consistent differences in the acoustic structure of the screams,

across individuals, depending on the social role the individual played during the conflict. A discriminant function analysis, based on the ten acoustic measures taken from the calls, was able to correctly classify calls according to the eliciting context on 93% of occasions (cross-validated). We observed a few instances of third party intervention in agonistic interactions, where the third party approached from out of sight. We suggest that the third party was using the information encoded in the screams of the fighting individuals to influence their decision to intervene. We conclude that these two distinct scream variants, produced by victims and aggressors during agonistic interactions, may therefore be promising candidates for functioning as referential signals.

We then examined the structure of victim screams in more detail to see if information about the severity of the attack or the relative rank of the opponent was also encoded in the screams. Chimpanzees produce victim screams that vary acoustically, according to the severity of the aggression the victim is experiencing. There was no evidence that screams varied according to the relative size of the difference in rank between the victim and aggressor. We conclude that victim screams are produced in a context-specific manner and as such have the potential to function referentially: despite the likely emotional basis for these calls, listeners could infer both the role of the individual and if a victim, the severity of the attack, from just hearing the screams.

Playback experiments are now needed to test whether listening individuals do extract and use this valuable information in naturally occurring situations. If they do then these calls will begin to address the current anomaly of the absence of evidence for naturally occurring functional reference in apes. This will strengthen the view that the semantic abilities of human listeners build on phylogentically old traits. In addition once we fully understand the function of these calls we can explore the possibility that these signals are being produced intentionally: the first step towards understanding the evolution of basic linguistic reference.

References

Byrne, R. W. (1995). *The thinking ape: evolutionary origins of intelligence.* Oxford, Oxford Univ. Press.

Macedonia, J. M. (1990). "What is communicated in the antipredator calls of lemurs: evidence from playback experiments with ring-tailed and ruffed lemurs." *Ethology 86*: 177-190.

Seyfarth, R. M., Cheney, D. L., Marler, P. (1980). "Vervet monkey alarm calls: Semantic communication in a free-ranging primate." *Animal Behaviour 28*(4): 1070-1094.

Zuberbühler, K., Cheney, D. L., Seyfarth, R. M. (1999). "Conceptual Semantics in a Nonhuman Primate." *Journal of Comparative Psychology 113*(1): 33-42.

AN INDIVIDUAL-BASED MECHANISM FOR ADAPTIVE SEMANTIC CHANGE

DANIEL W. SMITH

Biology Department, Woods Hole Oceanographic Institution, MS 34
Woods Hole, MA 02543 US

D.W. Smith (2004) argued that in the absence of countervailing perceptual or cognitive constraints, word meanings might be expected to shift over time. Such semantic change would arise because the meanings of many words extend over a continuous range, while any individual speaker's experience with referents across such ranges must be finite, and will lack precision regarding the ranges' exact boundaries. Because of this limited knowledge, individual speakers, in learning words' meanings, must either guess at their range boundaries or underestimate them. This could lead to both variation among idiolects and *diachronic semantic change.*

A pool of variation among individuals' meanings could provide the flexibility to allow adaptive change in an average, or population-level, meaning, much as the presence in some individuals of alleles for adaptive biological traits allows those alleles eventually to become fixed within populations. But here a problem analogous to that of the "hopeful monster" in biology rears its head: like an animal newly bearing a beneficial mutation, but lacking another such to mate with, might not a speaker with a changed meaning range prove unable to communicate with others, and thus unable, too, to propagate his or her innovation? Computer simulations using Matlab (The Mathworks, Natick, MA, US) suggest that this problem can be overcome by gradually varying individuals' meaning ranges so that some may happen to change in the same direction, while at the same time successful communication is "reinforced."

In the simulations, a test population of "speakers" was initialized to have a vocabulary of 3 words, with the meaning ranges for those words equally spaced across a one-dimensional space, each covering one third of it. However, the items provided as candidates for description by the words were randomly distributed through only the lower half of the overall space. Different real-world interpretations of this model could be chosen, but one possibility is to think of the space itself as a dimension of physical space, and each item placed in it as a prey item. The prey items should be imagined as somewhat elusive creatures—birds flitting through dense foliage, say, or small rodents popping briefly from

underground burrows—so that a hearer might typically learn (roughly) where to find one of these delicacies only from someone else's spoken sighting report.

The basic transaction in the simulation consisted of a speaker choosing a word to indicate a prey item's (approximate) location to a hearer. If the prey item was in fact located within a specified distance from the center of the *hearer's* meaning range for that word, it would be considered "caught," and both speaker and hearer would accrue one credit for the current model iteration. At the end of each iteration, "successful" meaning ranges (as measured by better-than-average total credits accrued to their users) were left unchanged. Meaning ranges for each speaker who had accumulated fewer credits than the average for that round were varied by addition of normally distributed random components. The typical result over 100 iterations of this model was that both the "lower" and "middle" words in the original space migrated lower, providing better coverage of that portion of the model space where prey items actually were.

Both variation among idiolects and diachronic semantic change are in fact observed in living languages (Reiter & Sripada 2002; Traugott & Dasher 2002). In the model, the coincidence of some speakers' meaning ranges varying in an adaptive direction, when "reinforced," could lead in the long run to a more productive population-level partitioning of the overall meaning range. Moreover, because ranges moved only incrementally, the failure rate for communication did not need to rise prohibitively during such a transition.

In the real world, semantic variation is more multifarious than in the model, and different individuals' meaning ranges may vary in the same direction less by coincidence than as a result of observational learning. But this model suggests that even a simple and "mindless" source of variation can provide the flexibility needed to support semantic change without making communication fail. Variation consciously engineered by individual speakers toward better communication would likely lead to even quicker overall change.

References

Smith, D. W. (2004). Range-estimation in learning word meanings: a recipe for semantic change? Poster presented at EVOLANG V (Fifth International Conference on the Evolution of Language), Leipzig, Germany.

Reiter, R., & Sripada, S. (2002). Human variation and lexical choice. *Computational Linguistics*, *28*, 545-553.

Traugott, E. C., & Dasher, R. B. (2002). *Regularity in semantic change.* Cambridge, UK: Cambridge University Press.

A HOLISTIC PROTOLANGUAGE CANNOT BE STORED, CANNOT BE RETRIEVED

MAGGIE TALLERMAN
Linguistics Section, University of Newcastle upon Tyne
Newcastle NE1 7RU, U.K.

A minimal assumption in language evolution has to be that the mental lexicon evolved, but at earlier stages of hominid evolution was less sophisticated than in *Homo sapiens*. It cannot conceivably be the case that the mental lexicon at any pre-*sapiens* stage was more COMPLEX than it is today. However, recent proposals by Mithen (2005) and Arbib (2005) for a holistic protolanguage, assumed to be in use at least 500kya, seem to imply exactly this: the presumed content of holistic messages requires a lexicon with storage and retrieval capacities vastly superior to those available to *sapiens*. Such a protolanguage cannot reasonably be attributed to hominids at a less advanced stage of linguistic evolution.

Arbib (2005) proposes a protolanguage "composed of mainly 'unitary utterances' that symbolized frequently occurring situations [...] without being decomposable into distinct words". He continues: "Unitary utterances such as 'grooflook' [...] might have encoded quite complex [...] commands such as 'Take your spear and go around the other side of that animal and we will have a better chance together of being able to kill it'" (Arbib 2005: 118). In a similar vein, Mithen (2005: 172) proposes such holistic messages as "Go and hunt the hare I saw five minutes ago behind the stone at the top of the hill". Obviously, for such utterances to be produced, it must be possible both to store and retrieve them. However, Arbib's example encodes (the meaning of) no less than five distinct predicates and nine arguments (some covert), yet is supposedly stored as a single LEXICAL CONCEPT. Compare sentence production by modern speakers:

> Utterances comprising several sentences are rarely laid out entirely before linguistic planning begins. Instead, all current theories of sentence generation assume that speakers prepare sentences incrementally. Speakers can probably choose conceptual planning units of various sizes, but the typical unit appears to correspond roughly to a clause. (Treiman et al. 2003)

Yet Arbib's example is the equivalent of five clauses; Mithen's is three. If modern speakers engage in conceptual *planning* only at the level of a single clause – a mental proposition – how could early hominids possibly have had the lexical capacity to store, retrieve (and execute) a single lexical concept which

corresponds to several clauses' worth of semantic content? And if they could, why has this amazing conceptual capacity been lost?

A proponent of holistic protolanguage may counter that the capacity has not been lost, but surfaces in the storage and production of formulaic utterances such as *kick the bucket, you can't have your cake and eat it.* However, the properties of idioms exactly demonstrate the COMPOSITIONALITY of modern language rather than a holistic nature. Discussing such 'single-concept-multiple-lemma' cases, Levelt et al. (1999: 12) note that "the production of *kick the bucket* probably derives from activating a single, whole lexical concept, which in turn selects for multiple lemmas". Crucially, the 'lemmas' (syntactic units) of idioms are treated separately for morphological/syntactic purposes, e.g.

(1) He may kick the bucket. (2) If he kick*s* the bucket...
(3) If he kick***ed*** the bucket... (4) I hope he kick*s* the ***bloody*** bucket soon.

So clause-length (or multi-clause) idioms are not the equivalent of Arbib's and Mithen's proposed holistic utterances, since idioms have component parts which are easily manipulated. Thus, even if holistic protolanguage contained only single-proposition utterances, these are quite distinct from modern idioms.

How, then, did speakers of holistic protolanguage achieve lexical access to their complex concepts, and how did they get from semantics straight to sound? In modern language, intermediate knowledge of a word's syntactic and semantic properties (word class, gender, semantic field) aid or hinder its recall. Features of lexical items and connections between them facilitate access. Since no such hooks can exist in a holistic protolanguage, lexical access would appear to be a task of immense cognitive difficulty, one that would surely have been beyond the capabilities of the early hominids envisaged to use such holophrases. Conversely, in a synthetic protolanguage, protoclasses for nouns and verbs would derive from primate cognitive structure, and semantic fields can accrue gradually as more symbols (protowords) are added.

References

Arbib, M.A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105-167.

Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.

Mithen, S. (2005). *The singing Neanderthals: The origins of music, language, mind and body.* London: Weidenfeld & Nicholson.

Treiman, R., Clifton, C., Jr., Meyer, A.S., & Wurm, L.H. (2003). Language comprehension and production. In A. F. Healy & R. W. Proctor (Eds.), *Experimental psychology.* Volume 4 in I. B. Weiner (Editor-in-Chief), *Handbook of psychology* (pp. 527-547). New York: Wiley.

RECOMBINANCE IN THE EVOLUTION OF LANGUAGE

LEONARD TALMY
Department of Linguistics, Center for Cognitive Science,
University at Buffalo, State University of New York
609 Baldy Hall, Buffalo, NY 14260
talmy@buffalo.edu

In pre-language hominids, the vocal auditory channel, as it was then constituted, may have been inadequate as a means of transmission for communication involving certain levels of thought and interaction. If this circumstance were regarded metaphorically in terms of conflicting evolutionary pressures or forces, it could be seen as a bottleneck. On the one hand, the capacity within individuals for thought, i.e., for conceptual content and its processing, perhaps was already relatively great -- or was developing or had the near potential to develop -- in its range of content, of granularity, and of abstractness, as well as in complexity and speed. The potential also existed for the development of the interaction among individuals, so that it included the communication of more advanced thought more quickly. Such developments in thought and in its communication would have had selective advantage. On the other hand, the vocal-auditory channel then had limitations that made it unable to represent enough advanced conceptual content with enough fidelity and speed.

Another means of transmission, the bodily-visual channel in general or the manual-visual channel in particular, had properties that might have allowed it to handle the new communicative load. Within modern-day sign languages, the so-called "classifier subsystem" presents a kind of existence proof for the cognitive feasibility of a manual-visual system conveying advanced conceptual content with fidelity and speed. It has two main enabling properties: its extensive parallelness, that is, its numerous concurrent parameters for representing different kinds of content at the same time, and its extensive iconicity. But these are minimal in the vocal-auditory channel. Nevertheless, for whatever reasons, the manual-visual channel did not follow an evolutionary path toward becoming the main means of communication for humans, while the vocal-auditory channel did. For this to happen, though, this channel had to acquire certain characteristics that could overcome its limitations. The proposal here is that it shifted from being a largely analog system to being a mainly digital system.

As analyzed here, digitalness has a lesser or greater extent, cumulatively built up from a succession of four factors: a) discreteness, b) categoriality, c) recombination, and d) emergentness. These can be characterized as follows. a) Distinctly chunked elements, rather than gradients, form the basis of some domain in question. b) The chunked elements function as qualitatively distinct categories rather than, say, merely as steps along a single dimension. c) These categorial chunks systematically combine with each other in alternative arrangements rather than occurring only at their home sites. d) These arrangements each have their own new higher-level identities rather than remaining simply as patterns. The term "recombinance" is here applied to any cognitive domain that includes both recombination and emergentness.

Human language is extensively recombinant. By one analysis, it has six distinct forms of recombination, of which three or possibly four also exhibit emergentness. In particular, there are four formal types of recombination: phonetic features combining into phonemes, phonemes combining into morphemes, morphemes combining into idioms, and morphemes and idioms combining into expressions -- with the first three of these producing new emergent identities. And there are two semantic types of recombination: semantic components combining into morphemic meanings, and morphemic meanings combining into expression meanings -- with the first of these perhaps yielding a new emergent identity.

A heuristic survey of various cognitive systems such as visual perception and motor control suggests that discreteness and categoriality appear in many of them. But candidates for recombination and emergentness in these systems seem rarer and more problematic. Language evolved recently and may have borrowed or tapped into organizational features of the extant cognitive systems. The cognitive system of language thus could have readily acquired its discrete and categorial characteristics from other systems. But language seems to be the cognitive system with the most types and the most extensive use of recombinance. The question thus arises whether language, as it evolved, adopted a full level of recombinance already present in another cognitive system, increasing it somewhat; adopted a minor level of recombinance from another cognitive system, elaborating it greatly; or developed full recombinance newly as an innovation.

In any case, the evolutionary development of digitalized recombinance in the cognitive system underlying the vocal-auditory channel rendered it capable of transmitting a greater amount of more complex conceptual content with greater fidelity and speed. Again in metaphoric terms, this development

resolved the bottleneck by loosening the prior constriction -- or rather, by circumventing the built-in limitations of the channel.

Finally, consideration can be given to whether thought coevolved with language in certain respects, such as in its degree of digitalness in general and recombination in particular, in its "crispness", and in its voluntariness.

APE GESTURES AND HUMAN LANGUAGE

MICHAEL TOMASELLO

Department of Developmental and Comparative Psychology
Max Planck Institute for Evolutionary Anthropology
Deutscher Platz 6, D-04103 Leipzig, Germany
tomas@eva.mpg.de

Apes and other nonhuman primates have very little voluntary control over their vocal signals. In contrast, apes have much more voluntary control over their gestures - using them flexibly as needed, even in combination, in different communicative circumstances. Moreover, in using many gestures a signaler must be concerned about whether a recipient is attending to the gesture visually, in a way that is not necessary for vocalizations broadcast indiscriminately. For these reasons and others, human cooperative communication most likely began in the gestural modality. An especially interesting and important gesture is pointing, which apes do not do for one another, but only for humans and only in one of its functions (requesting). Human infants use the pointing gesture spontaneously for at least three different functions from before language begins, two of them purely cooperative (sharing emotions and providing others with needed information). It is argued that the pointing gesture embodies many aspects of the human adaptation for cooperative interactions involving shared intentionality - and so it is the best candidate we have for an immediate precursor to human language.

PREHISTORIC HANDEDNESS: SOME HARD EVIDENCE

NATALIE UOMINI

Department of Archaeology, University of Southampton, School of Humanities, Southampton, SO17 1BJ, UK

It is often stated in language evolution research that right-handedness is connected to the emergence of language in the hominin lineage. Most often invoked is the linking mechanism of cerebral asymmetry, but the precise nature of this relationship is rarely specified. In the context of human evolution, I define the term *handedness* as a species-level tendency to coordinate the right and left hands in a consistent manner, not only individually but at a population level. Any archaeological evidence that bears on prehistoric handedness should provide indirect information about prehistoric brain structure and function. Handedness evolution can be traced in several ways, but the most direct evidence lies in the archaeological record (the skeletal and material cultural data have already been extensively reviewed by Steele (2000) and Steele & Uomini, in press).

This paper will present a concise and structured summary of the archaeological data for right- and left-handedness in hominins, including *Homo heidelbergensis*, Neanderthals, and living humans, with a special focus on hard-hammer and soft-hammer direct percussion on stone (i.e. knapping). We will include results from our own analyses of lithic material from the archaeological site of Boxgrove (UK), combined with an experimental study of knapping gestures which relates the observed laterality markers to the gestures that created them.

In order to reconcile at a theoretical level the archaeological evidence for handedness with laterality in language and the brain, we characterise stone knapping as a *skilled bimanual task*. In this context, *skilled* refers to motor learning, in that it involves neuronal reorganisation of motor cortex (Karni et al., 1998). We characterise this task in terms of the Frame/Contents model of handedness (MacNeilage 1986, Guiard, 1987). In this model, one upper limb performs movements which Guiard (1987) qualifies as high-frequency,

being more spatially and temporally precise, whereas the other hand is low-frequency, acting as a stabiliser or support. The uniquely human aspect of handedness is our tendency, for skilled tasks, to learn the frames with the left hand and the contents with the right hand. We attempt to integrate this model for stone knapping into Wray's (1992; 2002) Focusing Hypothesis of asymmetrical language functions, in which the right hemisphere manipulates the holistic or spatial elements, and the left hemisphere the analytical or sequential elements.

With this paper we hope to raise awareness of the archaeological evidence for handedness, which can give clues to the timing of the emergence of a potential marker for language.

References

Guiard, Y. (1987). Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behavior, 19(4)*, 486-517.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M.M., Turner, R., & Ungerleider, L.G. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences USA, 95*, 861-868.

Macneilage, P.F. (1986). Bimanual coordination and the beginnings of speech. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 189–204). New York: Stockton Press.

Macneilage, P.F. (1998). The frame/content theory of evolution of speech production. *Behavioral & Brain Sciences, 21(4)*, 499-511.

Steele, J., & Uomini, N. (in press). Humans, tools and handedness. In V. Roux & B. Bril (Eds.), *Stone Knapping: the necessary conditions for a uniquely hominid behaviour* (pp. 215–238). Cambridge: McDonald Institute Monograph series.

Steele, J. (2000). Handedness in past human populations: skeletal markers. *Laterality, 5(3)*, 193-220.

Wray, A. (1992). *The focusing hypothesis: the theory of left hemisphere lateralised language re-examined.* Amsterdam/ Philadelphia: John Benjamins.

Wray, A. (2002). Dual processing in protolanguage: perfomance without competence. In A. Wray (Ed.), *The transition to language* (pp. 113-137). Oxford: Oxford University Press.

LATERALIZATION OF INTENTIONNAL GESTURES IN NON HUMAN PRIMATES: BABOONS COMMUNICATE WITH THEIR RIGHT HAND

JACQUES VAUCLAIR AND ADRIEN MEGUERDITCHIAN

Center for Research in Psychology Cognition, Language & Emotion, Department of Psychology, University of Provence, 13621 Aix-en-Provence, France.

Comparative studies of nonhuman and human primates concerning intentional communicative gestures know a renewed interest regarding the evolution of communicatory systems, in particular language. Although gestures are true means of communication among groups of non human primates (e.g., Tomasello & al., 1997), they have been relatively little studied compared to vocalizations and facial expressions. Whether these communicative behaviours involved lateralized systems is still unclear. Humans are mainly right-handed for many actions including manual gesturing and such asymmetries are linked to a left cerebral hemispheric dominance for the perception and the production of language (Knecht & al., 2000). Thus, the study of communicative gestures and their asymmetries in nonhuman primates constitutes an ideal framework to clarify the hypothesis of the gestural origin of the language and its lateralization (Corballis, 2002).

Some investigations in manual gestural communication by humans reveal links between handedness and hemispheric specialisation for language. Firstly, it has been shown that the activity of the right-hand is predominant for manual movements when people are talking and for signing by deaf humans with left-hemispheric dominance for the control of sign language functions. Secondly, the degree of right-hand asymmetries for manual communication such as "pointing" increases during the development of speech in young children. Additionally, the use of right-hand is more pronounced for signing than for non communicative motor actions among children of deaf parents (see Vauclair 2004, for a review). Concerning nonhuman primates, studies have only concerned captive chimpanzees (*Pan troglodytes*) and have also shown populations-level right-handedness for communicative gestures (Hopkins & al., 1998), a bias which is stronger than the bias exhibited in manipulative tasks (Hopkins & al., 2005). Thus, such continuity between humans and pongidae support the view that lateralization for language may have evolved from a gestural system of communication lateralized in the left hemisphere, in the common ancestor as recently as 5 or 6 million years ago.

To our knowledge, no investigation has been undertaken in monkeys. Our research aims thus at describing several intentional communicative gestures and their lateralization in baboons (*Papio anubis*). One of these gestures is "the threat gesture" which is part of the specific gestural repertoire, as well as other gestures induced by humans, such as "requests" and "pointings". Hand preferences for these gestures were assessed by observing interactions between conspecifics and between a baboon and a human observer. The results showed significant population-level right-handedness for communicative gestures in the baboon. Moreover, these biases were stronger than those exhibited in a non communicative motor actions (uni- and bimanual manipulative tasks: Vauclair & al., 2005). The results will be discussed within a comparative and a speculative context regarding the evolution of language and its cerebral lateralization.

References

Corballis, M. C. (2002). *From Hand to Mouth. The Origins of Language.* Princeton, NJ : Princeton University Press.

Hopkins, W. D., & Leavens, D. A. (1998). Hand use and gestural communication in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 112,* 95-99.

Hopkins, W. D., Russel, J., Freeman, H., Buehler, N., Reynolds, E., & Schapiro, S. J. (2005). The distribution and development of handedness for manual gestures in captive chimpanzees (*Pan troglodytes). Psychological Science, 6,* 487-493.

Knecht, S., Deppe, M., Draeger, B., Bobe, L., Lohman, H., Ringelstein, E. B., & Henningsen, H. (2000). Language lateralization in healhty right-handers. *Brain, 123,* 74-81.

Tomasello, M., & Camaioni, L. (1997). A comparison of the gestural communication of apes and human infants. *Human Development, 40,* 7-24.

Vauclair, J. (2004). Lateralization of communicative signals in nonhuman primates and the hypothesis of the gestural origin of language. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems, 5,* 363-384.

Vauclair, J., Meguerditchian, A., & Hopkins, W. D. (2005). Hand preferences for unimanual and coordinated bimanual tasks in baboons *(Papio anubis). Cognitive Brain Research, 25,* 210-216.

EMERGENCE OF GRAMMAR AS REVEALED BY VISUAL IMPRINTING IN NEWLY-HATCHED CHICKS

ELISABETTA VERSACE
Department of Psychology, University of Trieste, Via S. Anastasio 12, Trieste, 34123, Italy

LUCIA REGOLIN
Department of General Psychology, University of Padua, Via Venezia 8, Padova, 35131, Italy

GIORGIO VALLORTIGARA
Department of Psychology and B.R.A.I.N Centre for Neuroscience, University of Trieste, Via S. Anastasio 12, Trieste, 34123, Italy

1. Introduction

We investigated possible precursors of grammar in a non-human species, the domestic chick (*Gallus gallus*), using filial imprinting as an experimental tool. This procedure is comparable to the habituation-dishabituation techniques, which assess whether subjects that do not possess language recognize and respond to unexpected change in an object/event (Vallortigara, 2006).

Newly-hatched chicks were imprinted by exposing them to visual stimuli whose components were arranged according to an $(AB)^n$, or to an $(A)^n(B)^n$ (Fitch & Hauser, 2004) or to an (A(BB)A) grammar. At test, chicks could associate, i.e. approach, either a stimulus whose components were arranged according to the familiar grammar or a stimulus whose components were arranged according to an unfamiliar grammar.

2. Material and methods

From day 1 to day 3 of life, chicks were exposed to an imprinting stimulus composed by several simultaneously presented units (3x5cm) whose colours were arranged according to an $(AB)^n$ or to an $(A)^n(B)^n$ or to an (A(BB)A) grammar – different letters indicate different colours: blue, green, red, and yellow.

On day 4, chicks were individually tested presenting them with a stimulus composed of a familiar grammar (the grammar of the imprinting stimulus) and a stimulus composed of an unfamiliar grammar, located at the opposite ends of a runway (72x30x25cm).

In **Experiment 1** chicks were either imprinted with an ABABAB stimulus and then tested with CDCDCD *vs.* CDDCDC stimuli, or were imprinted with an AAABBB stimulus and then tested with CCCDDD *vs.* CDDCDC stimuli.

In **Experiment 2** chicks were imprinted with an ABAB or ABBA stimulus and then tested with CDCD *vs.* CDDC stimuli.

Time spent close to the familiar and unfamiliar stimulus was recorded for each chick for 6 consecutive minutes and then computed for each chick and for each minute of the test, as in Eq.(1):

Time close to familiar/(Total time spent close to the two stimuli)x100 Eq.(1)

Departures from chance level (50%,) indicated preferences for the familiar (>50%) or the unfamiliar (<50%) stimulus, and were estimated by two-tailed one-sample t-test. An ANOVA was performed on the data following log transformation to account for any heterogeneity of variances with type of imprinting stimulus as between-subject factor, and time (from the first to the sixth minute) as a within-subject factor.

3. Results and discussion

In **Experiment 1** the ANOVA revealed a significant main effect of time ($F_{5,540}$= 4.92; $p<0.01$) but not of the type of imprinting stimulus ($F_{1,108}=1.81$; $p=0.18$) nor of the time x stimulus interaction ($F_{5,\ 540}=0.44$; $p=0.82$). One sample t-tests showed that chicks preferred the unfamiliar stimulus during the overall testing period (Mean= 45.21; SE= 2.46; $t_{111}=-3.77$; $p<0.01$). In **Experiment 2** the ANOVA revealed a significant main effect of time ($F_{5,\ 965}=6.20$, $p<0.01$), but not of the type of imprinting stimulus ($F_{1,\ 193}=0.55$; $p=0.46$) nor of any interaction ($F_{5,965}=0.41$, $p=0.84$). Again, chicks preferred the unfamiliar stimulus during the overall testing period (Mean= 45.67; SE=2.17 $t_{196}=-4.24$; $p<0.01$).

Differently from evidence obtained in cotton-top tamarins (Fitch & Hauser, 2004), young chicks appear to be able to process either the $(AB)^n$ and the $(A)^n(B)^n$ structures and the (A(BB)A) structure too. Further research is warranted to investigate the capability of generalization, the role of sequential stimuli and more sophisticated grammar rules, especially of stimuli whose grammar can be unambiguously interpreted.

References

Fitch, W. T. & Hauser, M. D. (2004). Computational Constraints on Syntactic Processing in a Nonhuman Primate. *Science* 303, 377-380.

Vallortigara, G. (2006) The Cognitive Chicken: Visual and Spatial Cognition in a Non-Mammalian Brain. In: E.A. Wasserman and T.R. Zentall, (Eds.) *Comparative Cognition: Experimental Explorations of Animal Intelligence*, Oxford: Oxford University Press, *in press*.

BEYOND THE ARGUMENT FROM DESIGN

WILLEM ZUIDEMA

Institute for Logic, Language and Computation, University of Amsterdam, Plantage Muidergracht 24, 1018 HG, Amsterdam, the Netherlands
jzuidema@science.uva.nl

TIMOTHY O'DONNELL

Primate Cognitive Neuroscience Laboratory, Harvard University, 33 Kirkland Street, Cambridge MA 02138, U.S.A.
timo@wjh.harvard.edu

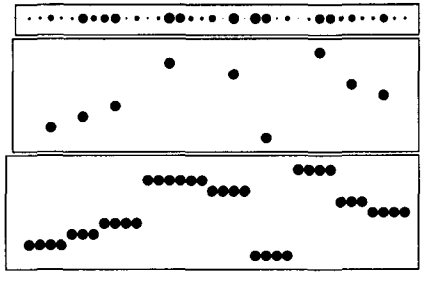
Many studies of the evolutionary origins of human language capabilities rely on what is sometimes called the "Argument from Design". Such studies attempt to establish that a given feature of that capacity is (i) too complex to have arisen by chance, and (ii) appears to be specifically designed for processing natural languages. It is argued that the theory of natural selection is the only scientific theory that can explain the appearance of complex, adaptive design, and, hence, that the conclusion that the feature evolved as an adaptation for language is unavoidable.

We will not, at this point, address the many disagreements about the linguistic data used in such studies, or questions about whether or not given processing abilities are specific for language, or about whether or not objective measures for complexity exist. Rather, we analyze the validity of reasoning with the argument from design when studying culturally transmitted systems such as natural language or music. We show that in these systems such reasoning is unsound, because there exists an alternative scientific explanation for the appearance of design that can be termed "cultural evolution".

As a simple example, consider the evidence reviewed in Pinker and Jackendoff (2005) showing that other primates, including chimpanzees, have difficulties distinguishing human phonemes and/or make phoneme boundaries differently from humans. Pinker & Jackendoff conclude that human speech perception is special, and must therefore, they imply, be adapted for language in the biological sense. However, it is easy to show – as we do in figure 1 using a variant of the model from Zuidema and Westermann (2003) – that if a language is transmitted and negotiated culturally, and allowed to change based on success and failure in recognition, any arbitrary features of the perceptual system will be reflected in the configuration of

signals. This suggests an alternative explanation for the fact that *humans* are much better than other species at recognizing *human* phonemes: human languages have evolved so as to exploit the accidental peaks in human auditory perception.

In our talk, we will look in detail at two other proposed adaptations, concerning compositional semantics and phrasal syntax, and summarize results from simulations studied by ourselves and others (e.g. Kirby, 1994). In all cases, we find that human languages can evolve to match idiosyncratic features of human language processing, giving humans the appearance of being designed for language without them having adapted in the biological sense. Hence, every time we observe the appearance of design for language, we need to ask: did it result from cultural or from biological adaptation? One important route for distinguishing between the two hypotheses is via falsification of the biological adaptation hypothesis by showing similar biases in animals. A second route, supporting the latter hypothesis, is via an optimality- (or game-) theoretic analysis showing that languages adapted to human biases are superior to languages adapted to non-human biases. We will present examples of both types of evidence, and conclude that language evolution research can and should move beyond the argument from design.



Legend: The top frame (auditory perception) shows for each of 36 possible signals, the randomly chosen probabilities of correct recognition. The middle frame (production) shows for each of 9 possible meanings (vertical axis), which signal (horizontal axis) is used to express it. The bottom frame (interpretation) shows for each of 36 possible signals (horizontal), which of 9 possible meanings (vertical) is chosen as its interpretation.

Figure 1. Through cultural evolution, languages emerge that reflect arbitrary features of the auditory perception. Shown are results from a simulation (a variant of the model described in detail in Zuidema & Westermann, 2003) where individuals, with given perceptual characteristics (top frame) learn their language (middle and bottom frame) from each other. The result of the simulation gives the appearance of design: the characteristics of perception are such that the signals used to express each possible meaning (middle frame) are all among the most reliably recognised signals (top frame). However, there has only been cultural adaptation: the language evolved to exploit the peaks in auditory perception.

References

Kirby, S. (1994). Adaptive explanations for language universals. *Sprachtypologie und Universalienforshung*, *47*, 186-210.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it. *Cognition*, *95*(2), 201-236.

Zuidema, W., & Westermann, G. (2003). Evolution of an optimal lexicon under constraints from embodiment. *Artificial Life*, *9*(4), 387-402.

Author Index

Arbib MA, 3
Arnold K, 389
Au C-P, 391

Baronchelli A, 11
Barrat A, 11
Bartlett M, 393
Belpaeme T, 395
Bergen BK, 35
Bleys J, 395
Bonaiuto J, 3
Bowie J, 397
Briscoe T, 19
Bryson JJ, 399
Byrne R, 401

Cartmill E, 401
Cavalli Sforza L, 255
Chater N, 27
Cheney D, 444
Christiansen MH, 27, 333, 430
Coupé C, 419
Cristianini N, 348
Crow TJ, 403

Dall'Asta L, 11
de Beule J, 35
de Boer B, 405
de Jager ST, 407
de Pauw G, 43
Di Chio C, 51
Di Chio P, 51
Dediu D, 59
Delvaux V, 67
Demolin D, 67
Dessalles J-L, 75
Dowman M, 83

Fitch WT, 409
Fontanari JF, 411, 438

Galantucci B, 413
Gil D, 91
Gong T, 99, 206, 419
Gontier N, 107
Griffiths TL, 83

Hashimoto T, 415
Hawkey DJC, 417
Hinzen W, 115
Hoefler S, 123
Hurford JR, 131

Jäger G, 139
Jeffreys M, 145
Johansson S, 152, 160

Karnik H, 222
Kazakov D, 393
Ke J, 419
Kirby S, 83, 283, 421
Knight C, 168
Kroos C, 413

Landsbergen F, 423
Lanyon SJ, 176
Liebal K, 267
Locke JL, 184
Loreto V, 11
Lupyan G, 190

Marocco D, 198
Meguerditchian A, 455
Merker B, 440
Minett JW, 99, 206
Mirolli M, 214
Mithen S, 425
Mitri S, 428
Mittal S, 222
Monaghan P, 430

Nakatsuka M, 415
Nasidze I, 432
Newmeyer FJ, 434
Nolfi S, 198

O'Donnell T, 459
Okanoya K, 440
Oudeyer P-Y, 436

Parisi D, 214, 230
Parker AR, 239
Perlovksy LI, 411, 438
Philps D, 247
Piazza A, 255
Pika S, 267
Poulshock J, 275

Reali F, 27
Regolin L, 457
Rhodes T, 413
Ritchie G, 283
Rosta E, 3

Sasahara K, 440
Schulz R, 291
Scott-Phillips T, 299
SEDSU Project, The 379
Seyfarth R, 442
Slocombe K, 443
Smith ADM, 307
Smith DW, 445
Smith K, 315
Steels L, 323
Sternberg DA, 333
Stockwell P, 291
Stoneking M, 432

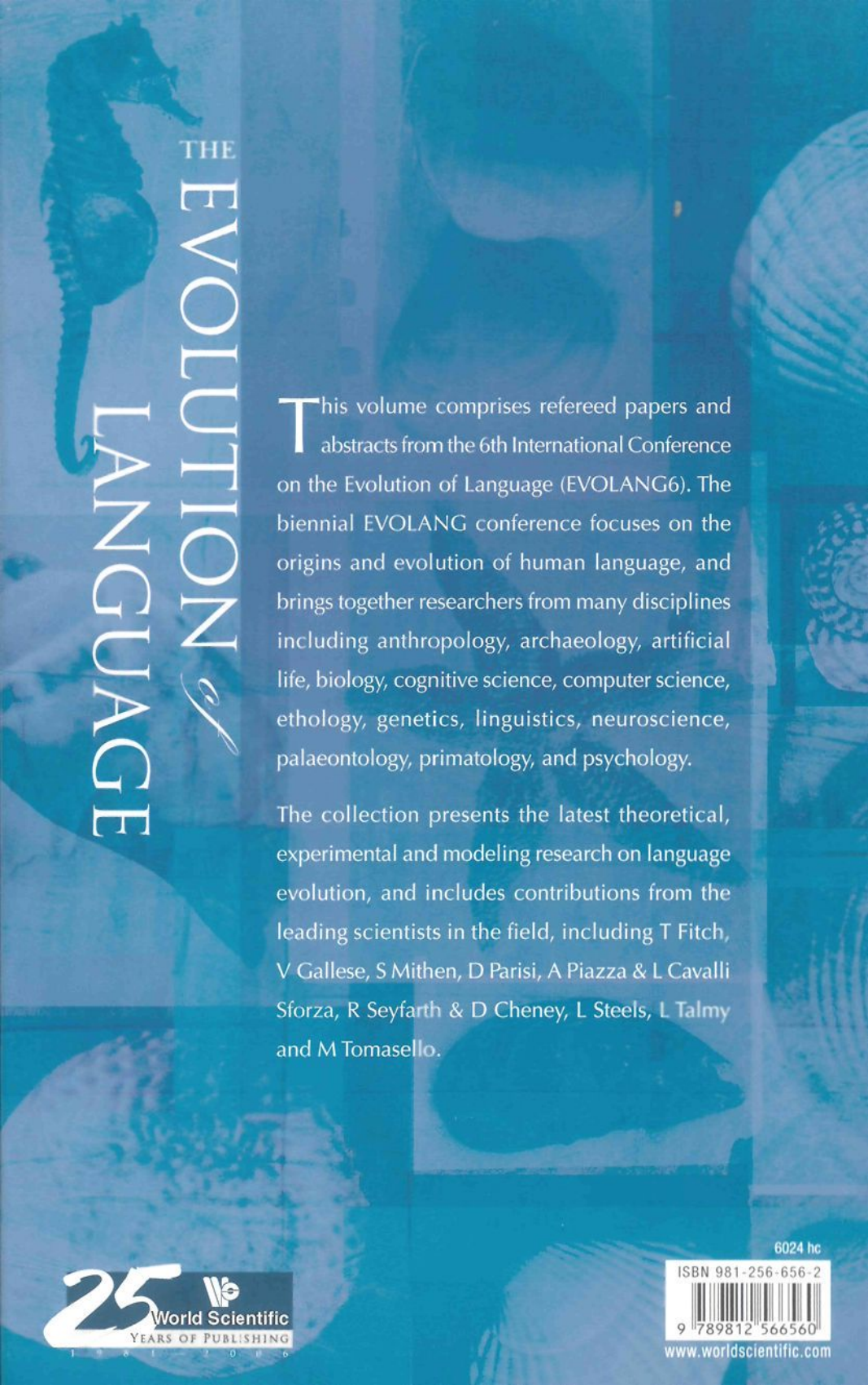
Tallerman M, 447
Talmy L, 449
Tamariz M, 341
Tomasello M, 452
Turchi M, 348

Uomini N, 453

Vallortigara G, 457
van Rooij R, 356
Vauclair J, 455
Versace E, 457
Vogt P, 364, 428

Wakabayashi M, 291
Wang WS-Y, 99, 206
Wiles J, 291

Zeevat H, 372
Zlatev J, 379
Zuberbühler K, 389, 443
Zuidema W, 459



THE EVOLUTION of LANGUAGE

This volume comprises refereed papers and abstracts from the 6th International Conference on the Evolution of Language (EVOLANG6). The biennial EVOLANG conference focuses on the origins and evolution of human language, and brings together researchers from many disciplines including anthropology, archaeology, artificial life, biology, cognitive science, computer science, ethology, genetics, linguistics, neuroscience, palaeontology, primatology, and psychology.

The collection presents the latest theoretical, experimental and modeling research on language evolution, and includes contributions from the leading scientists in the field, including T Fitch, V Gallese, S Mithen, D Parisi, A Piazza & L Cavalli Sforza, R Seyfarth & D Cheney, L Steels, L Talmy and M Tomasello.

6024 hc
ISBN 981-256-656-2
9 789812 566560
www.worldscientific.com

25 World Scientific
YEARS OF PUBLISHING
1981 2006