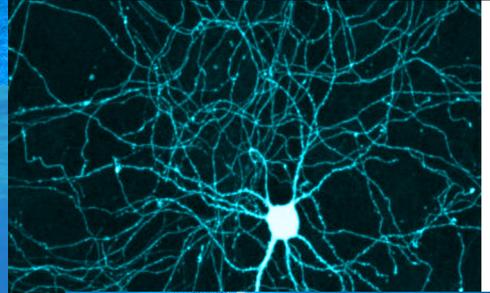
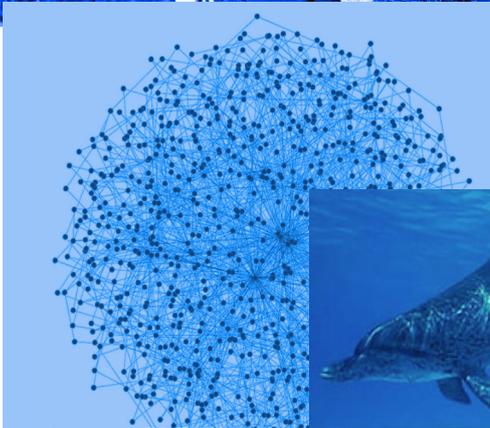


THE EVOLUTION *of*

Proceedings of the
11th International
Conference
New Orleans, LA
March 21-24 2016

LANGUAGE



Editors

Seán Roberts
Christine Cuskley
Luke McCrohon
Lluís Barceló-Coblijn
Olga Fehér
Tessa Verhoef



The Evolution of Language: Proceedings of the 11th International Conference.
New Orleans, USA.

Cover image credits: <https://www.flickr.com/photos/sjcockell/4684828794> :
Simon Cockell

<https://www.flickr.com/photos/69221340@N04/8652893210> : Papooga

<https://www.flickr.com/photos/jdebberly/2850385433> : Jay Ebberly

<https://www.flickr.com/photos/pustovit/14336071644> : Vladimir Pustovit

<https://www.flickr.com/photos/kittysfotos/10037699755> : Kitty Terwolbeck

<https://www.flickr.com/photos/ajmat/3730694413> : AJ.Mat

<https://www.flickr.com/photos/neilspicys/2348972925> : NeilsPhotography

<https://www.flickr.com/photos/kapten/1085243998> : Björn Söderqvist

<https://www.flickr.com/photos/thomasclaveirole/463202335> : Thomas Claveirole

<https://www.flickr.com/photos/deanwissing/3565255880> : Dean Wissing

<https://www.flickr.com/photos/mikeblogs/3101400087> : Mike Seyfang

Preface

This volume collects the refereed papers and abstracts of the 11th International Conference on the Evolution of Language (EVOLANG XI), held in New Orleans on 21st-24th March, 2016. Submissions to the conference were solicited in two forms, papers and abstracts, and this is reflected in the structure of this volume.

The biennial EVOLANG conference is characterised by an invigorating, multi-disciplinary approach to the origins and evolution of human language, and brings together researchers from many fields including anthropology, archaeology, artificial life, biology, cognitive science, computer science, ethology, genetics, linguistics, neuroscience, palaeontology, primatology, psychology and statistical physics. The multi-disciplinary nature of the field makes the refereeing process for EVOLANG very challenging, and we are indebted to our panel of reviewers for their very conscientious and valuable efforts.

For the first time, the proceedings of EvoLang XI are primarily available online in an open access format. Please visit <http://evolang.org/neworleans/> for up-to-date papers, workshop papers and supplementary materials. Thanks are due to the following people:

The EVOLANG committee:

- Andreas Baumann (University of Vienna),
- Rudolf Botha (Stellenbosch University),
- Christine Cuskley (Institute for Scientific Interchange),
- Erica Cartmill (University of California, Los Angeles) [First Contact/Co-chair],
- Jean-Louis Dessalles (Ecole Nationale Supérieure des Télécommunications, Paris),
- Ramon Ferrer i Cancho (Polytechnic University of Barcelona),
- Tecumseh Fitch (University of Vienna),

- Jim Hurford (University of Edinburgh),
- Simon Kirby (University of Edinburgh) [Co-chair],
- Chris Knight (University of East London),
- Heidi Lyn (University of Southern Mississippi),
- Luke McCrohon (University of Tokyo) [Webmaster],
- Kazuo Okanoya (University of Tokyo),
- Thom Scott-Phillips (Durham University),
- Andrea Ravignani (University of Vienna, University of Edinburgh),
- Nikolaus Ritt (University of Vienna),
- Kenny Smith (University of Edinburgh) [Treasurer],
- Maggie Tallerman (Newcastle University, UK),
- Natalie Uomini (University of Liverpool).

The local organising committee:

University of Southern Mississippi: Heidi Lyn (Chair), Megan Broadway, Myстера Samuelson, Jennie Christopher, and Beatrice Chenkin.

University of South Alabama: Stephanie Jett.

Tulane University: Olanike-Ola Orie.

Louisiana State University: XiaoChing Li

The workshop convenors: Christine Cuskley, Vittorio Loreto, Francesca Tria, Vito D.P. Servedio, Andrea Ravignani, Erin E. Hannon, Seán Roberts, Gregory J. Mills, Madza Farias-Virgens, Dillon Niederhut, Terrence Deacon, Tecumseh Fitch, Bart de Boer and Takeshi Nishimura.

The plenary speakers: Sharon Thompson-Schill, Thom Scott-Phillips, Ljiljana Progovac, Richard Moore, Erich Jarvis, Vincent Janik, Evelina Fedorenko, Dean Falk, Joan Bybee.

Finally, and most importantly, the authors of all the contributions collected here.

Seán Roberts, Christine Cuskley, Luke McCrohon,
Lluís Barceló-Coblijn, Olga Feher and Tessa Verhoef.
Scientific Committee,
February 2016

Reviewers

Alan Nielsen	Bill Thompson	Daniel Dor
Albert Naccache	Bodo Winter	Daniel Lenzen
Aleksandrs Berdicevskis	Brady Clark	Dankmar Enke
Alex Carstensen	Camilla Power	David Leavens
Alfredo Ardila	Carel ten Cate	David Lightfoot
Andrea Ravnani	Carl Vogel	Dillon Niederhut
Andreas Baumann	Carlos Santana	Elizabeth Irvine
Andreea Geambasu	Carmen Saldana	Erica Cartmill
Andrew Feeney	Catriona Silvey	Estefania Santacreu-Vasut
Andrew Wedel	Cedric Boeckx	Federica Cavicchio
Anne Marie Di Sciullo	Chris Knight	Freek Van de Velde
Anne van der Kant	Christian Abry	Fusa Katada
Anneliese Kuhle	Christina Behme	Gareth Roberts
Antoni Gomila	Christine Caldwell	Gary Lupyan
Antonio Benítez-Burraco	Christine Cuskley	Gerhard Jaeger
Aritz Irurtzun	Connie de Vos	Greg Bryant
Ashley Micklos	Cory M Cuthbertson	Gregory Mills
Barbora Skarabela	Damian Blasi	Hannah Little
Bart de Boer	Dan Dediu	Harald Hammarström
		Heidi Lyn
		Hope Morgan

Irene Pepperberg	Ljiljana Progovac	Michelle Spierings
Jérôme Michaud	Lluís Barceló-Coblijn	Mieko Ogura
James Kirby	Luke Fleming	Molly Flaherty
James Steele	Luke McCrohon	Molly Lewis
James Winters	Madza Farias-Virgens	Monica Tamariz
Jasmeen Kanwal	Maggie Tallerman	Natalie Uomini
Jean-Louis Dessalles	Marcus Perlman	Nicolas Fay
Jennifer Culbertson	Marie Coppola	Nikolas Gisborne
Jeremy Collins	Marie Montant	Nikolaus Ritt
Joan Bybee	Marieke Schouwstra	Olga Feher
Joana Rosselló	Marieke Woensdregt	Olga Vasileva
Jon Carr	Marion Laporte	Olivier Morin
Jordi Arranz	Mark Atkinson	Paul Vogt
Juan Carlos Moreno Cabrera	Mark Bartlett	Peeter Tinitis
Kang-Suk Byun	Mark Dingemanse	Phillip Alday
Katie Slocombe	Mark Ellison	Phillip M. Alday
Kazuo Okanoya	Maryia Fedzechkina	Przemysław Żywicznyński
Keelin Murray	Matt Cartmill	Ramon Ferrer I Cancho
Kenny Smith	Matt Spike	Remi Van Trijp
Kensy Cooperrider	Matthew Lou-Magnuson	Richard Blythe
Kenta Suzuki	Maurício Martins	Richard Moore
Kevin Stadler	Mauricio Martins	Rick Janssen
Klaas Seinhorst	Michael Arbib	Rie Asano
Kobin Kendrick	Michael Franke	Robert Berwick
Koji Fujita	Michael Pleyer	Robert Truswell
	Michael Spranger	Roland Mühlenbernd
		Rory Turnbull

Rose Stamp	Sergio Balari	Tecumseh Fitch
Roslyn Frank	Simon D. Levy	Ted Bayne
Ruth Sonnweber	Simon Kirby	Teresa Bejarano
Ryan Lopic	Simon Levy	Tessa Verhoef
Sławomir Waciewicz	Simon Townsend	Thom Scott-Phillips
Sabine van der Ham	Sonia Ragir	Till Bergmann
Sabrina Engesser	Stefan Hartmann	Timo Ahlers
Samarth Swarup	Takashi Hashimoto	Tom Sawallis
Savithry Namboodiripad	Takeshi Nishimura	Vanessa Ferdinand
Seán Roberts	Tao Gong	Willem Zuidema

Contents

Preface	3
Program committee	19
Papers	19
Brain Mechanisms Of Human Acoustic Communication: A Phylogenetic Approach And Its Ontogenetic Implications, <i>Hermann Ackermann, Wolfram Ziegler</i>	20
Towards A Rigorous Motivation For Zipf's Law, <i>Phillip M. Alday</i>	22
Evolution Towards An Optimal Management Of Linguistic Information, <i>Lluís Barcelo-Coblijn</i>	31
A Lotka-volterra Model Of The Evolutionary Dynamics Of Compositionality Markers, <i>Andreas Baumann, Christina Prömer, Kamil Kazmierski, Nikolaus Ritt</i>	40
The Antiquity Of Musicality And Its Role In Prehistoric Culture, <i>Ted Bayne</i>	49
The Low-complexity-belt: Evidence For Large-scale Language Contact In Human Prehistory?, <i>Christian Bentz</i>	59
Redundant Features Are Less Likely To Survive: Empirical Evidence From The Slavic Languages, <i>Aleksandrs Berdicevskis, Hanne Eckhoff</i>	70
A Scientometric Analysis Of Evolang: Intersections And Authorships, <i>Till Bergmann, Rick Dale</i>	79
Using Causal Inference To Detect Directional Tendencies In Semantic Evolution, <i>Johannes Dellert</i>	88
The Emergence Of The Progressive To Imperfective Diachronic Cycle In Reinforcement-learning Agents, <i>Dankmar Enke, Roland Mühlendernd, Igor Yanovich</i>	97

Kauffman's Adjacent Possible In Word Order Evolution, <i>Ramon Ferrer-I-Cancho</i>	106
Phoneme Inventory Size Distributions And The Origins Of The Duality Of Patterning, <i>Luke Fleming</i>	115
Migration As A Window Into The Coevolution Between Language And Behavior, <i>Victor Gay, Daniel Hicks, Estefania Santacreu-Vasut</i>	124
The Impact Of Communicative Network Structure On The Conventionalization Of Referring Expressions In Gesture, <i>Matt Hall, Russell Richie, Marie Coppola</i>	132
Deictic Tools Can Limit The Emergence Of Referential Symbol Systems, <i>Elizabeth Irvine, Sean Roberts</i>	140
Inferring The World Tree Of Languages From Word Lists, <i>Gerhard Jaeger, Soeren Wichmann</i>	149
Protolanguage Possibilities In A Construction Grammar Framework, <i>Sverker Johansson</i>	157
Evolution Of The Language-ready Brain: Warfare Or 'mother Tongues'?, <i>Chris Knight, Camilla Power</i>	163
The Emergence Of Argument Marking, <i>Sander Lestrade</i>	171
Differing Signal-meaning Dimensionalities Facilitates The Emergence Of Structure, <i>Hannah Little, Kerem Eryilmaz, Bart de Boer</i>	182
Correlated Evolution Or Not? Phylogenetic Linguistics With Syntactic, Cognacy, And Phonetic Data, <i>Giuseppe Longobardi, Armin Buch, Andrea Ceolin, Aaron Ecaj, Cristina Guardiano, Monica Irimia, Dimitris Michelioudakis, Nina Radkevich, Gerhard Jaeger</i>	190
Linking The Processes Of Language Evolution And Language Change: A Five-level Hierarchy, <i>Jérôme Michaud</i>	199
Interaction For Facilitating Conventionalization: Negotiating The Silent Gesture Communication Of Noun-verb Pairs, <i>Ashley Micklos</i>	209
How Selection For Language Could Distort The Dynamics Of Human Evolution, <i>William Mitchener</i>	217
Frequency-dependent Regularization In Iterated Learning, <i>Emily Morgan, Roger Levy</i>	227

Self-organization In Sound Systems: A Model Of Sound Strings Processing Agents, <i>Roland Mühlenbernd, Johannes Wahle</i>	236
Edward Sapir And The Origin Of Language, <i>Albert Naccache</i>	245
Quantifying The Semantic Value Of Words, <i>Dillon Niederhut</i>	254
Semantic Approximation And Its Effect On The Development Of Lexical Conventions, <i>Bill Noble, Raquel Fernández</i>	262
The Evolution Of Im/politeness, <i>Monika Pleyer, Michael Pleyer</i>	271
What Kind Of Grammar Did Early Humans (and Neanderthals) Command? A Linguistic Reconstruction, <i>Ljiljana Progovac</i>	280
Active Control Of Complexity Growth In Naming Games: Hearer’s Choice, <i>William Schueller, Pierre-Yves Oudeyer</i>	288
Mind The Gap: Inductive Biases In Phonological Feature Learning, <i>Klaas Seinhorst</i>	297
Against The Emergent View Of Language Evolution, <i>Maggie Tallerman</i>	304
Learning To Learn From Similar Others: Approximate Bayesian Computation Through Babbling, <i>Bill Thompson, Heikki Rasilo</i>	312
Genetic Drift Explains Sapir’s “drift” In Semantic Change, <i>Igor Yanovich</i>	321
Abstracts	330
Triadic Ontogenetic Ritualization: An Overlooked Possibility, <i>Ekaterina Abramova</i>	331
Pre And Post Partum Whistle Production Of A Bottlenose Dolphin (<i>Tursiops Truncatus</i>) Mother-calf Dyad, <i>Audra Ames, Sara Wielandt, Dianne Cameron, Stan Kuczaj</i>	334
Noise In Phonology Affects Encoding Strategies In Morphology, <i>David Ardell, Noelle Anderson, Bodo Winter</i>	337
Evolution Of Language From The Aphasia Perspective, <i>Alfredo Ardila</i>	339
Towards An Action-oriented Approach To The Evolution Of Language And Music, <i>Rie Asano</i>	342

On A Music-ready Brain: Neural Basis, Mechanisms, And Their Contribution To The Language Evolution, <i>Rie Asano, Edward Ruoyang Shi</i>	345
Adult Language Learning And The Evolution Of Linguistic Complexity, <i>Mark Atkinson, Kenny Smith, Simon Kirby</i>	350
Evolution Of What?, <i>Christina Behme</i>	352
Spontaneous Dialect Formation In A Population Of Locally Aligning Agents, <i>Richard A. Blythe, Alistair H. Jones, Jessica Renton</i>	362
How The Brain Got Grammaticalized: Globularization And (self-)domestication, <i>Cedric Boeckx, Constantina Theofanopoulou, Antonio Benítez-Burraco</i>	365
Signature Whistles In An Introduction Context, <i>Megan Broadway, Jamie Klaus, Billie Serafin, Heidi Lyn</i>	368
How Do Laughter And Language Interact?, <i>Greg Bryant</i>	371
Cultural Evolution And Communication Yield Structured Languages In An Open-ended World, <i>Jon W. Carr, Kenny Smith, Hannah Cornish, Simon Kirby</i>	374
Lasting Impacts Of The Code Model On Primate Communication Research, <i>Erica Cartmill</i>	376
Are Emotional Displays An Evolutionary Precursor To Compositionality In Language?, <i>Federica Cavicchio, Livnat Leemor, Simone Shamay-Tsoory, Wendy Sandler</i>	379
Functionally Flexible Vocalizations In Wild Bonobos (<i>pan Pansicus</i>), <i>Zanna Clay, Jahmaira Archbold, Klaus Zuberbuhler</i>	382
Relationship Between Nonverbal Social Skills And Language Development, <i>Hélène Cochet, Richard Byrne</i>	385
Dwarf Mongooses Combine Meaningful Alarm Calls, <i>Katie Collier, Andrew N. Radford, Balthasar Bickel, Marta B. Manser, Simon W. Townsend</i>	388
Word Order Universals Reflect Cognitive Biases: Evidence From Silent Gesture, <i>Jennifer Culbertson, Simon Kirby, Marieke Schouwstra</i>	391
The Emergence Of Rules And Exceptions In A Population Of Interacting Agents, <i>Christine Cuskley, Vittorio Loreto</i>	394

The Evolution Of Collaborative Stories, <i>Christine Cuskley, Bernardo Monechi, Pietro Gravino, Vittorio Loreto</i>	397
Empirically Assessing Linguistic Ability With Stone Tools, <i>Cory Cuthbertson</i>	399
Anatomical Biasing Of Click Learning And Production: An MRI And 3d Palate Imaging Study, <i>Dan Dediu, Scott Moisik</i>	401
The Fidelity Of Iterated Vocal Imitation, <i>Pierce Edmiston, Marcus Perlman, Gary Lupyan</i>	403
Meaningful Call Combinations And Compositional Processing In A Social Bird, <i>Sabrina Engesser, Amanda R. Ridley, Simon W. Townsend</i>	405
Using HMMs To Attribute Structure To Artificial Languages, <i>Kerem Eryilmaz, Hannah Little, Bart de Boer</i>	407
Stick Or Switch: A Simple Selection Heuristic May Drive Adaptive Language Evolution, <i>Nicolas Fay, Shane Rogers</i>	410
Processing Preferences Shape Language Change, <i>Maryia Fedzechkina, Becky Chu, T. Florian Jaeger, John Trueswell</i>	412
Communicative Interaction Leads To The Elimination Of Unpredictable Variation, <i>Olga Feher, Kenny Smith, Elizabeth Wonnacott, Nikolaus Ritt</i>	415
Word Learners Regularize Synonyms And Homonyms Similarly, <i>Vanessa Ferdinand, Matt Spike</i>	417
Humans Recognize Vocal Expressions Of Emotional States Universally Across Species, <i>Piera Filippi, Jenna V. Congdon, John Hoang, Daniel Liu Bowling, Stephan Reber, Andrius Paukonis, Marisa Hoeschele, Sebastian Ocklenburg, Bart de Boer, Christopher B. Sturdy, Albert Newen, Onur Güntürkün</i>	419
Multimodal Processing Of Emotional Meanings: A Hypothesis On The Adaptive Value Of Prosody, <i>Piera Filippi, Sebastian Ocklenburg, Daniel Liu Bowling, Larissa Heege, Albert Newen, Onur Güntürkün, Bart de Boer</i>	422
Do Lab Attested Biases Predict The Structure Of A New Natural Language?, <i>Molly Flaherty, Katelyn Stangl, Susan Goldin-Meadow</i>	425

Cooperative Communication And Communication Styles In Bonobos And Chimpanzees In The Wild: Same Same But Different?, <i>Marlen Fröhlich, Paul H Kuchenbuch, Gudrun Müller, Barbara Fruth, Takeshi Furuichi, Roman M Wittig, Simone Pika</i>	427
Integration Or Disintegration?, <i>Koji Fujita, Haruka Fujita</i>	430
Effects Of Task-specific Variables On Auditory Artificial Grammar Learning And Generalization, <i>Andreea Geambasu, Michelle J. Spierings, Carel ten Cate, Clara C. Levelt</i>	433
Intentional Meaning Of Bonobo Gestures, <i>Kirsty Graham, Catherine Hobaiter, Richard Byrne</i>	435
Plain Simple Complex Structures: The Emergence Of Overspecification In An Iterated Learning Setup, <i>Stefan Hartmann, Peeter Tunits, Jonas Nölle, Thomas Hartmann, Michael Pleyer</i>	437
Language Origins In Light Of Neuro-atypical Cognition And Speech Profiles, <i>Wolfram Hinzen, Joana Rosselló</i>	440
Effort Vs. Robust Information Transfer In Language Evolution, <i>T. Florian Jaeger, Maryia Fedzechkina</i>	442
Simple Agents Are Able To Replicate Speech Sounds Using 3d Vocal Tract Model, <i>Rick Janssen, Dan Dediu, Scott Moisik</i>	445
Nonlinear Biases In Articulation Constrain The Design Space Of Language, <i>Rick Janssen, Bodo Winter, Dan Dediu, Scott Moisik, Sean Roberts</i>	448
Modeling Language Change Triggered By Language Shift, <i>Anna Jon-And, Elliot Aguilar</i>	451
The Evolution Of Zipf's Law Of Abbreviation, <i>Jasmeen Kanwal, Kenny Smith, Jennifer Culbertson, Simon Kirby</i>	454
The Spontaneous Emergence Of Linguistic Diversity In An Artificial Language, <i>Deborah Kerr, Kenny Smith</i>	457
A General Auditory Bias For Handling Speaker Variability In Speech? Evidence In Humans And Songbirds., <i>Buddhamas Kriengwatana, Paola Escudero, Anne Kerkhoven, Carel ten Cate</i>	459
Cumulative Vocal Cultures In Orangutans And Their Ontogenetic Origin, <i>Adriano Lameira, Jeremy Kendal, Marco Gamba</i>	461

Learnability Pressures Influence The Encoding Of Information Density In The Lexicon, <i>Molly Lewis, Michael C. Frank</i>	464
A Developmental Perspective On Language Origin: Children Are Old Hands At Gesture, <i>Casey Lister, Tiarn Burtenshaw, Nicolas Fay, Bradley Walker, Jeneva Ohan</i>	466
Emergence Of Signal Structure: Effects Of Duration Constraints, <i>Hannah Little, Kerem Eryilmaz, Bart de Boer</i>	468
The Evolution Of Redundancy In A Global Language, <i>Gary Lupyan, Justin Sulik</i>	470
Nonhuman Animals' Use Of Ostensive Cues In An Object Choice Task, <i>Heidi Lyn, Stephanie Jett, Megan Broadway, Mystera Samuelson</i>	472
Language Adapts To Signal Disruption In Interaction, <i>Vinicius Macuch Silva, Sean Roberts</i>	475
Biological Systems Of Interest To Researchers Of Cultural Evolution, <i>Luke Mccrohon</i>	478
Preliminary Results From A Computational Multi Agent Modelling Approach To Study Humpback Whale Song Cultural Transmission, <i>Michael Mcloughlin, Luca Lamoni, Ellen Garland, Simon Ingram, Alexis Kirke, Michael Noad, Luke Rendell, Eduardo Miranda</i>	480
Human-like Brain Specialization In Baboons: An Invo Anatomical MRI Study Of Language Areas Homologs In 96 Subjects, <i>Adrien Meguerditchian, Damien Marie, Konstantina Margiotoudi, Scott A. Love, Alice Bertello, Romain Lacoste, Muriel Roth, Bruno Nazarian, Jean-Luc Anton, Olivier Coulon</i>	483
The Evolution Of Repair: Evidence From Online Conversations, <i>Gregory Mills</i>	485
Arbitrary Hierarchy: A Precedent For Language?, <i>Dominic Mitchell</i>	487
Make New With Old: Human Language In Phylogenetically Ancient Brain Regions, <i>Marie Montant, Johannes Ziegler, Benny Briesemeister, Tila Brink, Bruno Wicker, Aurélie Ponz, Mireille Bonnard, Arthur Jacobs, Mario Braun</i>	490
The Effect Of Modality On Signal Space In Natural Languages, <i>Hope Morgan</i>	491
Linguistic Structure Emerges In The Cultural Evolution Of Artificial Sign Languages, <i>Yasamin Motamedi, Marieke Schouwstra, Kenny Smith, Simon Kirby</i>	493

A Social Dimension Of Language Evolution, <i>Albert Naccache</i>	496
Shared Basis For Language And Mathematics Revealed By Cross-domain Syntactic Priming, <i>Tomoya Nakai, Kazuo Okanoya</i>	499
Measuring Conventionalization In The Manual Modality, <i>Savithry Namboodiripad, Daniel Lenzen, Ryan Lopic, Tessa Verhoef</i>	502
The Arbitrariness Of The Sign Revisited: The Role Of Phonological Similarity, <i>Alan Nielsen, Dieuwke Hupkes, Simon Kirby, Kenny Smith</i>	505
Domestication And Evolution Of Signal Complexity In Finches, <i>Kazuo Okanoya</i>	507
Parrot "Phonological Regression": Expanding Our Understanding Of The Evolution Of Vocal Learning, <i>Irene M. Pepperberg, Katia Zilber-Izhar, Scott Smith</i>	509
Early Learned Words Are More Iconic, <i>Lynn Perry, Marcus Perlman, Gary Lupyan, Bodo Winter, Dominic Massaro</i>	511
Cooperative Communication: What Do Primates And Corvids Have To Tell?, <i>Simone Pika</i>	514
Construction Grammar For Apes, <i>Michael Pleyer, Stefan Hartmann</i>	517
The Cultural Evolution Of Structure In Music And Language, <i>Andrea Ravignani, Tania Delgado, Simon Kirby</i>	520
Strategies In Gesture And Sign For Demoting An Agent: Effects Of Language Community And Input, <i>Lilia Rissman, Laura Horton, Molly Flaherty, Marie Coppola, Annie Senghas, Diane Brentari, Susan Goldin-Meadow</i>	523
Social Biases Versus Efficient Communication: An Iterated Learning Study, <i>Gareth Roberts, Mariya Fedzechkina</i>	526
Vocal Learning And Homo Loquens, <i>Joana Rosselló</i>	528
The Cultural Evolution Of Complexity In Linguistic Structure, <i>Carmen Saldana, Simon Kirby, Kenny Smith</i>	530
Skepticism Towards Skepticism Towards Computer Simulation In Evolutionary Linguistics, <i>Carlos Santana</i>	533

From Natural Order To Convention In Silent Gesture, <i>Marieke Schouwstra, Kenny Smith, Simon Kirby</i>	536
Children's Production Of Determiners As A Test Case For Innate Syntactic Categories, <i>Catriona Silvey, Christos Christodoulopoulos</i>	539
Vocal Learning In Functionally Referential Chimpanzee Food Calls, <i>Katie Slocombe, Stuart Watson, Anne Schel, Claudia Wilke, Emma Wallace, Leveda Cheng, Victoria West, Simon Townsend</i>	541
Chimpanzees Process Structural Isomorphisms Across Sensory Modalities, <i>Ruth Sonnweber, Andrea Ravignani</i>	543
Rule Learning In Birds: Zebra Finches Generalize By Positional Similarities, Budgerigars By The Structural Rules., <i>Michelle Spierings, Carel ten Cate</i>	546
Information Dynamics Of Learned Signalling Games, <i>Matthew Spike, Simon Kirby, Kenny Smith</i>	548
Minimal Pressures Leading To Duality Of Patterning, <i>Matthew Spike, Kenny Smith, Simon Kirby</i>	550
Metalinguistic Awareness Of Trends As A Driving Force In Language Change: An Empirical Study, <i>Kevin Stadler, Elyse Jamieson, Kenny Smith, Simon Kirby</i>	552
The Grammar Of The Body And The Emergence Of Complexity In Sign Languages, <i>Rose Stamp, Wendy Sandler</i>	554
Failures Of Perspective Taking In An Open-ended Signaling Task, <i>Justin Sulik, Gary Lupyan</i>	556
What Is Unique About The Evolution Of Language Compared To Other Cultural Domains? An Experimental Study Of Language, Technology And Art, <i>Monica Tamariz, Jon W. Carr</i>	558
Evidence Of Descent With Modification And Selection In Iterated Learning Experiments, <i>Monica Tamariz, Joleana Shurley</i>	560
Interpreting Silent Gesture, <i>Bill Thompson, Marieke Schouwstra, Henriëtte de Swart</i>	563
Arbitrariness Of Iconicity: The Sources (and Forces) Of (dis)similarities In Iconic Representations, <i>Oksana Tkachman, Carla L. Hudson Kam</i>	566
Experimental Evidence For Phonemic-like Contrasts In A Nonhuman Vocal System, <i>Simon Townsend, Andrew Russell, Sabrina Engesser</i>	568

Modeling The Emergence Of Creole Languages, <i>Francesca Tria, Vittorio Loreto, Vito Servedio, S. Mufwene Salikoko</i>	570
Dendrophobia In Bonobo Comprehension Of Spoken English, <i>Robert Truswell</i>	574
A Constant Rate Effect Without Stable Functions, <i>Robert Truswell, Nikolas Gisborne</i>	576
Norms For Constructing Language In Humans And Animals, <i>Robert Ullrich</i>	578
Addressees Use Zipf's Law As A Cue For Semantics, <i>Freek Van de Velde, Dirk Pijpops</i>	580
A Continuum Of Human Cognitive-linguistic Evolution, <i>Olga Vasileva</i>	583
Language Evolution In Ontogeny And Phylogeny, <i>Olga Vasileva</i>	586
Constituent Order In Pictorial Representations Of Events Is Influenced By Language, <i>Anu Vastenius, Jordan Zlatev, Joost Van de Weijer</i>	589
Iconicity, Naturalness And Systematicity In The Emergence Of Sign Language Structure, <i>Tessa Verhoef, Carol Padden, Simon Kirby</i>	591
Language Evolution And Language Origins In Teaching Linguistics At The University Level, <i>Stawomir Waciewicz, Przemysław Żywicznyński, Arkadiusz Jasiński</i>	594
Languages Prefer Robust Phonemes, <i>Andrew Wedel, Bodo Winter</i>	597
Rethinking Zipf's Frequency-meaning Relationship: Implications For The Evolution Of Word Meaning, <i>Bodo Winter, David Ardell</i>	599
The Structure Of Iconicity In The English Lexicon, <i>Bodo Winter, Lynn Perry, Marcus Perlman, Gary Lupyan</i>	602
Signal Autonomy Is Shaped By Contextual Predictability, <i>James Winters, Simon Kirby, Kenny Smith</i>	605
The Cultural Co-evolution Of Language And Mindreading, <i>Marieke Woensdregt, Kenny Smith, Chris Cummins, Simon Kirby</i>	607
A Game Theoretic Account Of Semantic Subjectification In The Cultural Evolution Of Languages, <i>Eva Zehentner, Andreas Baumann, Nikolaus Ritt, Christina Prömer</i>	609

Deep Learning Models Of Language Processing And The Evolution Of Syntax, <i>Willem Zuidema</i>	611
Language-biology Coevolution Fixation Times, <i>Bart de Boer</i>	613
Category Learning In Audition, Touch, And Vision, <i>Sabine van der Ham, Bill Thompson, Bart de Boer</i>	616
Index of authors	618

Papers

BRAIN MECHANISMS OF HUMAN ACOUSTIC COMMUNICATION: A PHYLOGENETIC APPROACH AND ITS ONTOGENETIC IMPLICATIONS

HERMANN ACKERMANN

*Hertie Institute for Clinical Brain Research, University of Tuebingen,
Tuebingen, Germany
hermann.ackermann@uni-tuebingen.de*

WOLFRAM ZIEGLER

*Institute of Phonetics and Speech Processing, LMU Munich
Munich, Germany
wolfram.ziegler@ekn-muenchen.de*

A recent comparative analysis of the neurobiological bases of vocal behavior suggests the basal ganglia to provide the crucial phylogenetic platform for the integration of primate-general mechanisms of acoustic communication with the human-unique faculty of articulate speech (Ackermann et al., 2014). More specifically, structural refinement of the basal ganglia and their connections with the cerebral cortex (corticostriatal loops) – driven, conceivably, by human-specific mutations of the *FOXP2* gene – might have represented a pivotal step towards the emergence of spoken language in our hominin ancestors, allowing the recruitment of the larynx as an articulatory organ. Though the notion that ontogeny recapitulates the adult stages of phylogeny has been dismissed decades ago, the constraints that canalized spoken language evolution must, nevertheless, be expected to have an impact upon speech acquisition. Against this background, a closer look at the contribution of the basal ganglia to the ontogenetic development of vocal behavior appears warranted.

Since congenitally deaf and blind children are capable to cry and laugh, even in the absence of tactile / haptic exploratory capabilities, these vocalizations have been classified – in a human-ethological perspective - as innate “fixed action patterns”, and a hierarchically organized network of, presumably, mammalian-general brainstem control mechanism appears to orchestrate the acoustic components of multimodal emotional displays (Jürgens, 2002). Besides innate affective utterances, the vocal behavior of neonates encompasses speech-related sounds in terms of quasi-resonant nuclei (quasi-vowels), produced in a laryngeal state of “normal phonation”, i.e., the supraglottal vocal tract being “at rest”. As a

1

rule, thus, the physiological prerequisites for the elaboration of a fully operational glottal sound source in our species – such as elaborate monosynaptic cortical projections to the brainstem nuclei that steer the laryngeal muscles (Kuypers/Jürgens hypothesis; Fitch et al., 2010) – must have been established already during (late) fetal life. By contrast, the production of mature consonant-vowel syllable sequences in terms of canonical or reduplicated babbling only arises months later (Oller, 2000). This temporal delay between an early development of speech-related laryngeal functions – starting already at birth - and a protracted emergence of adequately sequenced articulatory (supralaryngeal) constrictions parallels, within some limits, the available data on the maturation of the central nervous system. First, the corticostriatal circuits - especially, the efferent projections of the putamen to its pallidal targets - show a later onset and a prolonged time course of myelin formation as compared to the motor and sensory roots of the cranial nerves as well as the parastriatal aspects of the posterior limb of the internal capsule, encompassing the corticobulbar fibers (Gilles & Nelson Jr., 2012). In addition, second, the striatum displays a still “irregular cytoarchitectonic organization” at the time of delivery, by contrast to an already “very mature” texture in neonates of the infratentorial cranial nerve nuclei, engaged in the innervation of vocal tract muscles (Kostović, 1993).

In conclusion, preverbal vocal behavior appears to evolve across two levels: (i) neonates already master relatively well the operation of a glottal sound source, (ii) the subsequent myelogenetic and cytoarchitectural elaboration of corticostriatal networks then seems to allow for the implementation of syllabic vocal tract movement sequences, based upon the precise adjustment of laryngeal functions and supralaryngeal articulatory excursions.

References

- Ackermann, H., Hage, S.R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, 37, 529-604.
- Fitch, W.T., Huber, L., & Bugnyar, T. (2010). Social cognition and the evolution of language: Constructing cognitive phylogenies. *Neuron*, 65, 795-814.
- Gilles, F.H., & Nelson Jr., M.D. (2012). *The developing human brain: Growth and adversities*. London: Mac Keith Press.
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews*, 26, 235-258.
- Kostovic I. (1993). Zentralnervensystem. In K.V. Hinrichsen (Ed.), *Humanembryologie* (pp. 381-448). Berlin: Springer Verlag.
- Oller, D.K. (2000). *The emergence of the speech capacity*. Mahwah/NJ: Lawrence Erlbaum Associates.

TOWARDS A RIGOROUS MOTIVATION FOR ZIPF'S LAW

PHILLIP M. ALDAY

*Cognitive Neuroscience Laboratory, University of South Australia
Adelaide, Australia
phillip.alday@unisa.edu.au*

Language evolution can be viewed from two viewpoints: the development of a communicative system and the biological adaptations necessary for producing and perceiving said system. The communicative-system vantage point has enjoyed a wealth of mathematical models based on simple distributional properties of language, often formulated as empirical laws. However, beyond vague psychological notions of “least effort”, no principled explanation has been proposed for the existence and success of such laws. Meanwhile, psychological and neurobiological models have focused largely on the computational constraints presented by incremental, real-time processing. In the following, we show that information-theoretic entropy underpins successful models of both types and provides a more principled motivation for Zipf's Law.

1. Introduction

There are two distinct developments that the “evolution of language” refers to, namely (1) the biological, and especially the neurobiological, adaptations necessary for producing, perceiving and processing language and (2) the development of the communication system, abstracted away from generators and receivers of the signal. The development of the communication system has proven remarkably easy to study using rather simple mathematical models (e.g. Ferrer-i-Cancho, 2015b; Lieberman, Michel, Jackson, Tang, & Nowak, 2007; Pagel, Atkinson, & Meade, 2007; Piantadosi, Tily, & Gibson, 2011), but finding an equally parsimonious quantitative model for the neurobiology of language has proven surprisingly difficult. Despite advances on both sides, a combined approach has not been widely adopted, with the mathematical community at times dismissing the “psychological bias”, much less the neuroscientific one (cf. Ferrer-i-Cancho, 2015a). A current neurobiological theory for cortical responses (Friston, 2005) provides the necessary unifying perspective for the evolution of language both as a communication system and as a neurobiological system. More precisely, the diachronic development is shaped by the synchronic constraints arising from basic neuro-computational principles. In the following, we will use this to derive Zipf's Law in the abstract from neurobiologically motivated first principles as well as provide a prediction about the form of its parameters.

2. Zipf's Law and Psychological Tradeoffs in Communication

Frequency-based explanations are common in empirical linguistics, yet they provide few deep, causal explanations (e.g. what drives the development of the frequency distribution?). Nonetheless, distributional statistics provide a convenient, largely theory agnostic method for modelling properties of a language. Zipf (1929, 1935, 1949) demonstrated that distributional statistics in language often follow a Pareto-like distribution (although that is not the terminology he used). Zipf suggested a number of power laws in language, but in the following we will focus on the relationship between frequency (f) and rank (r):

$$f \propto \frac{1}{r} \Leftrightarrow f = \frac{c}{r} \quad (1)$$

for some constant c . This is often extended via an exponent, empirically observed to be near 1, allowing for a slope parameter when plotted log-log:

$$f = \frac{c}{r^\alpha} \Rightarrow \log f = \log \frac{c}{r^\alpha} = \log c - \alpha \log r \quad (2)$$

Now, the probability density function (PDF) for the Pareto distribution is given by

$$P(x) = \frac{(\alpha - 1)x_0^{\alpha-1}}{x^\alpha}, x \geq x_0 \quad (3)$$

where $x_0 > 0$ is the location parameter and expresses the minimum possible value and $\alpha > 1$ is the shape parameter^a and expresses how “bent” the distribution is. When x 's are ranks, then $x_0 = 1$ and this reduces to

$$P(x) = \frac{\alpha - 1}{x^\alpha}, \quad x \geq 1 \quad (4)$$

which we recognize as a special case of Equation (2) when $c = \alpha - 1$.

Zipf postulated a principle of least effort as the motivation for his empirical laws, and indeed this matches well with the “80-20” laws often associated with the Pareto distribution. Ferrer-i-Cancho and Solé (2003) added mathematical rigor to this intuition via simultaneous optimization of hearer and speaker effort when operating on signal-object associations and showed that Zipfian distributions emerge naturally when hearer and speaker effort are weighted equally.

3. Linking Brains and Behavior: Words as Experiments

Friston (2005, 2009) proposed a theory of neurocomputation based on the fitting of generative models of upcoming perceptual stimuli via expectation maximization. Friston, Adams, Perrinet, and Breakspear (2012) expanded upon this proposal by incorporating action into the model-fitting process, focusing on saccades

^aTraditionally, the PDF is expressed with α and $\alpha + 1$ such that $\alpha > 0$, but our presentation makes the notation more compatible with the literature on Zipf's Law.

(eye movements) in visual processing. An accurate model follows from minimizing the (information-theoretic) free energy and surprisal in the generative models. However, in order to best improve the generative model, the most informative, and therefore the *most surprising* stimuli, are sought out.

3.1. Information-theoretic Surprisal

In information-theoretic terms, *surprisal* is also called self-information and is defined as

$$I(x) = -\log P(x) \quad (5)$$

i.e. the self-information of a specific element, class or form is the negative logarithm of the probability of its occurrence. The logarithmic transform provides power-law type scaling and turns additive effects on this scale into multiplicative effects on the original scale. Because probabilities are always between zero and one (inclusive), the logarithm is always negative and thus the negative sign in the definition places self-information on a non-negative scale. Although “self-information” and “surprisal” are technical terms with a precise definition, they nonetheless correspond roughly to intuition. The less probable a certain element is (i.e. the less expected it is), the closer its probability is to zero and hence the further its logarithm is away from zero, i.e. the greater its surprisal. Moreover, they contain more information in themselves because they are not as easily predictable.

3.2. Information-theoretic Entropy

We can also consider the amount of information contained in an entire set, or, equivalently, how much surprisal we should expect from a “typical” or “average” element. In technical terms, the expected value is given by:

$$H(X) = E[I(x)] = -\int P(x) \log P(x) dx, \quad x \in X \quad (6)$$

This value is commonly called *entropy*.

3.3. Maximizing Entropy in Language

If we assume that language is optimized for the balance between hearer and speaker, then we can replace $P(x)$ by the Pareto PDF (3,4) above and can maximize the entropy of language, i.e. the average surprisal, by optimizing the parameter α .

In particular, the entropy of the Pareto distribution (with $x_0 = 1$) is given by:^b

$$H(X) = \log\left(\frac{1}{\alpha - 1}\right) + \left(\frac{\alpha}{\alpha - 1}\right) \quad (7)$$

^bThe derivation of this result is beyond the scope of this paper. Again, we use a slightly non traditional parameterization to better match the literature on Zipf’s Law.

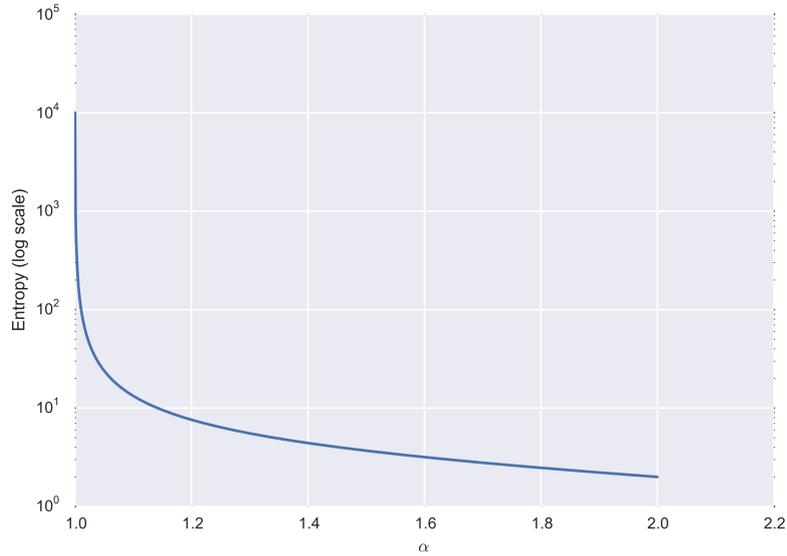


Figure 1. The entropy of the Pareto distribution decreases with increasing α

Figure 1 shows the relationship between α and the entropy of the Pareto distribution. As $\alpha \rightarrow \infty$, $P(x)$ converges to the Dirac delta-function $\delta_{x_0=1}(x)$ and entropy drops as only one symbol (word) from a large pool is meaningful. Intuitively, this would happen when a language consists of nearly only filler words and one meaningful word — if that word occurs exclusively, then it is not very informative in itself because there is no surprise, but if that word occurs rarely, then it is very informative but this contribution is lost in the average. However, as $\alpha \rightarrow 1$, the distribution becomes successively flatter, but maintaining a spike-like structure with a thick tail. Intuitively, this occurs when a small number of words are highly informative but all words have non-vanishing information content. As such, we expect that $\alpha = 1$ is near optimal when $c = \alpha - 1$ and that languages will have evolved (in the communicative sense, but following from the biological evolution) to have near optimal α .

4. Empirical Estimates Across Languages

Baixeries, Elvevg, and Ferrer-i-Cancho (2013) have previously shown that α decreases during first language acquisition on the basis of data from four Germanic languages, generally converging towards a value slightly below 1, with a fair

amount of inter-language variability. In the following, we examine α across a sample of 310 languages using the translations for Universal Declaration of Human Rights provided by the `nltk.corpus` Python package (Bird, Klein, & Loper, 2009) (see Table 1). We use ordinary least-squares regression to obtain estimates for the intercept ($\log c$) and slope ($-\alpha$) from Equation (2). Source code for the analysis can be found on Bitbucket.^c

Table 1. Estimation of α by encoding. Error is standard error of the mean across single-language estimates. Encoding serves as a proxy for writing system; for this corpus, UTF-8 is typically used for ideographic scripts, while *Other* includes Hebrew and Arabic scripts.

Encoding	n	α
Latin1	190	0.90 ± 0.01
Cyrillic	10	0.74 ± 0.08
UTF8	86	0.97 ± 0.03
Other	110	0.93 ± 0.03
All	310	0.90 ± 0.01

4.1. Constant of Proportionality

In hypothesizing that $\alpha = 1$ is optimal, we assumed a proper Pareto distribution, i.e. that $c = \alpha - 1$. Figure 2 shows that this is not quite true, with $\alpha - c = 0.8$ perhaps representing a more realistic assumption. As such, we expect that α will accordingly be shifted away from 1. In particular, we can consider accommodate this shift by setting α in Equation (4) equal to $\alpha' + k$ for some constant k . Then we have

$$P(x) = \frac{\alpha' + k - 1}{x^{\alpha' + k}}, \quad x \geq 1 \quad (8)$$

which implies that our estimate for α should be shifted away from one by the same amount as $\alpha - c$, i.e. we should expect $\alpha \approx 0.8$ to be near optimal.

4.2. Exponent

Figure 3 shows the distribution for estimates of α across languages, with a mean of about 0.9 (cf. Table 1). This is somewhat less than the original predicted idealized value of $\alpha = 1$; however, it is line with our updated estimate based on the bias in $\alpha - c$. Moreover, our c -corrected estimate provides an explanation of why previous work has found α to be near one, but rarely exactly one, even when corrected for observation error.

^c<https://bitbucket.org/palday/evolang/>

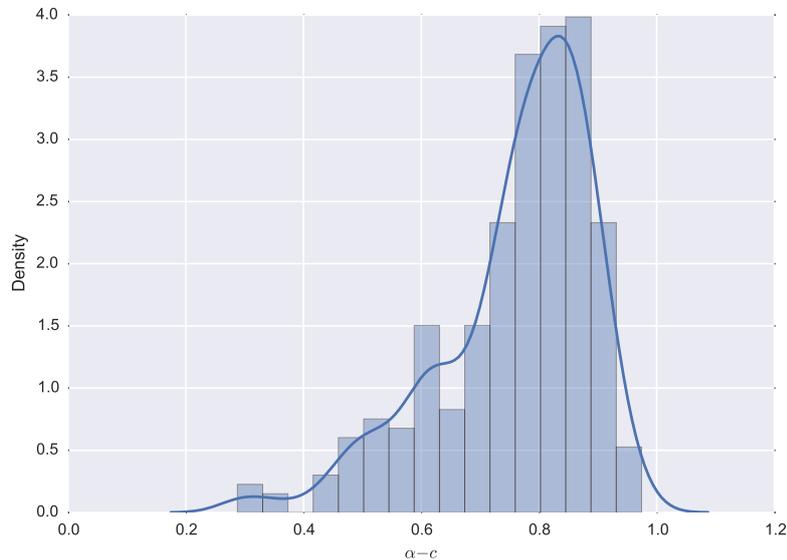


Figure 2. The difference between the constant of proportionality and α . If Zipfian distributions are exactly Pareto, then $\alpha - c = 1$, but this is not the case. Instead, the distribution is shifted left, with a mode of around 0.8

As writing systems may have an impact on blind orthographic measures (i.e. notions of “word” based purely on white-space delimited tokenization), we provide additional estimates divided by text encoding, which stands as a proxy for orthographic system, see Table 1 and Figure 4. Although the shape of the distribution varies across orthographic systems, the distributions all have a mode near 0.8, which suggests that the writing system does not lead to differences larger than those previously observed between closely related languages (cf. Baixeries et al., 2013).

5. Conclusion

Frequency-based explanations abound in empirical linguistics, from corpus linguistics to psycho- and neurolinguistics. Yet, they often suffer from a bit of a chicken and egg problem: X does this because X is more frequent, but how did X become more frequent in the first place? The results presented here provide a first step towards grounding empirical laws in the processing constraints and strategies of individual language users. We have shown how neurocomputational principles can motivate empirical laws via processing strategies, but not yet provided a direct

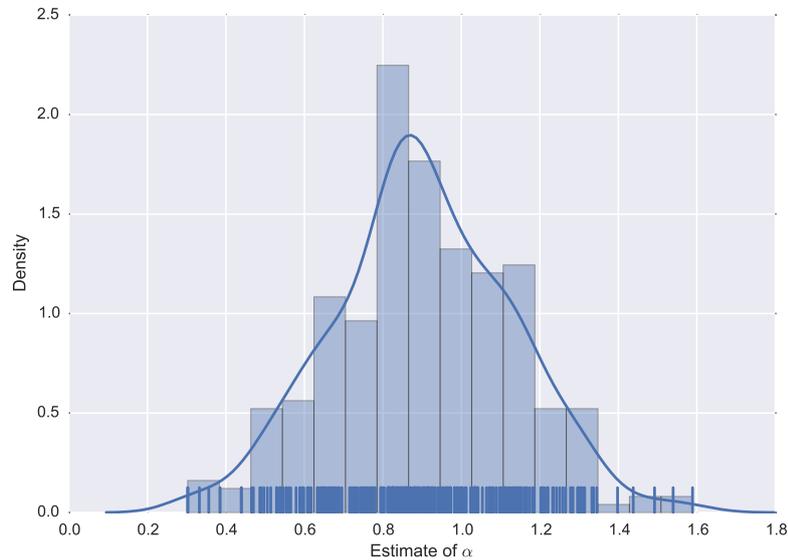


Figure 3. Distribution of α across languages. The surprisal-maximizing Pareto model predicts that $\alpha = 1$ should be ideal

derivation — our model is compatible with the principles but does not require neurobiological specifics and thus remains psychological. Nonetheless, we are able to formulate hypotheses in a principled way about the ideal values for parameters, which bear out in empirical testing. Having parameters that relate back to assumptions about basic cognitive strategies and processing constraints are far more valuable than parameters related to uninformed curve fitting. We can and should have both quantity and explanatory quality. Theories of language evolution need to be motivated by the biological entities doing the evolving.

Acknowledgements

I would like to thank Oliver Schallert for engaging discussions of empirical laws in synchrony and diachrony as well as two reviewers for invaluable suggestions on the textual presentation.

References

- Baixeries, J., Elvevg, B., & Ferrer-i-Cancho, R. (2013). The evolution of the exponent of Zipf's law in language ontogeny. *PLoS One*, 8(3), e53227.

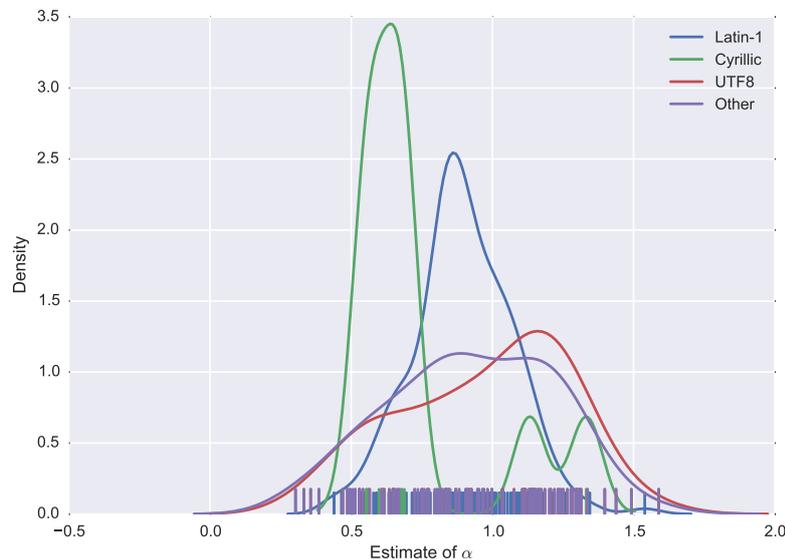


Figure 4. Distribution of α across languages. There is some variation in the shape of the distribution across languages, but the location seems similar with α near 0.8

- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. O'Reily Media.
- Ferrer-i-Cancho, R. (2015a). Reply to the commentary “be careful when assuming the obvious” by P.M. Alday. *Language Dynamics and Change*, 5, 147–155.
- Ferrer-i-Cancho, R. (2015b). The placement of the head that minimizes online memory: a complex systems approach. *Language Dynamics and Change*, 5, 114–137.
- Ferrer-i-Cancho, R., & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3), 788-791.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836.
- Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, 13(7), 293–301.
- Friston, K., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Frontiers in Psychology*, 3(151).
- Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449(7163),

713–716.

- Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout indo-european history. *Nature*, 449(7163), 717–720.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*.
- Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, 40, 1–95.
- Zipf, G. K. (1935). *The psycho-biology of language*. Boston, MA: Houghton Mifflin Company.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.

EVOLUTION TOWARDS AN OPTIMAL MANAGEMENT OF LINGUISTIC INFORMATION

LLUÍS BARCELÓ-COBLIJN

*Department of Psychology, University of the Balearic Islands,
Palma, Spain
lluis.barcelo@uib.cat*

Network science shows great applicability, also in cognitive science related areas. Network-based approaches to the brain are able to extract a pattern, called the connectome. Typically, humans develop a small-world (SM) brain pattern, which seems to be an optimal network in terms of information management. However, in atypical development the pattern changes (or deviates). In the case of network-based approaches to language development, syntactic networks show that children develop their language capability till reaching a stage characterized by a SM network. It is suggested that during evolution human brains evolved in order to develop an optimal brain able to deal with a huge quantity of linguistic information. In order to better understand the typical path, information about atypical language development should be recovered. To date, there is still no information about atypical linguistic networks. In the present work three different linguistic disorders (down syndrome, hearing impairment and language specific impairment) are put to the test. This allows comparisons between different biological conditions that affect the global patterns of linguistic phenotypes.

1. Introduction

Network Science (Newman, 2010) has shown a great applicability, also in cognitive science (Baronchelli et al., 2013). In the last years, a new impulse from the interaction of network science and linguistic theory has emerged, so that new information – invisible to the unaided eye – about the speaker's linguistic capability is now available. Networks provide global information, which is only perceptible once many pieces of a system are connected; therefore, to focus directly on words or on individual sentences is not enough to extract this particular kind of information. In other words, the kind of information that is sought is related to the complexity of the system as a whole. After several network-based studies of language ontogeny – focused on the syntactic capability to combine words –, the study of atypical language development has also been addressed (Barceló-Cobljin et al. 2015), putting forth a research agenda which envisages syntactic networks as endophenotypes. The ultimate goal will be the configuration of a phenotypic morpho-space in which

each linguistic phenotype – either typical or atypical – can be placed Barceló-Coblijn & Gomila, 2014). Three sets of syntactic networks – including Down syndrome (DS), Specific language impairment (SLI) and Hearing Impairment (HI) – will be compared to a control group and to longitudinal studies of language acquisition. These analyses represent the first steps towards the morpho-space of linguistic phenotypes.

2. Complexity

Neil Fraser Johnson (2007) defines complexity science as the “study of the phenomena which emerge from a collection of interacting objects”. To adopt this general definition implies that complexity could be applied to language studies too, and so complex systems theory offers a promising new approach to language. Networks can represent any system in which there are elements interacting to each other. A crucial and basic aspect of Network science is that two networks could contain the same number of nodes, but a different number of edges. Hence, these two networks would show a completely different behavior. By analyzing them, one would see radical different scores in the network measures: a network with k nodes could be *regular*, if all nodes have exactly the same number of connections; however, it could be *scale-free*, if there appear a small number of highly connected nodes (also known as *hubs*). A *scale-free* network of k nodes can develop the characteristic of *small-world*, if it has a high *cluster coefficient* (indicating how many neighbours of a given node are also neighbours of each other) and a low *path length* (how many connections must be traversed from a randomly chosen node to another randomly selected node). Thus, the n of nodes is of course important, but even more important is the particular connectivity a network displays (Watts & Strogatz, 1998).

3. Network Science and Linguistic studies

Applications of network approaches to language studies are easy to figure out: nodes can be phonemes, or words, or morphemes, etc, while edges would be interactions between those elements. Phonology has received quite a bit of attention (e.g., Vitevitch 2008): the goal was to capture particular aspects of the mental lexicon, where *Neighbourhood density* is an important concept: a word is considered a “phonological neighbour” if one single phonological change transforms that word into the target word. Thus, the attested similarities between languages as different as Spanish, Mandarin, Hawaiian and Basque strongly suggested commonly shared psycholinguistic mechanisms used in the architecture of the mental lexicon (Arbesman, et al. 2010). Its application is also being expanded to clinical studies (Vitevitch & Castro, 2015). Interestingly, the application of Network science to language is reaching other linguistic areas:

e.g., based on word co-occurrence, Beckage, Smith & Hills (2011) analyzed the semantic network growth in typically developed (from now on TD) and late talkers. Thus, their results support the thesis that *small-world* connectivity is somehow linked individual children's lexical development. Moreover, it has also been shown that language corpora let see a complex system behavior, and could be represented into a graph of word interactions (Ferrer-i-Cancho & Solé 2001). Thus, a novel like *Moby Dick* shows precisely a small-world network pattern of connectivity (Solé, et al. 2010). A different technique – this time combining syntactic theory and network science – shows that syntactically annotated corpora can have a network representation (Corominas-Murtra, et al. 2009). This work showed that it is possible to apply a network-based technique to the longitudinal stages during language development – as a way to go beyond the analysis of isolated sentences or utterances. Corominas-Murtra and collaborators' work represented in fact the birth of a particular technique of linguistic analysis, which is expanded here.

4. Syntactic networks

Following the line of research sketched out in the previous section, a new technique of linguistic analysis has emerged, which includes the syntactic analysis of each sentence in a corpus, phrase or utterance – following the Dependency Grammar (Hudson, 1990). Analyses of linguistic corpora covering one year of children's life have shown that children, regardless the language they acquire, follow the very same developmental stages (Corominas-Murtra, et al. 2009; Barceló-Coblijn et al. 2012). These three stages can be represented by a type of network: first a *tree-like* network, then a *scale-free* network and finally a *small-world* network. The syntactic analysis of language in combination with graph representation, not only applies the formal procedure of networks, but these are networks confidently encoded by a human linguist. Thus, lexical categories are differentiated, as well as syntactic relationships (i.e. in $a > b$, element “*a*” *syntactically depends on* element “*b*”). This technique uncovers how speakers combine words: which are the most connected words, what kind of relations they establish to each other and how the system behaves. The integration of all manually annotated sentences from a spontaneous speech corpus into a graph typically lets see that the graph is made up of one or more networks (called *components*). Metaphorically speaking, it is like a picture of a city taken from a hot-air balloon in the air. While standing on the street one can see the houses of that street – and many details of each house –, nothing can be said about how streets connect each other. Instead, a macroscopic picture of the town lets see how streets connect to each other, which houses are centric and which peripheral to the structure. It is the very same object, though from a

different perspective. Thus, a syntactic network becomes a kind of routing map of the speaker's public language.

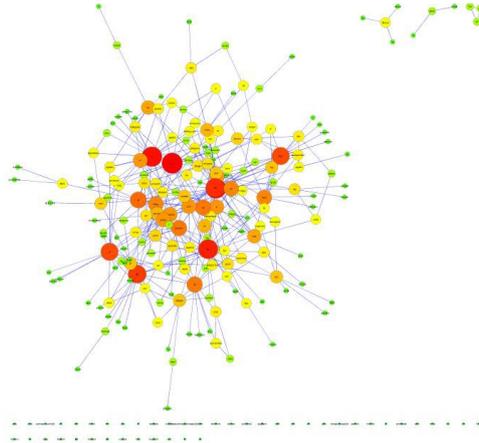


Figure 1. Graph from a sample of a Catalan speaker, age 2 years and 8 months.

5. Brain networks: the connectome

Network approaches have also dealt with brain structures, given the fact that the brain can also be considered a complex system. By means of combinations of neuroimaging techniques (e.g., diffusion tensor imaging) and network calculations, brain networks have become a real, useful tool in clinical studies too. According to Dehaene & Changeux (2011), higher order cognitive functions depend on a more integrative network topology. The structural configuration of the brain has indeed caught many researchers' attention, and several techniques have been developed in order to get more insight about it (Sporns, Tononi & Kötter, 2005). Following this point of view, van den Heuvel & Sporns (2011) could show that human brain networks have a kind of inner sub-structure only made of *hubs*: Up to 12 strongly connected bihemispheric hub regions. This structural characteristic has been dubbed as the *rich-club* effect. Collin & van den Heuvel (2013) review the connectome approach showing that there is a good deal of evidence pointing to the usefulness of network-based approaches to the study of both brain development and brain connectivity, which can be applied not only to ontogeny but also to ageing studies. Why could the macroscopic view be interesting for cognitive science? To be sure, it is a different perspective and hence the information quality will also be different than (and complementary to) the information from micro- or even mesoscopic approaches. Perhaps the most interesting thing of network-based techniques is

their capability to capture global dynamics. It seems that the *H. sapiens'* brain is structurally organized as a *small-world* network (henceforth SMN). This structure seems to be more suitable than others for dealing with huge quantities of information. As it has been mentioned before, a SMN has a high *clustering coefficient* and a low *path length*. Among other factors, the presence of *hubs* in the SMN structure has been related to those characteristics. Interestingly, Crossley and colleagues (2014) have observed that hubs of brain networks are generally central to clinical disorders like Alzheimer or schizophrenia. In their study, Crossley and colleagues could observe that “lesions were also disproportionately concentrated in rich-club regions rather than in the much larger number of peripheral regions”. On another front, Ahmadlou et al. (2013) made an analysis of the functional brain connectivity in participants with Down syndrome (DS). The authors used a technique that combined the information from electroencephalogram frequency bands in a framework of Small-Worldness network. Results suggested that the connectivity of DS patients was disrupted in the whole brain in alpha and theta bands and that the global organization of the DS brain is closer to a *random* network than to a *SMN*.

Taking into account these data, it is worth thinking about the brain structure in pathological conditions, and how these conditions affect brain structure development. And then, how such an atypical brain development leads to an atypical linguistic profile. Are there similarities between the atypical brain networks and the atypical syntactic networks?

5.1. Evolution towards an optimal structure?

From the above mentioned studies on the connectome, the conclusion is that TD humans develop a brain structure reaching *small-worldness*. When particular biological factors intervene in the developmental path, the brain phenotype can develop atypically. It has repeatedly been observed that when a deviation from the typical developmental path takes place, the final connectome is also different. The above mentioned network-based studies on language ontogeny seem to agree with the connectome studies: TD children's ontogeny leads to a language capacity characterized by a *small-world* pattern, where for each word, two syntactic links are present. The syntactic network developed there can then be conceived as a reflection of the linguistic phenotype and therefore, for every person, a phenotype (a connectome) – and hence for every connectome a linguistic phenotype too. Thus, there are a (huge) limited number of brain phenotypes according to biological theory (in the lines of the Evo-Devo perspective), and therefore there will also be a (huge) limited number of linguistic phenotypes (Lorenzo & Balari, 2012).

This altogether suggests that there is in general a statistical tendency for brain developing towards the *small-world* pattern. The SM pattern, according to

physics, seems to be an optimal pattern for information processing. In the case of language development, current results also point to the same direction: TD speakers develop their capability for syntactically combining words following the same stages and reaching a characteristic SM pattern. It seems, thus, that something happened during the evolution of hominins that lead *H. sapiens* to develop so fast (during ontogeny) into that specific, linguistic, brain phenotype: A brain able to deal with an enormous amount of information, in an optimal way. However, in order to reinforce this hypothesis, it is necessary to check what happens in syntactic networks from atypically developing speakers.

6. Atypical development of language and networks

It has been observed since long time ago in biology that many morphological outcomes are possible, one for each biological phenotype. Biological variability can be represented by a set of possibilities that are the most probable within a population from the statistical point of view. There exist biological factors that provoke a deviation from the typical developmental path. Some deviations can be more abrupt than others (and hence in some cases the deviation of the linguistic phenotype would seem – from the external / observational viewpoint – more or less close to the typical linguistic phenotype of healthy speakers). We know that for brains, and for language too, thanks to decades of fruitful studies and discussions on linguistic disorders. However, there is still no information about atypical linguistic networks. In fact, if TD children develop a typical brain and a typical linguistic capability that both can be characterized by a *small-world* pattern, a simple prediction would be that any atypical connectome should correlate with an atypical linguistic network. This is something that has not been addressed until now.

6.1. Procedure

In this work, new information on networks is now presented, showing interesting aspects that uncover information about the linguistic capability under three different pathological conditions: Down syndrome, Hearing Impairment and Specific Language Impairment, in comparison to a control group. Each group has 20 Dutch speakers. The corpora were created by Bol & Kuiken (1990) and are available at CHILDEs data base. The syntactic analyses of the present work have been done by means of new software, *Netlang*, specifically created for this purpose (see <http://neurolang.wordpress.com>).

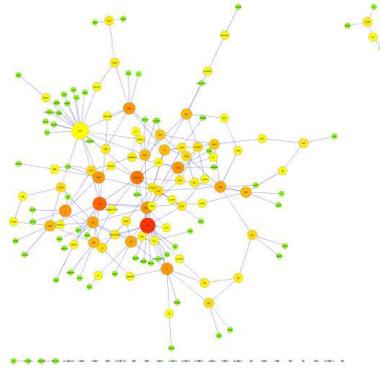


Figure 2. Graph from a sample of a speaker affected by SLI, age 5 years, 1 month and 2 days.

In contrast to previous works, the analyses made with this software has allowed encoding syntactic relationships in the edges: the relationship “subject” can be now seen “inside” the link between two nodes. An additional enhancement is that the software allows the inclusion of lexical categories too. So, the expression “house” can be categorized as “noun”, and this information can be integrated into the network.

6.2. Results and first conclusions

Among other things, it will be shown that many of the atypical linguistic networks present a *non*-small-world pattern like in the TD children's (typically characterized by the 1:2 relation of words/syntactic links; that is to say, for each word, the child produces two syntactic links). Moreover, differences hubs connectivity – like in the atypical connectome. Thus, it seems that also in the case of atypical syntactic networks, hubs are specially affected, in comparison to TD syntactic networks.

There are a number of reasons to pursue the study of networks as endophenotypes, i.e. as biological markers that help to see the connection between genotype and phenotype (Barceló-Coblijn, et al. 2015). To develop such an enterprise, it is first necessary the development of new computational tools that reduce the number of problems due to typological differences between languages. Additionally, larger groups of speakers must be analyzed. Those seminal works included few speakers since these were longitudinal studies (up to 17 files for each speaker, covering between the second and third year of life) and hence focused on ontogeny. Cross-sectional studies can also be informative,

since they have a larger number of participants – and hence statistical calculations are more reliable (this does not mean that a longitudinal study are worse than a cross-sectional one, when in fact they complement each other).

The present work represents a step forward in the study of linguistic networks, since it begins to put to the test the hypothesis put forward here about brain evolution towards an optimal structure for the management of linguistic information. By including three different linguistic disorders and larger sets of speakers, it gives way to compare between different biological conditions that affect the global patterns of the linguistic phenotypes.

Acknowledgements

This work has been supported by funds from the Spanish Ministry of Economy and Competitiveness (grant number FFI2013-44007-P).

References

- Ahmadlou, M., Gharib, M., Hemmati, S., Vameghi, R., & Sajedi, F. (2013). Disrupted small-world brain network in children with Down Syndrome. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 124(9), 1755-1764.
- Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, 20(03), 679-685.
- Balari, S. & Guillermo L. (2012). *Computational Phenotypes: Towards an Evolutionary Developmental Biolinguistics*. Oxford: Oxford University Press.
- Barceló-Coblijn, L., Benítez-Burraco, A., & Irurtzun, A. (2015). Syntactic Networks as an Endophenotype of Developmental Language Disorders: An Evo-Devo Approach to Clinical Linguistics. *Biolinguistics*, 9: 43-49.
- Barceló-Coblijn, L. & Gomila, A. (2014) Syntactic development in phenotypic space. In Cartmill, Roberts, Lyn & Cornish (eds.) *The Evolution of Language. Proceedings of the 10th International Conference (EVOLANG 10)*, pp. 33-40. Singapur, World Scientific.
- Barceló-Coblijn, L., Corominas-Murtra, B., & Gomila, A. (2012). Syntactic trees and small-world networks: syntactic development as a dynamical process. *Adaptive Behavior*, 20(6), 427-442.
- Baronchelli A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N. & Christiansen, M. Networks in Cognitive Science, *Trends in Cognitive Science*, 17(7): 348-360.
- Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS One*, 6(5), e19348.
- Bol, G. W. & Kuiken, F. (1990). Grammatical analysis of developmental language disorders: A study of the morphosyntax of children with specific

- language disorders, with hearing impairment and with Down's syndrome, *Clinical Linguistics and Phonetics*, 4, 77–86.
- Collin, G., & van den Heuvel, M. P. (2013). The ontogeny of the human connectome: development and dynamic changes of brain connectivity across the life span. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 19(6), 616-628.
- Corominas-Murtra, B., Valverde, S., & Solé, R. V. (2009). The Ontogeny Of Scale-Free Syntax Networks: Phase Transitions In Early Language Acquisition. *Advances in Complex Systems (ACS)*, 12(03), 371-392.
- Crossley, N. A., Mechelli, A., Scott, J., Carletti, F., Fox, P. T., McGuire, P., & Bullmore, E. T. (2014). The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain: A Journal of Neurology*, 137(Pt 8), 2382-2395.
- Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200-227.
- Ferrer-i-Cancho, R., & Solé, R. V. (2001). The small world of human language. *Proceedings. Biological Sciences / The Royal Society*, 268(1482), 2261-5.
- Goñi, J., Avena-Koenigsberger, A., Velez de Mendizabal, N., van den Heuvel, M. P., Betzel, R. F., & Sporns, O. (2013). Exploring the Morphospace of Communication Efficiency in Complex Networks. *PLoS ONE*, 8(3), e58070.
- Hudson, R. A. (1990). *English Word Grammar*. Blackwell Pub.
- Johnson, N. F. (2007). *Two's company, three is complexity: A simple guide to the science of all sciences*. Oxford: Oneworld.
- Newman, M. (2010). *Networks: An Introduction*. OUP Oxford.
- Solé, R. V., Corominas-Murtra, B., Valverde, S., & Steels, L. (2010). Language networks: Their structure, function, and evolution. *Complexity*, 15(6), 20-26.
- Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1(4), e42.
- van den Heuvel, M. P., & Sporns, O. (2011). Rich-club organization of the human connectome. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(44), 15775-15786.
- Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research: JSLHR*, 51(2), 408-422.
- Vitevitch, M. S., & Castro, N. (2015). Using network science in the language sciences and clinic. *International Journal of Speech-Language Pathology*, 17(1), 13-25.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of «small-world» networks. *Nature*, 393(6684), 440-442.

A LOTKA-VOLTERRA MODEL OF THE EVOLUTIONARY DYNAMICS OF COMPOSITIONALITY MARKERS

ANDREAS BAUMANN, CHRISTINA PRÖMER

*Department of English, University of Vienna,
Vienna, Austria
andreas.baumann@univie.ac.at, christina.proemer@univie.ac.at,*

KAMIL KAŻMIERSKI

*Faculty of English, Adam Mickiewicz University,
Poznań, Poland
kamil.kazmierski@wa.amu.edu.pl*

NIKOLAUS RITT

*Department of English, University of Vienna,
Vienna, Austria
nikolaus.ritt@univie.ac.at*

Morpho-syntactic boundaries can either be signaled by alignment to boundaries in regular prosodic patterns or by being ‘irregularly’ misaligned, in which case they are often signaled instead through highly dispreferred, or marked, structures such as consonant clusters. In some languages these structures additionally appear in simple forms, which compromises their compositionality-signaling function. This paper models the dynamics of such structures in complex and simple forms by means of a Lotka-Volterra model, which is analyzed evolutionarily. Finally, the evolutionary dynamics of the model are tested against diachronic language data.

1. Introduction

Syntactic and morpho-semantic compositionality is often mirrored, and thereby signaled, prosodically: for example, morpho-syntactic constituents may correspond to rhythmic feet (i.e. alternating sequences of stressed and unstressed syllables), or to regular sequences of (typically CV-) syllables (Schlüter 2005). In a number of languages, however, compositionality is signaled by irregular and phonologically marked structures, such as consonant clusters spanning

inflectional or derivational morpheme boundaries (English /nd/, as in *seem-ed*, for instance). Their signaling function thus compensates for the comparably high costs involved in their production and perception. Nevertheless, such structures can also appear in simple forms (e.g. English /nd/ both in *plann-ed* and in *hand*). Since this obviously compromises their signaling function, the prediction has been made that clusters which span (and signal) morpheme boundaries select against morpheme internal homophones, and vice versa (Dressler & Dziubalska-Kořaczyk 2006, Calderone et al. 2014). On the other hand, however, it has been argued that consonant clusters in simple and complex forms benefit from each other through analogy effects (Hogg & McCully 1987).

Conceptualizing structures acting as compositionality markers as linguistic replicators (Croft 2000, Kirby 2013), the present paper introduces a mathematical model of the dynamics in the cultural evolution of these structures, which takes both of the cognitive mechanisms described above into account. It models both short-term dynamics on the ecological time scale – i.e. the speech-based interaction of potential compositionality markers in complex and simple forms – as well as long-term dynamics on the evolutionary time scale – i.e. the change in the distribution of potential compositionality markers in complex and simple forms. The model is then tested against historical language data.

2. Formulation and analysis of the model

2.1. Ecological dynamics

The ecological short-term dynamics of potential compositionality markers are modeled in terms of a two-dimensional Lotka-Volterra dynamical system in continuous time. Lotka-Volterra models are simple and have been fruitfully applied in linguistics and cultural evolution (Cavalli-Sforza & Feldman 1981, Nowak et al. 2001, Solé et al. 2010, Zhang & Gong 2013). While of limited use for precise quantitative predictions, Lotka-Volterra equations are highly valuable for qualitative analyses. We distinguish between items that actually are signaling complexity (e.g. instances of the consonant cluster type /nd/ spanning a morpheme boundary), and items with the very same structure, that do not (e.g. instances of /nd/ within morphemes). The dynamics of the size of the population of morphologically complex tokens C and the size of the population of tokens of their morphologically simple counterparts S is given by

$$\begin{pmatrix} \dot{C} \\ \dot{S} \end{pmatrix} = A(\lambda) \cdot \begin{pmatrix} C \\ S \end{pmatrix},$$

where the matrix

$$A(\lambda) = \begin{pmatrix} r_c (1 - \lambda) \left(1 - \frac{c}{K}\right) - d & a \\ a & r_s \lambda \left(1 - \frac{s}{K}\right) - d \end{pmatrix}$$

is a function of λ , parametrizing the amount of simple items the structure occurs in (see below). Consider the dynamics of S first. The per token growth rate \dot{S}/S measures the number of tokens in simple items produced or lost per time unit. It depends on three components: (i) production due to replication (uttering and recognizing tokens), (ii) production due to analogy, and (iii) disappearance. The intrinsic reproduction rate $r_s > 0$ measures the number of offspring tokens in simple items produced per token per time unit. It depends on (a) the average frequency of each item type the structure occurs in, (b) the probability of uttering the structure correctly if it occurs in a simple item, (c) the probability of recognizing and processing the uttered structure, and (d) the number of simple items in which the structure occurs, i.e. the number of structure hosts. Let $\lambda \in [0,1]$ be the fraction of actual structure host items. It can increase when potential hosts become actual ones, e.g. through vowel deletion (/nəd/ → /nd/), and decrease when hosts for instance undergo vowel insertion (/nd/ → /nəd/). Finally, the production of such structures in simple items is bounded by a maximal number of possible tokens K , reflecting the total number of potential occurrences in speech utterances (which is clearly limited, not least due to physical reasons). It is assumed that the reproduction rate in the number of simple host structures decreases linearly in K .

In addition to reproduction, tokens in simple items can be produced via analogy to tokens in compositional items at a rate of $a > 0$ (Hogg & McCully 1987). Finally, structures can disappear at a rate of $d > 0$ tokens per time unit. This may happen (a) due to the death of a speaker or (b) when a memorized token is forgotten. Since speakers die, the disappearance rate must be strictly positive, even if speakers never forget tokens.

For complex items, the parameters $r_c, d, a, K > 0$ are defined – *mutatis mutandis* – just as for simple ones. It is assumed that the type frequency of simple items, represented by the fraction of actual structure hosts λ , affects the successful processing (i.e. the decomposition) of complex ones: if the same structure occurs in a large number of simple types, it will be more difficult to recognize the compositionality of the complex items in which it occurs (Dressler & Dziubalska-Kolaczyk 2006). This affects their reproduction rate. For the sake of simplicity we assume it to decrease linearly in λ .

Since $a > 0$, the system above describes a mutually supporting relationship between the two subpopulations. Hence, there exist four population-dynamical equilibria, one of which is an internal sink $(\hat{C}, \hat{S}) > 0$, i.e. both populations stably coexist, as long as a and d are sufficiently small (Hofbauer & Sigmund 1998).

2.2. Evolutionary dynamics

Since the number of actual structure hosts may change diachronically, the model in the previous section above does not suffice for investigating the evolutionary long-term dynamics of potential compositionality markers. Suppose the fraction of actual hosts λ changes to a new value λ' due to replication errors, such as vowel insertion or deletion in the case of consonant clusters. In this way, the configurations of replicating structures and their hosts defined by λ and λ' , respectively, can be conceptualized as two different strategies which define pairs of subpopulations (C_λ, S_λ) and $(C_{\lambda'}, S_{\lambda'})$.

The question is now, whether the new mutant strategy λ' successfully spreads and establishes itself in the language, and if so, what the evolutionary trajectory of this proportion looks like in the long run.

In order to tackle this question, an evolutionary invasion analysis is conducted on the Lotka-Volterra system above (Geritz et al. 1998, Dercole & Rinaldi 2008). For each pair of strategies (λ, λ') the so-called invasion fitness $r(\lambda, \lambda')$ is defined as the expected growth-rate of an initially rare mutant population characterized by a strategy λ' exposed to an environment determined by a resident population characterized by strategy λ at its population dynamical attractor $(\hat{C}_\lambda, \hat{S}_\lambda)$. In the case of the present dynamical system it is sufficient to derive the sign equivalent and algebraically simple expression

$$s(\lambda, \lambda') := \det(A(\lambda))|_{(C_\lambda, S_\lambda) = (\hat{C}_\lambda, \hat{S}_\lambda)},$$

(see Hoyle & Bowers 2008: Appendix B). It can be shown that if the difference between two successive strategies is sufficiently small, then $s(\lambda, \lambda') > 0$ implies that the mutant population successfully invades and replaces the resident population and the mutant population converges to its population-dynamical attractor $(\hat{C}_{\lambda'}, \hat{S}_{\lambda'})$, becoming the new resident population to be invaded by successive mutant populations (Geritz et al. 1998). If $s(\lambda, \lambda') \leq 0$ then nothing changes and the strategy stays the same. The evolutionary dynamics of strategy λ are thus fully determined by $s(\lambda, \lambda')$ and evolutionary equilibria are those points in which

$$\partial s(\lambda, \lambda') / \partial \lambda' |_{\lambda' = \lambda} = 0,$$

i.e. the fitness landscape of λ' determined by the resident λ is locally flat. A search for evolutionary equilibria and a subsequent stability analysis reveals that there exists an evolutionarily stable and attracting equilibrium at

$$\lambda_{int} = \frac{1}{2}$$

and two evolutionary repellors close to the boundaries. Hence the strategy evolves either towards the internal state λ_{int} or towards the boundaries $\lambda_0 = 0$ or $\lambda_1 = 1$. For a given λ one can easily compute the fraction of structures in simple items

$$f(\lambda) = \frac{\hat{S}_\lambda}{\hat{S}_\lambda + \hat{C}_\lambda},$$

shown in Figure 1a. Given the population dynamical equilibria for each λ it can be derived that $0 < f(\lambda_{int}) < 1$, $f(\lambda_0) = 0$ and $f(\lambda_1) = 1$, which means that – depending on their respective starting points – structures will evolve such that they (a) mark complexity unambiguously, (b) mark complexity ambiguously, or (c) never mark complexity (Fig. 1). While possibilities (a) and (c) are what is expected in the literature (Dressler et al. 2010), the possibility of diachronically stable ambiguous complexity markers is less expected. In the model, this is a reflex of the mutual relationship between the two subpopulations.

The fraction f can be estimated from diachronic language data and compared to the evolutionary trajectory of $f(\lambda)$ which is predicted from the evolutionary invasion analysis of the model. This is what will be done in the subsequent section.

3. Evaluating the model: the case of word-final consonant clusters

3.1. Data description

The model is tested against historical language data drawn from a diachronically layered data set of Middle English and Early Modern English words ending in consonant clusters. As already outlined in the introduction, English final consonant clusters are good examples of (potential) compositionality markers. In some cases they unambiguously signal morphological complexity (e.g. /md/ as in *seem-ed*, which never occurs in simple forms), while in other cases they never signal complexity (e.g. /mp/ as in *lamp*). In addition, there are word-final consonant clusters that can occur both in complex as well as in simple forms (e.g. /nd/ as in *plann-ed* and *hand*).

The Middle English period serves as a suitable testing ground for the present model, since due to the process of schwa loss, which started in the 12th century, a huge number of word-final consonant clusters were newly created. Due to this

abrupt change in the language a number of speech-driven repair processes are to be expected. Hence, if the interplay of compositionality marking sequences and those appearing in simple items indeed depends on the cognitive forces outlined in the introduction and incorporated into the model, then the predictions from the model and the development of the historical English data should coincide.

The words containing word-final consonant clusters were extracted from the Penn Helsinki Corpora of Middle English (Kroch & Taylor 2000) and Early Modern English (Kroch et al. 2004) and manually labeled as either complex (i.e. signaling compositionality) or simple (i.e. occurring within lexical items) (see Dressler et al. 2010). In total, approximately 240.000 tokens were included in the data set. The data set was then divided into 6 discrete subsets, one for each century from the 12th to the 17th century.

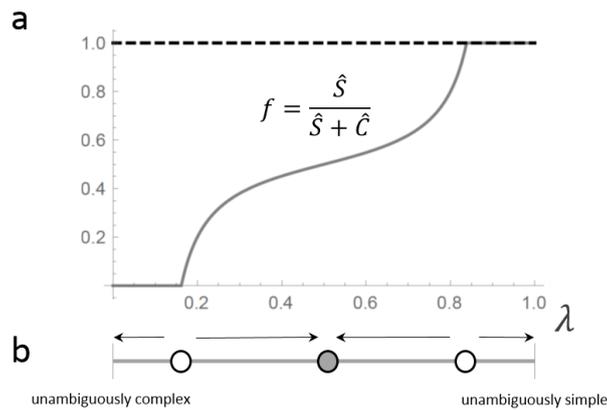


Figure 1. (a) The proportion of the token frequency of replicators in simple items among replicators appearing in simple and complex forms depending on the evolving parameter λ ($r_c = r_s = 10, a = .01, d = 2, K = 50$). (b) Representation of the evolutionary dynamics of λ .

3.2. Historical trajectories and invasion-analysis simulations

For each century and for each consonant cluster type (e.g. /md/, /mp/, and /nd/) the fraction $f = S/(S + C)$ was estimated in order to compute the respective historical trajectories (Fig. 2a). After that the frequency distributions of all f -values were determined for each century (upper histograms in Fig. 2c).

The initial values of the simulated trajectories were drawn from the empirical distribution of the 12th century data. The trajectories were then simulated according to the invasion-analysis procedure outlined in 2.2. As expected, the trajectories approach the boundaries as well as a medial strategy (Fig. 2b). Figure

2c illustrates that the final distribution of the simulated f -values is similar to the distribution of the 17th century data (small to medium difference at $V^2 = .18, \chi^2 = 16.86, N = 160$, after 15 simulation steps). For pairwise *post-hoc* comparisons of the distributions, f -values were grouped together into 5 equally sized bins. This is motivated by the classification of complexity markers proposed by Dressler et al. (2010), in which consonant clusters are either (i) reliably signaling complexity, (ii) signaling complexity by weak default, (iii) ambiguous, (iv) simple by weak default, or (v) almost always simple.

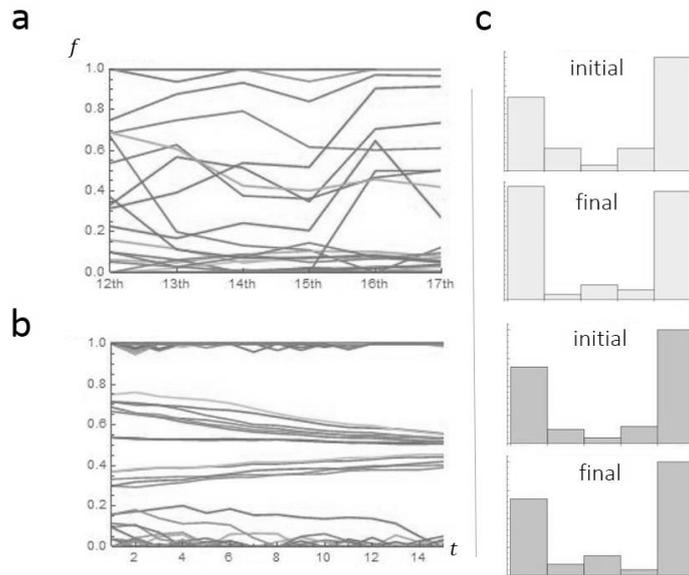


Figure 2. (a) Historical trajectories of the fractions f of $N_{hist} = 42$ to 93 different word-final consonant cluster types (period-wise sample sizes N_{hist} depending on the available data). (b) Invasion-analysis simulation of the evolution of $f(\lambda)$ of $N_{sim} = 80$ independent sequence types for $t = 15$ simulation steps ($r_c = r_s = 10, a = .01, d = 2, K = 50$; population dynamical parameters were chosen *ad hoc*; random drift according to a uniform distribution $\sim U(\pm .01)$ was included to allow for more realistic random fluctuations). Initial values are drawn from the initial historical distribution. (c) Initial and final distributions of the historical (light gray) and simulated (dark gray) f -values, respectively. The difference between the empirical distributions ($V^2 = .18, \chi^2 = 16.86$) is larger than the difference between the final ones ($V^2 = .10, \chi^2 = 15.33$).

The majority of cluster types is located at the boundaries, thus either unambiguously marking complexity or unambiguously appearing in simple forms only. However, as expected, a number of consonant clusters, namely those starting at intermediate f -values, approach the medial ambiguous strategy.

Note that in contrast to the simulated trajectories, the historical English data exhibit properties which contradict the predictions from the model, such as cluster specific trajectories that cross the threshold in the middle of the $[0,1]$ range of f . This can be attributed to the fact that single linguistic replicators, such as certain consonant-cluster types, are also influenced by factors other than analogy and complexity signaling, as for instance major changes in the inflectional system. Although it cannot be guaranteed, we hope that by studying a diverse and comparably large sample of different cluster types, such effects are largely ruled out when looking at the English consonant-cluster inventory as a whole.

4. Conclusion

In this paper a model of the dynamics of linguistic structures that potentially signal compositionality and assist the speaker in the decomposition of larger units into subunits has been proposed. Two counteracting cognitive forces, namely inhibiting semiotic ambiguity effects and analogy, have been incorporated into the model. An analysis of the model as well as simulations reveal that there are three possible stable strategies: potential compositionality-signaling sequences (a) solely appear in complex forms, (b) appear in a fair share of complex forms, or (c) never appear in complex forms. This result is supported by diachronic English language data. In particular, our findings suggest that the coexistence of certain ambiguous configurations – which are *prima facie* disadvantaged in terms of language production and perception – and structurally similar non-ambiguous configurations in language does not necessarily represent an accidental inconsistency, but can rather be explained as a natural reflex of the interaction of well-known cognitive mechanisms.

Acknowledgements

This research was supported by the Austrian Science Fund (FWF, grant No. P27592-G18) and the Österreichischer Austauschdienst (ÖAD, grant No. PL8/2014). We would like to thank Ines Nitsche, Paula Orzechowska and Paulina Zydorowicz for their contribution to the data analysis, as well as three anonymous reviewers for valuable comments and suggestions.

References

- Cavalli-Sforza, L.; M. Feldman (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton, NJ: Princeton University Press.
- Calderone, B., Celata, C., Korecky-Kröll, K., Dressler, W.U. (2014). A computational approach to morphonotactics: evidence from German. *Language Sciences* 46: 59-70.
- Croft, William (2000). *Explaining language change: An evolutionary approach* (Longman linguistics library). Harlow, England, New York: Longman.
- Dercole, F., Rinaldi, S. (2008). *Analysis of evolutionary processes*. Princeton University Press, Princeton, New York.
- Dressler, W. U., Dziubalska-Kolaczyk, K. (2006). Proposing morphonotactics. *Wiener Ling. Gazette* 73: 69-87.
- Dressler, W.U., Dziubalska-Kolaczyk, K., Pestal, L. (2010). Change and variation in morphonotactics. *Folia Linguist. Hist.* 31, 51-67.
- Geritz, S. A. H., Kisdi, E., Meszéna, G., Metz, J. A. J. (1998). Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology* 12: 35–57.
- Hofbauer, J., Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge: CUP.
- Hogg, R., McCully, C. (1987). *Metrical phonology: a course book*. Cambridge: CUP.
- Hoyle, A., Bowers, R. G. (2008). Can possible evolutionary outcomes be determined directly from the population dynamics? *Theoretical Population Biology* 74: 311-323.
- Kroch, A., Taylor, A. (2000). *The Penn-Helsinki Parsed Corpus of Middle English* (PPCME2). Department of Linguistics, University of Pennsylvania. CD-ROM, second edition, (<http://www.ling.upenn.edu/hist-corpora/>).
- Kroch, A., Santorini, B., Delfs, L. (2004). *The Penn-Helsinki Parsed Corpus of Early Modern English* (PPCEME). Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, (<http://www.ling.upenn.edu/hist-corpora/>).
- Kirby, S. (2013). Transitions: The evolution of linguistic replicators. In: P.-M. Binder, K. Smith (eds.), *The language phenomenon*, pp. 121–138. Berlin: Springer.
- Nowak, M. A., Plotkin, J., Jansen, V. (2001). The evolution of syntactic communication. *Nature* 404: 495–498.
- Solé, R. V., Corominas-Murtra, B., Fortuny, J. (2010). Diversity, competition, extinction: the ecophysics of language change. *J. R. Soc. Interface* 7: 1647–1664.
- Schlüter, J. (2005). *Rhythmic grammar: the influence of rhythm on grammatical variation and change in English*. Berlin: Mouton de Gruyter.
- Zhang, M., Gong, T. (2013). Principles of parametric estimation in modeling language competition. *PNAS* 110: 9698–9703.

The Antiquity of Musicality and its Role in Prehistoric Culture

Ted Bayne
Independent Researcher
West Tisbury, MA 02575 USA

Musicality consists of a distributed array of faculties and general substrates, each with its own deep history. In **part one**, a few key components of this array are reviewed to appreciate their antiquity including the salient neurobiological precursors to the underlying sociality. In **part two**: the role of this musicality in hunter-gatherer culture is examined whose peoples also have deep histories and leverage these faculties in polyphonic-polyrhythmic proficiency (Arom, 2004). It is assumed that between 200 and 100 kya, this array of faculties morphed into enculturated musical forms without leaving a trace. But in extant hunter-gatherers a *semiotic plasticity* is found where spoken language is just one modality blended with others. Musicality is critical to rituals that form the praxis of social memory. In collective performance, song/dance integrates the worlds of the spirits, the forest, morality, the hunt, and social homeostasis. Words alone could not achieve the affective and symbolic efficacy required. Very old anthropomorphic structures coalesce the seen natural world and the unseen (but “experienced”) spirit world. Ritualized musical forms demand a semiotic *channel* to bind human care to this humanized cosmology. A final section considers the implications of these topics for language evolution.

1. Antiquity of Musicality’s Underpinnings

1.1 Four phylogenetic precursors critical to pro-social communicativeness

(a) An autonomic social-engagement neurological subsystem developed on top of older sympathetic and much older parasympathetic systems. Evolved in the primate line over 15 my. (Porges & Carter in Narvaez et al, 2013:ch.5) (see §1.2 below); (b) The human “interaction engine”: a pre-linguistic set of capacities and drives (Levinson in Enfield & Levinson, 2006:ch.1; Levinson & Holler, 2014; Levinson in Arbib, 2013:ch.3); (c) Hierarchical Predictive Coding (HPC): predictive processes within cognitive supports of perception and motor, critical for communication (e.g. in Gricean turn-taking, a turn may take 200 ms yet single word latency may be 600 ms) (Shipp et al, 2013; Clark, 2013; Vuust et al, 2014); (d) Extended altriciality and neoteny: caregiver-infant preverbal/tonal IDS processed as “music” (Schoore in Narvaez et al, 2013:ch.2; Koelsch, 2013:244-248; Chang in Fisher et al, 2013:ch.8).

1.2 Vocalization: cortical control, tracking melodic contour, pitch cognition

Morley (2013:144-153, 214-227) carefully works through the evidence for when *Homo* was vocalization-ready. “...both early anatomically modern humans and Neanderthals *do* show modern thoracic vertebral canal development suggesting that such control over extended-duration vocalization has an ancient provenance and is not a feature solely of modern humans.” These and related developments occurred sometime after *H. ergaster*. For archeological evidence and a timeline see Tomlinson (2015:209-237).

Primate and early *Homo* lines saw the loss of fixed calls and a gain in cortical control producing wider range/repertoire in the context of sociality. The development of this sub-system is part of the larger neurobiological development of what has been called the social-engagement autonomic nervous system (ANS). Porges and Carter (in Narvaez et al, 2013:145) summarize the character of this latest layer in the ANS:

This synergism of neural mechanisms in mammals down-regulated defensive systems and promoted proximity by providing social cues (e.g., intonation of vocalization, facial expressivity, posture, and head gesture) that the organism was not in a physiological state that promoted aggressive and dangerous behaviors. Detection of these social cues allowed for symbiotic regulation of behavior and the elaboration of reciprocal care-giving.

Even in primates such as bonobos, vocal utterance is part of an integrated cluster of expressive physical features. Morley (2013:207-208) reminds us of this larger interactive process:

Amongst bonobos, utterances always occur along with facial expressions, and tactile communication, which is true for human ID speech too. Humans also use facial affect (emotional expression) and vocal affect to inform judgment about affective content of each other; they seem to be interdependent systems in both production and perception, with the former having had significant impact on the development of the nature of affective communication in the latter.

Gesture and speech systems co-evolved and leverage common neural/cognitive components. Evidence clearly points to vocalization being very old (Morley, 2014:148). And the notion of “speech vs. gesture” is a false dichotomy. Gesture and speech are a co-existing integrated system. (McNeill, 2012; Kendon in Dor et al, 2014:§6.2; Levinson, 2014; Cole in Radman, 2013:ch. 1; Ghazanfar & Takahashi, 2014; Willems & Hagoort, 2007). Vocalization and bodily gesture grew up together and mirror each other. Selective pressure operated through social-engagement challenges. This aggregated a correlated set of neurobiological developments including gesture and vocalization as well as enhancements to the ANS.

Human processing of pitches is ancient. Archeological dating of the Swabian Jura pipes at 35-40 kya indicates that *H. sapiens* possessed a capacity to perceive, produce, and measure segmented pitches at a time considerably before those dates (Tomlinson, 2015:257-261; Morley, 2013:42-51).

1.3 Rhythmic entrainment and dance

The concept of *entrainment* covers a manifold continuum of inter-coordinated sociality addressing, for example, lithic technology and its related social learning as well as many other forms of shared temporality and rhythmicity (Tomlinson, 2015:76-84). Dance leverages bipedal rhythmic capacities that emerged before modern *H. sapiens sapiens*. Brown (2006), Merker et al (2009), and Ghazanfar & Takahashi (2014) have researched the origins of general rhythmicity and entrainment. Dance is intertwined with musical production and perception (Janata in Arbib, 2013:326-327). It is a cultural invention like song that ritually reinforces group cohesion, sacred beliefs, and social knowledge. "... dance, in the ritual setting, *is* a literature of the non-literate cultures." (Snyder cited in Royce, 2002:154) "Dance may in fact be as old as the human capacities for bipedal walking and running, which date back 2-5 million years." (Brown, 2006:1163)

1.4 The multi-modal brain: music, language, and action processing

The flexibility and domain-general plasticity of the human brain is widely accepted, an ancient adaptive advantage that builds and accumulates culture. The MNS (mirror neuron system) provides an imitative and interpretive power whether modeled on the target goal or the methodical steps. Wakita (2014) drove an investigation (using near-infrared spectroscopy - NIRS) that combined visual interpretation of *silent* action patterns along with the interpretation of the musical content *implied* by that silent action. The study found that "the notion that hierarchical processing in Broca's area is a common function shared between the language and music domains may help explain the role of Broca's area in action perception." Janata et al concur (in Arbib, 2013:325): "The involvement of Broca's area in language production, processing of musical syntax, abstract rule-action relationships in general, and connections to the lateral temporal lobe, makes it a likely substrate for controlling the integration of music, language, and meaning across multiple timescales."

And, finally, Koelsch (2013:§13.2) describes a single continuum and domain:

...overlaps of the cognitive operations (and neural mechanisms) underlying music- and language-processing indicate that 'music' and 'language' are different aspects, or two poles, of a single continuous domain ...the human brain, particularly at an early age, does not treat language and music as strictly separate domains but rather treats 'language as a special case of music' (Koelsch & Siebel, 2005).

2. Musicality's role in prehistoric cultures

Sub-Saharan hunter-gatherers have been studied since the 1960's and 1970's (e.g. Turnbull, Blacking, Arom, and others) with an emphasis on their musicality. Hewlett (2014:Intro) provides a survey and discussion of scholarship in this area since 1970 (see

also the studies referenced here). The very deep history of these peoples has re-emerged through recent mtDNA genetic study, particularly around the “out of Africa” question. Musicality has been a major factor in the *resilience* of these cultures over the course of tens of millennia in three ways: (1) a *semiotic plasticity*, (2) *social memory*, and (3) the *symbolic efficacy* of ritual in an anthropomorphic cosmology.

2.1 Congo Basin hunter-gatherers: first peoples?

These bands are archetypically aboriginal. Genetic data shows limited intermarriage with non-Pygmies until recently (Hewlett, 2014:Intro). They exhibit cultural practices that merge musical and verbal utterance and movement into a communicative continuum (Lewis in Arbib, 2013:47-50).

These [genetic] data showed that the Biaka [Aka] Pygmies have one of the most ancient ... sublineages observed in African mtDNA and, thus, that they could represent one of the oldest human populations. In addition, the !Kung [Bushmen] exhibited a set of haplotypes that were positioned closest to the root of the human mtDNA phylogeny, suggesting that they, too, represent one of the most ancient African populations. (Yu-Sheng Chen, 2000:1362 cited in Grauer, 2011:20-21)

There is significant cultural sharing of fundamentals among the peoples (“Pygmies”) that have been studied in this region. Moïse (in Hewlett, 2014:ch.4) provides a review of current scholarship on Pygmy *history* (after ~70 kya) while also critiquing tired sentimental perspectives that lack evidence. Pygmies have shown hardiness in the modern period despite Bantu/farmer and then colonial impacts. The Bantu period begins somewhere around 3.5-6 kya.

2.2 Antiquity of these hunter-gatherers: the consensus on “out of Africa”

A growing consensus has emerged on when the main migration occurred (~65-55 kya) (Mellars et al, 2013, 2013a), the genetic identity of these groups, the routes taken, and settlement areas (Dennell & Porr, 2014). It is tempting (Grauer, 2011:119) to draw inferences from the presence of the signature musical forms of Pygmy/Bushmen in places as far away as Papua New Guinea (specifically Lihir Island, New Ireland), especially when the genetic identity of the particular PNG people (Hg Q2. Corser, 2012) is a subgroup of the initial migrant parent group (Hg L3). Of course, those who migrated brought their culture with them. And it is likely that the original migrants had developed musical forms long before migration. Ritual musicality can have resilience similar to that of lithic technology itself due to its adhesion to survival activities. Given the wide variety of world musical forms with ancient roots (Jordania, 2011), the *re-creation* of these very same polyphonic/polyrhythmic forms seems unlikely. Lewis (in Arbib, 2013:64) shares a relevant anecdote: “Upon hearing [a recording of] Mbuti

music, the BaYaka immediately recognized that the Mbuti were ‘real forest people’ like themselves, even though genetic studies suggest that they last lived together around 18-20 thousand years ago (Bahuchet 1996).” They clearly detected shared structural-musical fundamentals after a separation of many millennia.

2.3 Semiotic plasticity: an articulatory continuum

Jerome Lewis in his study of the Mbendjele Pygmies goes right to the heart of what can be called a *semiotic plasticity*:

“Mbendjele have developed specific styles of communication for different audiences and situations. They mix words with sung sounds, ideophones, expletives, whistles, signs, hand signals, gestures, vocabulary from other people’s languages, animal sounds, and other environmental sounds, sometimes in a single speech act. In the context of forest hunting and gathering, the role of different language styles and communicative strategies suggests that diverse styles of communicating could have been crucial to the survival of early humans...” (Lewis in Botha & Knight, 2009:ch.13)

[For the Mbendjele ...] *It is not what people are singing but the polyphonic yodelling singing style, not which dances they dance or which spirits they call but the ritual structures they follow, not the language they speak but how it is spoken. The perception of what it means to be Yaka is based on an aesthetic quality and lifestyle as much as on genealogical accident, a distinctive sense of style in which music is more central to culture than language.* (Lewis 2002:68. Emphasis added.)

Some groups are so eclectic in language-use that a specific language cannot be used as a reliable group identifier. Their repertoire of communicative-expressive forms is an extremely versatile, speaking and listening to wide spectrum of beings: animals, spirits, farmers, etc. As Severi (2014:46,57) contends, this semiotic continuum blurs the distinctions between articulatory domains. In what has been called a “synesthetic fusion”, song and dance and visual narrative can all have shared meaning and common set of referents (see also Cross et al in Arbib, 2013:544). The modern sequestering of verbal language away from this articulatory continuum obscures its *roots*: an early development in a rich semiotic *mélange* containing gesture, dance, mime, song, polyrhythm, ridicule, mythology, and dialogic sociality. (see Arom, 2004:§2.4-§2.6).

2.4 Musicality and Social Memory

As Severi (2015:3) stresses, the ritual act itself is the sink of social memory not the “lyrics”. *Oral* cultures are not dependent on the memories of individuals thereby making them fragile. “[in these societies] the use of memory is extremely elaborate...

[these cultures] are founded upon ritual gestures and images as much as upon the use of speech.” (Severi, 2015:14) In the oral culture of these immediate-return hunter-gatherers, non-verbal expressive rituals (dance/song) are forms of kinesthetic social memory. There is no shamanic figure prompting people to “remember”. Daily collective practice touches on the status of the spirit world, re-asserts gender balance, and protects the moral homeostasis of the community. Egalitarianism along with continual vigilance about *autonomy* persists thanks to daily dialogic contention. The *very long-term* endurance of this culture grows out of an embodied syncretism of sacred narratives with the ethologies that express them providing a safe context for vivid differences while rarely turning to aggression. The egalitarian ethos wards off toxic or dangerous people. (Boehm, 2001; Lewis, 2002; Finnegan, 2008, 2013; Kisliuk, 2001).

Lewis (in Hewlett, 2014:ch.8) points to a nonlinguistic core of the Mbendjele form of social memory that blocks emergence of an *expert* who would exercise *authority*:

...institutions such as massana or ekila show how values and meanings can be condensed to establish a cultural store that ensures internal continuity between generations without attributing special status or authority to individuals. Their basically nonlinguistic nature means that it is difficult to articulate them explicitly as a coherent body of ideas.

2.5 Musicality and Ritual: a Cognitive Symbiosis with Symbolic Efficacy

The very *anthropomorphic* nature of the spirit world strengthens the attunement of ritual symbolic expression to the spiritual entities themselves. This anthropomorphic epistemic response is quite ancient and is more experiential than ideological:

...there is virtually no phenomenon to which humans do not ascribe human features, either as persons or as artifacts of personal action. These features may be physical, but more importantly include theory of mind, with human linguistic and symbolic capacities; narrativity; teleology; and mind-body dualism. (Guthrie in Hodder ed., 2014:98)

Ritualized musicality has the heightened power to affect “agencies” in the experienced spirit world, thus having *efficacy* (Lewis in Arbib, 2013:53; Arom, 2004:7). Mbendjele *ekila* (Lewis in Hewlett, 2014:ch. 7; Finnegan, 2008:83) and supernatural notions about the forest and its creatures (Lewis, 2002:95) consist of anthropomorphic attributions of human qualities. The supernatural is blended seamlessly with the seen quotidian world such that named ritual song/dances address spiritual entities in a highly congruent way and the resulting ethologies become increasingly engrained via a social feedback loop.

Dance/song performance seen as an *isolate* is not symbolic but in the larger context it requires symbolic cognition for its production and perception. Musical forms address

the *displacement* inherent in an *imagined* spiritual and natural world. Musicality and ritual forms are symbiotic complements where dance/song is embodied kinesthetic memory grounding and perpetuating the metaphysics of cosmology and sacred propositions. This culturally embedded musicality is tens of thousands of years old. Susanne FÜRNISS describes its role as the *channel* between natural and supernatural:

*Music making is an essential activity, both for the society as a whole and for the individuals that are its foundation, each **servicing in their own way as intermediaries between the natural and supernatural worlds...*** (FÜRNISS in Hewlett 2014:ch.7 [emphasis added])

... music is the very means man has to make contact with these [supernatural] forces... [Each] music may be said to correspond to a precise function, a function to which the 'liturgical' ensemble in which it is inserted corresponds. (Arom, 2004:7)

The named dance/song is the ancient ancestor of **prayer** in the Common Era. It is presumed that the dance/song projects non-verbal signals to (“theory-of-mind”) animals and spiritual figures that receive them as their semiotic *interpretants*. Hunter-gatherers are peers of the animals and spirits and act as the caretakers of “peace in the forest” whereas, with Common Era **prayer**, the supplicant is *lowly* compared with the “creator/lord” being beseeched. Tomlinson, drawing on Steven Mithen, (2015:278) argues for a cognitive symbiosis between musicality (mind/body) on one hand and religious-ritual thought on the other:

The connection of musicking to ritual, religion, metaphysics ... came about as an unfolding congruency between this musical cognition and the similar cognition that made them [religion, etc.] possible. [M]inds capable of formalized abstraction ... linked transcendentalized sociality to musicking in a dance coforming the two.

2.6 Implications for Language Evolution: Multi-Domain Mapping

Michelle Kisliuk’s monograph (2001:ch.4) on BaAka music and life tells of “Seeking the Mother of Elamba” – a “sacred” named dance widely revered. The “mother” is the source of the dance – a curator not an elevated “expert”. The polyrhythmic Elamba dance has a cognitive-kinesthetic complexity. It is loaded with nuance of body language, gesture, and timing. Kisliuk is a young musicologist-anthropologist with unique social and musical skills.

With a difficult kinesthetic structure, the Elamba dance is one half of the cognitive symbiosis discussed above in §2.5: a domain unto itself with its own techniques and semantics. The other half is the spirit world, its entities, and the spiritual access people

have to the animals. As a cognitive domain in its own right, these sacred narratives only exist as embodied and embedded in the ritual praxis of the named dance/songs. There is no verbal recitation of the “beliefs”. The two domains are intertwined. The symbiosis produces a symbolic efficacy that can soften up animals for the kill and beckon or banish quirky anthropomorphic spiritual entities. The cultural practice provides social cohesion as well as an extraordinary perseverance through tens of millennia. Brown et al (2014) have found genetic markers that correlate with this deep-time persistence.

Fifteen years ago Brown (in Wallin et al, 2000:ch.16) proposed “musilanguage” as a common ancestor for music and language (verbal speech). But since 2000, neuroscience has found that music and language are jointly parasitic on many of the very same neurobiological underpinnings (Koelsch, 2013:§13.2,§9.6) as well as riding on the precursors listed in §1.1 above. Emerging in *parallel* in the same biocultural niche (Cross et al in Arbib, 2013:ch.21), music and language were formed out of the friction of use attending to very different social needs yet were part of a single articulatory continuum. Positing a “common ancestor” only spawned more unanswerable questions. Careful critiques of this, like Botha (2009), helped get us past this. Both musical forms and verbal speech deal in different yet analogous ways with the generational issues of social memory (see §2.4 above).

The cognitive domain mapping going on in ritual dance/song is interesting because of its overlap with the generational problem humans must have faced in the Paleolithic. Lithic and related technologies gradually grew more complex. Steps increased and logistics in production and use grew in complexity. At some point it didn’t work to just do a “look at me now”. Dual domains of the embedded gestures of work and symbolic articulations about it had to be cognitively cross-correlated. Music/dance/song was recruited for spiritual matters. Referential articulation with minimal ambiguity can train the next generation about tool production and related skills. Dual-domain mapping was essential in both scenarios and implicated broadly in symbolic behavior.

For Sterelny (2012:ch.3) the generational accumulation of culture is inherently symbolic: “...the distinctive character of human social life depends on the accumulation, preservation, and intergenerational transmission of cognitive capital.” (ibid:§3.5) Tomlinson (2015:181) concurs: “Since forms of production transmitted from generation to generation in the guise of stone tools reach back to the earliest Acheulean traditions we can trace, stone tool production itself – *culture* itself – comes from its first appearance to be pervasively symbolic.” Elsewhere, Sterelny (2012a:2148) argues that “...the elaboration of skilled artisanship with the evolution of context independent and syntactically structured signals, [shows] that the skilled action–gesture co-evolutionary feedback loop helps us explain not just a general increase in communicative capacity ... [but also certain] features central to language as a communication system.”

References

- Arbib**, M.A. ed. (2013). *Language, music, and the brain: A mysterious relationship*. Cambridge, MA: MIT Press.
- Arom**, S. (2004). *African polyphony and polyrhythm*. Cambridge, UK: Cambridge University Press.
- Boehm**, C. (2001). *Hierarchy in the forest*. Cambridge, MA: Harvard University Press.
- Botha**, R. & Knight, C. (2009). *The cradle of language*. Oxford: Oxford University Press.
- Botha**, R. (2009). On musilanguage/“Hmmmmm” as an evolutionary precursor to language. *Language & Communication*, 29(1), 61-76.
- Botha**, R. (2016). *Language evolution: The windows approach*. Cambridge, UK: Cambridge University Press.
- Brown**, S. et al (2006). The neural basis of human dance. *Cerebral Cortex*, August 2006;16;1157-1167.
- Brown**, S. et al (2014). Correlations in the population structure of music, genes and language. *Proceedings of the Royal Society of London B: Biological Sciences* 281.1774 (2014): 20132072.
- Chen**, Y.S. et al (2000). mtDNA variation in the south african kung and khwe. *The American Journal of Human Genetics*, 66(4), 1362-1383.
- Clark**, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* (2013) 36, 181-253.
- Corser**, C.A. et al (2012). The Q2 mitochondrial haplogroup in oceania. *PLOS ONE*, Dec. 2012. doi:10.1371/journal.pone.0052022.
- Dennell**, R. & Porr, M. eds. (2014). *Southern asia, australia and the search for human origins*. NYC: Cambridge University Press.
- Deutsch**, D. ed. (2013). *The psychology of music (cognition and perception), third edition*. London: Elsevier.
- Dor**, D., Knight, C. & Lewis, J. (2014). *The social origins of language*. Oxford: Oxford University Press.
- Enfield**, N.J. & Levinson, S.C. eds. (2006). *Roots of human sociality*. Oxford: Berg.
- Finnegan**, M. (2008). *The political is personal*. PhD dissertation, University of Edinburgh.
- Finnegan**, M. (2013). The politics of eros: ritual dialogue and egalitarianism. *Journal of the Royal Anthropological Institute*, 19(4), 697-715.
- Fisher**, L.F. & Garcia J. R. & Chang, S. (2013). *Evolution's empress*. Oxford: Oxford University Press.
- Ghazanfar**, A. A., & Takahashi, D. Y. (2014). The evolution of speech: vision, rhythm, cooperation. *Trends in cognitive sciences*, 18(10), 543-553.
- Grauer**, V.A. (2011). *Sounding the depths: Tradition and the voices of history*. CreateSpace. Audio at: <http://soundingthedepths.blogspot.com>.
- Hewlett**, B.S. ed. (2014). *Hunter gatherers of the congo basin*. New Brunswick: Transaction Publishers.

- Hodder, I.** ed. (2014). *Religion at work in a neolithic society*. Cambridge, UK: Cambridge University Press.
- Jordania, J.** (2011). *Why do people sing? Music in human evolution*. Tblisi: Logos.
- Kisliuk, M.** (2001). *Seize the dance! BaAka musical life and the ethnography of performance*. Oxford: Oxford University Press.
- Koelsch, S.** (2013). *Brain and music*. Hoboken: Wiley-Blackwell.
- Levinson, S. C., & Holler, J.** (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1651), 20130302.
- Lewis, J.** (2002). *Forest hunter-gatherers and their world*. PhD thesis. Univ. of London
- McNeill, D.** (2012). *How language began: Gesture and speech in human evolution*. Cambridge, UK: Cambridge University Press.
- Mellars, P.** et al (2013). Genetic and archaeological perspectives on the initial modern human colonization of southern asia. *PNAS*, 110 (26), 10699-10704.
- Mellars, P.** et al (2013a). Supporting information for Mellars (2013). *PNAS*, 110 (26).
- Merker, B. H., Madison, G. S., & Eckerdal, P.** (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, 45(1), 4-17.
- Morley, I.** (2013). *The prehistory of music*. Oxford: Oxford University Press.
- Morley, I.** (2014). A multi-disciplinary approach to the origins of music. *Journal of anthropological sciences*. 92, 147-177.
- Narvaez, D.** et al eds. (2013). *Evolution, early experience and human development*. Oxford: Oxford University Press.
- Radman, Z.** ed. (2013). *The hand, an organ of the mind*. Cambridge, MA: MIT Press.
- Royce, A.P.** (1977, 2002). *The anthropology of dance*. Hampshire: Dance Books Ltd.
- Severi, C.** (2014). Transmutating beings: A proposal for an anthropology of thought. *HAU: Journal of Ethnographic Theory*, 4(2), 41-71.
- Severi, C.** (2015). *The chimera principle: An anthropology of memory and imagination*. Malinowski Monographs: HAU.
- Shipp, S., Adams, R.A., & Friston, K.J.** (2013). Reflections on agranular architecture: predictive coding in the motor cortex. *Trends in neurosciences*, 36(12), 706-716.
- Sterelny, K.** (2012) *The evolved apprentice*. Cambridge, MA: MIT Press.
- Sterelny, K.** (2012a). Language, gesture, skill: the co-evolutionary foundations of language. *Phil Trans R Soc B* 2012 25 June 2012. doi:10.1098/rstb.2012.0116
- Tomlinson, G.** (2015). *A million years of music*. Cambridge, MA: Zone/MIT Press.
- Vuust, P., & Witek, M.A.** (2014). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Frontiers in psychology*, 5
- Wakita, M.** (2014). Broca's area processes the hierarchical organization of the observed action. *Front. Hum. Neurosci.* 7:937. doi:10.3389/fnhum.2013.00937
- Wallin, N.L., Merker, B. & Brown, S.** (2000). *The origins of music*. Cambridge, MA: MIT Press.
- Willems, R.M. & Hagoort, P.** (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Lang.* Vol. 101, No. 3, 2007.

THE LOW-COMPLEXITY-BELT: EVIDENCE FOR LARGE-SCALE LANGUAGE CONTACT IN HUMAN PREHISTORY?

CHRISTIAN BENTZ

*Department of Linguistics, University of Tübingen
Tübingen, Germany*

*Department of Theoretical and Applied Linguistics, University of Cambridge
Cambridge, United Kingdom
chris@christianbentz.de*

The quantitative measurement of language complexity has witnessed a recent rise of interest, not least because language complexities reflect the learning constraints and pressures that shape languages over historical and evolutionary time. Here, an information-theoretic account of measuring language complexity is presented. Based on the entropy of word frequency distributions in parallel text samples, the complexities of overall 646 languages are estimated. A large-scale finding of this analysis is that languages just above the equator exhibit lower complexity than languages further away from the equator. This geo-spatial pattern is here referred to as the *Low-Complexity-Belt (LCB)*. The statistical significance of the positive latitude/complexity relationship is assessed in a linear regression and a linear mixed-effects regression, suggesting that the pattern holds *between* different families and areas, but not *within* different families and areas. The lack of systematic within-family effects is taken as potential evidence for a phylogenetically “deep” explanation. The pressures shaping language complexities probably pre-date the expansion of language families from their proto-languages. Large-scale prehistoric contact around the equator is tentatively given as a possible factor involved in the evolution of the LCB.

1. Introduction

Languages are cultural “tools” shaped to successfully transmit information. Due to different pathways and pressures of cultural evolution, they can differ widely with regards to their exact structural characteristics. In this context, there has been a rise of interest in the description, measurement and modelling of language complexity (Sampson, Gil, & Trudgill, 2009; Dahl, 2004; Newmeyer & Preston, 2014; Trudgill, 2011; Baerman, Brown, Corbett, et al., 2015).

This contribution focuses on a variant of information-theoretic complexities (Ehret & Szmrecsanyi, in press; Juola, 2008, 1998; Sadeniemi, Kettunen, Lindh-Knuutila, & Honkela, 2008; Geertzen, Blevins, & Milin, n.d.; Bane, 2008; Moscoso del Prado Martín, 2011), and its implications for the evolutionary pressures that have shaped languages. *Information-theoretic complexity* is here defined with reference to the distribution of word tokens over word types - often

called *lexical diversity*. It is measured across 1155 parallel texts - i.e. translations of the same content - into 885 different languages (see Section 2).

Imagine a language that uses a single word type over and over again, thus having *minimum* information-theoretic complexity. The word type effectively tells us nothing about the meaning encoded. In contrast, a language using a new word type for any conceivable meaning has *maximal* information-theoretic complexity. Every word type is exactly paired with one meaning, and is hence maximally informative.

Note that “complexity” can here be interpreted in two different senses: namely as *learning difficulty* and as *information encoding potential*. A minimum complexity language is extremely easy to learn, but meaningless. A maximum complexity language is hard (or impossible) to learn, but meaningful. The evolutionary trade-off between these two aspects of information encoding has been modelled computationally, and tested experimentally (Kirby, Cornish, & Smith, 2008; Kirby, Tamariz, Cornish, & Smith, 2015; Berdichevskis, 2012; Berdichevskis & Semenuks, in press). Human languages range in between these extremes (Bentz, Verkerk, Kiela, Hill, & Buttery, 2015), falling on a limited spectrum between minimum and maximum complexity. This has far-reaching implications. Minimum and maximum complexities of languages reflect the limits of human learning capacities, and the distribution of complexities across languages gives us a window into the interplay of language learning, usage and linguistic structure on historical and evolutionary timescales.

This study illustrates a systematic geo-spatial pattern relating to information-theoretic complexities across languages of the world. Namely, languages close to the equator have systematically lower information theoretic complexity than languages further away from the equator - given constant content of texts. This phenomenon is called the *Low-Complexity-Belt (LCB)*, and is illustrated in Section 3.1. Its statistical significance is tested in Section 3.2. Moreover, it is shown that though the pattern holds *between* language families and areas, there are differences *within* families and areas (Section 3.3).

Finally, it is argued that the presence of *between-family* correlations - and the absence of reliable *within-family* correlations - suggest that the LCB is a phenomenon with a “deep” phylogenetic explanation. Prehistoric *language contact* is given as a promising candidate for explaining the evolution of the LCB (Section 4).

2. Materials and methods

2.1. Parallel corpora

The parallel corpora used here come from the *Universal Declaration of Human Rights* (UDHR) in unicode,^a the *Parallel Bible Corpus* (PBC),^b and the *European Parliament Corpus* (EPC).^c These add up to an overall sample of around 200 million words, 1529 texts, and 1050 languages (i.e. unique ISO-639-3 codes).

Each text is tokenized by using an algorithm that splits strings of unicode characters on non-alphanumeric characters (i.e. white spaces, punctuation, special characters, etc.).^d The resulting tokens are then added up to the frequency per unique type. For example, the word type *right* occurs 33 times in the English UDHR. Note that this process does not involve lemmatization or stemming, i.e. *right* and *rights* are counted as two separate types here.

2.2. Estimating entropies

For each text the information-theoretic complexity is then calculated as the entropy of the distribution of word tokens over word types. The classic Shannon entropy (Shannon & Weaver, 1949) is defined as

$$H = -K \sum_{i=1}^r p_i \log_2(p_i). \quad (1)$$

Where K is a positive constant determining the unit of measurement (which is bits for $K=1$ and log to the base 2), r is the number of ranks (or different word types) in a word frequency distribution, and p_i is the probability of occurrence of a word of i^{th} rank. According to the maximum likelihood account, the probability p_i is simply the frequency of a type divided by the overall number of tokens in a text. However, it has been shown that the maximum likelihood method is somewhat unreliable, especially for small texts (Hausser & Strimmer, 2009; Nemenman, Shafee, & Bialek, 2001). To estimate entropies reliably, the *James-Stein shrinkage* estimator (Hausser & Strimmer, 2009) is used here instead.

Moreover, texts are taken from three different corpora (UDHR, PBC, EPC) with vastly differing average numbers of tokens (ca. 2K, ca. 10K, ca. 7M), which can additionally bias the estimation of entropy values. To reduce this bias, entropy values are centered and scaled per corpus.

^a<http://www.unicode.org/udhr/>

^b(Mayer & Cysouw, 2014), <http://paralleltxt.info/data/>

^c(Koehn, 2005), <http://www.statmt.org/europarl/>

^dNote that in the PBC - due to careful automated processing and manual double-checking - word types are reliably delimited by white spaces. This makes tokenization fairly robust across many different scripts. The UDHR and the EPC texts have not yet been pre-processed this way. This means they are more prone to errors when splitting strings of characters into word types, especially when problematic characters such as the apostrophe or tone numbers are involved.

Finally, information on latitudes and longitudes per language, as well as information on language stocks (i.e. language families) and language areas, is taken from the AUTOTYP database (Bickel & Nichols, 1999). Merging the scaled entropy values per language (i.e. ISO code) with AUTOTYP information reduces the sample to 1422 texts of 646 languages.

3. Results

3.1. The Low-Complexity-Belt

In Figure 1 entropy values are plotted on a world map using the latitudes and longitudes from the AUTOTYP database. In a) the size of dots reflects entropy values, and their colour reflects area membership. Visual inspection reveals that texts of languages located just above the equator (0° to ca. 30° north) are systematically represented by smaller dots, i.e. lower entropy. This is even more apparent in b), where the longitude is replaced by scaled entropy (and all dots are of the same size now). A loess smoother (black line) again indicates that texts of languages falling on the “belt” between the equator and a latitude of 30° north have systematically lower entropies, with the lowest point at around 15° north. The statistical significance of this pattern is assessed in the following subsections.

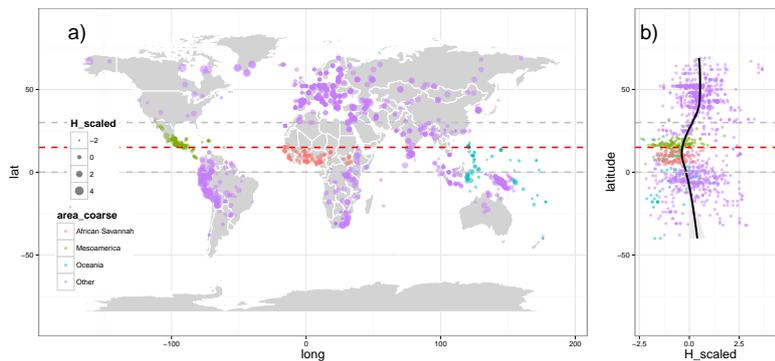


Figure 1. The LCB on a world map. a) World map with scaled entropy values (dots) for 1422 texts (646 languages). The size of dots reflects the scaled entropy. The colour of dots reflects selected language areas: African-Savannah (red), Mesoamerica (green), Oceania (blue), and all others (purple). The LCB is indicated by grey dashed lines at a latitude of 30° north and 0° (i.e. the equator). The core belt is at around 15° north (red dashed line). b) Cross-section with the x-axis reflecting scaled entropy values, and the y-axis representing latitude. A loess smoother with confidence intervals (black line with grey areas) is overlaid to illustrate the systematically lower entropy values around 15° north.

3.2. Simple regressions

3.2.1. Individual languages

If the LCB is an empirical phenomenon that does not derive from random fluctuations in entropies, then the distance from the core of the belt should be a significant predictor of entropy values: bigger distance from latitude 15° north should predict higher entropies.

This is tested in a simple linear regression model run in *R* (R Core Team, 2013), with scaled entropies per text as dependent variable, and distance from 15° north as predictor variable. Homoscedasticity, linearity and normality of residuals are checked visually. In this model the positive association is highly significant ($\beta = 0.023, p < 2.2e^{-16}, R^2 = 0.10$),^e with distance from 15° north explaining 10% of the variance in scaled entropies (see also Figure 2).

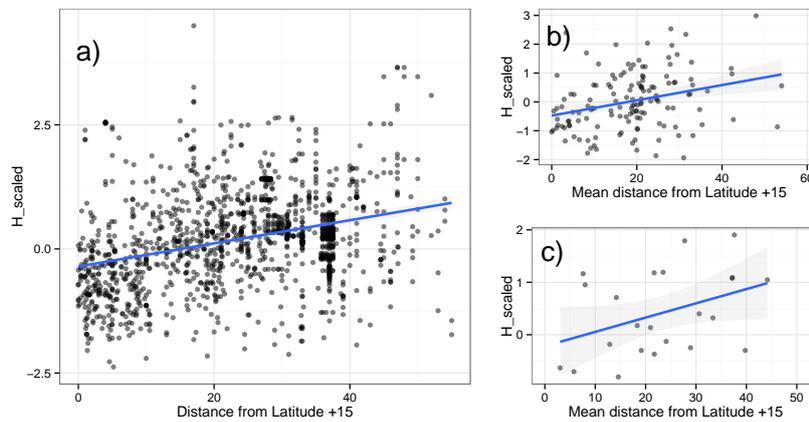


Figure 2. Entropy and distance from the LCB. a) Scaled entropy values (y-axis) for 1422 texts (646 languages) as a function of the distance from 15° north (x-axis). Note that 0 does not indicate the equator here, but distance from 15° north. The positive trend is indicated by a linear regression line with 95% confidence intervals (blue line with grey areas). b) Mean scaled entropy values and mean distances for 140 families. c) Mean scaled entropy values and mean distances for 23 areas.

^e $p < 2.2e^{-16}$ is the smallest p-value that *R* reports, i.e. effectively 0.

3.2.2. Mean entropies per family and area

The significant positive association might be driven by specific language families and areas, rather than being a pattern holding across different families and areas. A way to test this is to use mean entropy and mean distance values per family and area, rather than individual languages. This method is illustrated in Figure 2, panels b) and c).

This time, two simple linear regression models are fitted, with mean entropies per family and area as dependent variables, as well as mean distance from 15° north per family and area as predictors. The β -coefficients for both regressions per families ($\beta = 0.026$) and areas ($\beta = 0.027$) are very similar to the original one (0.023). Note that only for families the positive coefficient is significant ($p = 0.0004$, $R^2 = 0.08$), for areas it is not ($p = 0.06$, $R^2 = 0.12$). The non-significance of this p-value is certainly related to the drastic reduction of sample size from originally 1422 texts to just 23 areas. This is corroborated by a much higher power for the family-wise regression ($\pi = 0.93$) compared to the area-wise regression ($\pi = 0.34$). In fact, a power of at least 0.8 should be aimed for to rule out a high Type II error rate (Cohen, 1992).^f This criterion is met for families but not for areas. However, the positive β -coefficients still indicate that the pattern holds both across different families and across different areas.

3.3. Mixed-effects regression

If the positive association between distance from the core of the LCB and scaled entropies holds *between* different families and areas, does it also hold *within* different families and areas? To further assess this, we can fit linear mixed-effects models (Baayen, Davidson, & Bates, 2008; Jaeger, Graff, Croft, & Pontillo, 2011; Winter, 2013) with distance from 15° north as fixed effect, and family, area, text type and ISO code^g as random effects.

A “maximal” model according to Barr, Levy, Scheepers, and Tily (2013) is fitted with package *lme4* (Bates, Maechler, & Bolker, 2012) in *R*. This is a model with random slopes and intercepts per family, area and text type, and random intercepts for ISO codes.^h This maximal model converges, hence there is no need for a more parsimonious model (Bates, Kliegl, Vasishth, & Baayen, 2015). Again, linearity, homoscedasticity and normality of residuals are checked visually.

It turns out in a likelihood ratio test that this model is not significantly better than a null model without the fixed effect (distance from the LCB) ($\chi^2(13) =$

^fThanks to Johannes Wahle for pointing this out to me.

^g“Text type” here refers to whether the text is from the PBC, UDHR or EPC. ISO codes have to be included as random effects since there are sometimes multiple texts per ISO code and it is conceivable that there is within-language variation.

^hFor ISO codes only random intercepts make sense, since entropy values can only vary for constant distances, and distances can only vary for constant entropies

2.45, $p = 0.12$). This means that when adjusting for idiosyncratic variation within families, areas, text types and languages, the positive association between distance from 15° north and scaled entropy vanishes. In other words, though this association holds *between* families and areas, it does not hold *within* families and areas.¹

A visual way of illustrating this is to plot data points for families and areas separately, as seen in Figure 3. Here, it is apparent that though the positive relationship holds for Benue-Congo languages, it does not hold - and is even inverted - for Austronesian and Indo-European languages.

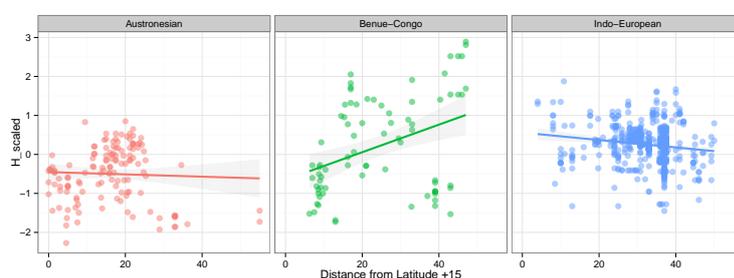


Figure 3. The entropy/latitude relationship for different language families. Scatterplots are faceted by the biggest language families (with more than 50 members): Austronesian (red), Benue-Congo (green), and Indo-European (blue). Linear regression lines with 95% confidence intervals (lines with transparent areas) are overlaid.

4. Discussion

Languages falling inside the *Low-Complexity-Belt*, spanning an area from the equator to ca. 30° north, have significantly lower entropies than languages north and south of them. This pattern is strongly significant in a simple regression model across 1422 texts and 646 languages, and it also holds (with reduced significance) for average values of 140 language families and 23 areas. However, it does not hold if slopes and intercepts per families, areas, texts and languages (random effects) are adjusted. Hence, whatever *causally* explains the positive relationship between latitude and information-theoretic complexity, it is an effect that seems to work at the between-family and between-area level, but is strongly weakened at the within-family and within-area level.

A possible explanation for this could be that the effect had an impact in prehistory on proto-languages of modern day language families, before they started to

¹Note that this is not due to different slopes (i.e. coefficients) per text type, since they are all positive if we look at the PBC, UDHR and EPC separately.

fan out into different branches and wider areas, explaining the between-family and between-area variation. As the effect started to cease or change in recent history, it left no systematic traces at the within-family and within-area level.

4.1. *Language contact*

A potential effect on complexity that has been proposed in the literature is the proportion of non-native adults (L2 speakers) learning a language, i.e. *language contact*. Lopyan and Dale (2010) illustrated that morphological complexity is linked with population sizes in a sample of more than 2000 languages. Population size was here taken as an approximation for language contact. Bentz and Winter (2013) tested this hypothesis more explicitly with regards to nominal case morphology and L2 speaker ratios. Furthermore, a direct link between L2 speaker proportions and entropy (as lexical diversity measure) was established recently (Bentz et al., 2015). Namely, languages with higher proportions of L2 speakers tend to be those with lower entropies. Potential mechanisms of entropy reduction by means of imperfect learning were elicited in a series of iterated learning experiments (Kirby et al., 2008, 2015; Berdichevskis, 2012; Berdichevskis & Semenuks, in press).

4.2. *Deep phylogenetic signals of complexity*

Based on these findings, it is conceivable that areas and families that contribute most to the LCB are those that had the biggest potential in terms of language contact in human prehistory. This makes sense, for instance, for the Benue-Congo family in the African Savannah (and South Africa). It is known as a “deep” family, with migrations and language contact in its early history, such as the Bantu expansion 3000 BC (Pereltsvaig, 2012, p.118). It might be worth considering similar scenarios for languages in Mesoamerica and Oceania. Interestingly, it was shown that entropies in Bantu languages, as well as Austronesian and Indo-European languages, have relatively strong “phylogenetic signals”, meaning that they follow closely the evolution reconstructed from cognate data (Bentz et al., 2015). In other words, entropies of extant languages are “conservative”. They reflect the situation of the past, going back to the roots of the language families several thousand years ago. This suggests that the pressures of the deep phylogenetic past - such as early language contact - might still be reflected in language complexities of today, even if the pressure has ceased to be relevant in recent history.

5. Conclusions

Languages tend to have lower information-theoretic complexity closer to the equator (around 15° north). This pattern is statistically strongly significant, and requires explanation. Subtle differences in significance, relating to variation within and between families and areas, suggest that the effect causing this pattern might

work at deep timescales. Language contact was here proposed as a possible explanation. It is attested as a factor driving lower morphological complexity and lower information-theoretic complexity in large-scale statistical studies, and tested in the lab via iterated learning experiments. However, further studies are necessary to illustrate comprehensively the link between latitude and higher language contact at different time depths. If this link is confirmed, it would constitute important evidence for explaining the evolution and diversity of human languages.

Acknowledgements

This project was funded by an Arts and Humanities Research Council (UK) doctoral grant and Cambridge Assessment (reference number: RG 69405), as well as a grant from the Cambridge Home and European Scholarship Scheme. At a later stage, this work was also supported by the EVOLAEMP project and the *Words, Bones, Genes, Tools* project at the University of Tübingen. I would like to thank Michael Cysouw for letting me use the latest version of the PBC, and Daniela Schwarzmaier de Recacoechea as well as Johannes Wahle for their helpful comments.

References

- Baayen, H. R., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. doi: 10.1016/j.jml.2007.12.005
- Baerman, M., Brown, D., Corbett, G. G., et al. (2015). *Understanding and measuring morphological complexity*. Oxford University Press.
- Bane, M. (2008). Quantifying and measuring morphological complexity. In *Proceedings of the 26th west coast conference on formal linguistics* (pp. 69–76).
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arxiv pre-print*, 1506.04967.
- Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. Retrieved from <http://cran.r-project.org/package=lme4>
- Bentz, C., Verkerk, A., Kiela, D., Hill, F., & Buttery, P. (2015). Adaptive communication: Languages with more non-native speakers tend to have fewer word forms. *PLoS ONE*, 10(6), e0128254. doi: 10.1371/journal.pone.0128254
- Bentz, C., & Winter, B. (2013). Languages with more second language speakers tend to lose nominal case. *Language Dynamics and Change*, 3, 1–27. doi: 10.1163/22105832-13030105

- Berdicevskis, A. (2012). Introducing pressure for expressivity into language evolution experiments. In T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, & J. R. Hurford (Eds.), *The Evolution of Language. Proceedings of the 9th International Conference (EVLANG9)*. Singapore: World Scientific.
- Berdicevskis, A., & Semenuks, A. (in press). Non-native speakers wreak havoc, native speakers clean it up: experimental modeling of contact-induced language simplification.
- Bickel, B., & Nichols, J. (1999). *The AUTOTYP database*. Retrieved from <http://www.autotyp.uzh.ch/>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.
- Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity*. John Benjamins Publishing.
- Ehret, K., & Szmrecsanyi, B. (in press). An information-theoretic approach to assess linguistic complexity. In R. Baechler & G. Seiler (Eds.), *Complexity and isolation*. Berlin: de Gruyter.
- Geertzen, J., Blevins, J. P., & Milin, P. (n.d.). Informativeness of linguistic unit boundaries.
- Hausser, J., & Strimmer, K. (2009). Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *The Journal of Machine Learning Research*, *10*, 1469–1484.
- Jaeger, T. F., Graff, P., Croft, W., & Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, *15*, 281–320. doi: 10.1515/LITY.2011.021
- Juola, P. (1998). Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics*, *5*(3), 206–213.
- Juola, P. (2008). Assessing linguistic complexity. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), *Language complexity: typology, contact, change* (pp. 89–108). Amsterdam: John Benjamins.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.
- Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Mt summit* (Vol. 5, pp. 79–86).
- Lupyan, G., & Dale, R. (2010, January). Language structure is partly determined by social structure. *PLoS ONE*, *5*(1), e8559.
- Mayer, T., & Cysouw, M. (2014). Creating a massively parallel bible corpus. In N. Calzolari et al. (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, Reykjavik,

- Iceland, May 26-31, 2014* (pp. 3158–3163). European Language Resources Association (ELRA).
- Moscoso del Prado Martín, F. (2011). The mirage of morphological complexity. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 3524–3529).
- Nemenman, I., Shafee, F., & Bialek, W. (2001). Entropy and inference, revisited. *arXiv preprint*, physics/0108025.
- Newmeyer, F. J., & Preston, L. B. (2014). *Measuring grammatical complexity*. Oxford University Press.
- Pereltsvaig, A. (2012). *Languages of the world: an introduction*. Cambridge University Press.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org/> doi: ISBN3-900051-07-0
- Sadeniemi, M., Kettunen, K., Lindh-Knuutila, T., & Honkela, T. (2008). Complexity of european union languages: A comparative approach. *Journal of Quantitative Linguistics*, 15(2), 185–211.
- Sampson, G., Gil, D., & Trudgill, P. (2009). *Language complexity as an evolving variable*. Oxford University Press.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: The University of Illinois Press.
- Trudgill, P. (2011). *Sociolinguistic typology: social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint*, 1308.5499.

REDUNDANT FEATURES ARE LESS LIKELY TO SURVIVE: EMPIRICAL EVIDENCE FROM THE SLAVIC LANGUAGES

ALEKSANDRS BERDICEVSKIS, HANNE ECKHOFF

*Department of Language and Linguistics, UiT The Arctic University of Norway
Tromsø, Norway*

aleksandrs.berdicevskis@uit.no, hanne.eckhoff@uit.no

We test whether the functionality (non-redundancy) of morphological features can serve as a predictor of the survivability of those features in the course of language change. We apply a recently proposed method of measuring functionality of a feature by estimating its importance for the performance of an automatic parser to the Slavic language group. We find that the functionality of a Common Slavic grammeme, together with the functionality of its category, is a significant predictor of its survivability in modern Slavic languages. The least functional grammemes within the most functional categories are most likely to die out.

1. Introduction

Many explanations of language evolution and change involve (either explicitly or implicitly) the concept of *redundancy*, especially morphological redundancy. The assumption that redundant features are more likely to disappear has played an important role in historical linguistics for decades (see Kiparsky 1982: 88–99 for an example; Lloyd, 1987: 33–35 for a brief overview). More recently, several influential theories have emerged (Sampson et al., 2009; Lupyán and Dale, 2010; Trudgill, 2011) that refine this assumption, claiming that it does not apply in equal measure to all languages. It is hypothesized that languages under certain sociocultural conditions (such as large population size or a large share of adult learners) will tend to shed excessive (i.e. redundant) complexity.

A serious problem with the notion of redundancy, however, is that it is difficult to operationalize and measure quantitatively, which means that theories such as those cited above must to some extent rest on assumptions or indirect qualitative estimates. In this paper, we improve on a method of measuring morphological redundancy proposed by Berdicevskis (2015).

The key idea behind the method is that the identification of syntactic structure by an automatic parser can be taken as a model of how human beings

understand meaning (i.e. identify semantic structure). While the model is not necessarily ecologically valid (parsers and humans process information differently), it is externally valid: given the same input (text to process) as humans, parsers can approximate the output (correct structure) very well. The main benefit of the model is that it makes it possible to run experiments, manipulating the input. If we, for instance, artificially distort the input, removing the information about a given morphological feature, and then compare the performance of the parser *before* and *after* removal, we can estimate how important the feature is for the identification of the underlying structure, how necessary for the understanding of the meaning and hence, how functional (non-redundant).

Importantly for the study of language change and evolution, this ablation technique can be applied both to extant and extinct languages, as long as there exists a decent treebank. We present a case study where we apply the method to the Slavic language group. We estimate the functionality of morphological categories and grammemes in Common Slavic and test how well this information predicts the survival and death of those features in modern Slavic languages.

2. Materials and methods

2.1. *The Slavic group*

The Slavic language group is divided into three branches: South, West and East. All extant languages have rich inflectional morphology, mostly inherited from Common Slavic. In this section, we describe how Common Slavic grammemes survive across Slavic languages (Table 1). In the following two sections we describe how we measure the functionality of these grammemes.

The earliest Slavic texts were written in Old Church Slavonic (OCS), a literary language based on a South Slavic dialect of Late Common Slavic. We use OCS as a proxy for Common Slavic, as is often done in historical linguistics.

We exclude the following from the analysis: mood, finiteness, voice, degree of comparison, adjective long/short form, synthetic future tense (which exists only for the verb 'be'), non-indicative and non-finite verbal forms.^a The tense grammeme coded as **res** in Table 1 stands for the Common Slavic perfect, pluperfect and conditional that consisted of an auxiliary verb and a so-called **resultative** participle. We do not take into account any morphological

^a The reasons for exclusion range from theoretical (there is no unified view on the structure of some categories, e.g. finiteness) to methodological (our experiments in their current form do not work with binary categories, e.g. adjective form, or categories that are too heterogeneous, e.g. mood).

innovations. Decisions represented in Table 1 largely follow Comrie & Corbett (1993).

Table 1. Common Slavic grammemes across modern Slavic languages

Cate- gory	Gram- meme	CF	GF	freq	South branch				West branch					East branch			
					bul	mkd	hbs	slv	ces	slk	hsb	dsb	pol	csb	rus	bel	ukr
Case	nom	36	6.3	20.5	1	1	1	1	1	1	1	1	1	1	1	1	1
Case	acc	36	5.2	16.1	0	0	1	1	1	1	1	1	1	1	1	1	1
Case	dat	36	4.8	8.4	0	0	1	1	1	1	1	1	1	1	1	1	1
Case	gen	36	4.2	11.1	0	0	1	1	1	1	1	1	1	1	1	1	1
Case	ins	36	2.8	3.1	0	0	1	1	1	1	1	1	1	1	1	1	1
Gend	m	4	2.5	35.4	1	1	1	1	1	1	1	1	1	1	1	1	1
Pers	1	3	2.0	4.8	1	1	1	1	1	1	1	1	1	1	1	1	1
Pers	2	3	2.0	6.1	1	1	1	1	1	1	1	1	1	1	1	1	1
Pers	3	3	2.0	22.4	1	1	1	1	1	1	1	1	1	1	1	1	1
Gend	n	4	2.0	10.3	1	1	1	1	1	1	1	1	1	1	1	1	1
Numb	pl	4	2.0	20.2	1	1	1	1	1	1	1	1	1	1	1	1	1
Numb	sg	4	2.0	60.7	1	1	1	1	1	1	1	1	1	1	1	1	1
Tens	res	8	2.0	0.4	1	1	1	1	1	1	1	1	1	1	1	1	1
Case	loc	36	1.7	3.3	0	0	0	1	1	1	1	1	1	1	1	1	1
Case	voc	36	1.7	0.9	1	1	1	0	1	0	1	0	1	1	0	0	1
Gend	f	4	1.5	10.6	1	1	1	1	1	1	1	1	1	1	1	1	1
Tens	pres	8	1.3	15.6	1	1	1	1	1	1	1	1	1	1	1	1	1
Numb	du	4	1.0	2.1	0	0	0	1	0	0	1	1	0	0	0	0	0
Tens	aor	8	0.7	7.4	1	1	1	0	0	0	1	0	0	0	0	0	0
Tens	impf	8	0.7	2.1	1	1	0	0	0	0	1	0	0	0	0	0	0

CF = category functionality, GF = grammeme functionality (see section 2.3), freq = absolute frequency (in OCS). The table is sorted first by GF (descending), then by CF (ascending), i.e. in the approximate descending order of survivability (see section 3). Languages are denoted by their ISO 639-3 codes. 0 means that a grammeme is (almost) extinct, 1 means that it is extant.

2.2. *Treebank and parser*

We extracted OCS data from the Tromsø Old Russian and OCS Treebank (TOROT),^b using the two largest documents, the Codex Marianus and the Codex Suprasliensis, both dated to the beginning of the 11th century. The joint TOROT file contains 13308 manually annotated (and double-checked) sentences.

^b <https://nestor.uit.no/>

The TOROT is a dependency treebank with morphological and syntactic annotation according to the PROIEL scheme (Haug et al., 2009). For our experiments, we converted the files to the CONLL format (Table 2).

For the parsing experiments we used MaltParser (Nivre et al., 2007), version 1.8.1.^c The parser was optimized using MaltOptimizer (Ballesteros and Nivre, 2012), version 1.0.3,^d optimization was performed on the original text, before any changes (see section 2.3).

Table 2. Example OCS sentence ('He said to them', from Matthew 12:11) in the PROIEL scheme and CONLL format.

ID	Form	Lemma	CPOS	FPOS	Features	Head	DREL
1	on" <i>he</i>	on"	P	Pd	NUMBs GENDm CASEn	3	sub
2	že <i>but</i>	že	D	Df	INFLn	3	aux
3	reče <i>say</i>	reši	V	V-	PERS3 NUMBs TENSa MOODi VOICa	0	pred
4	im" <i>them</i>	i	P	Pp	PERS3 NUMBp GENDm CASEd	3	obl

CPOS/FPOS = coarse/fine-grained part-of-speech tag; DREL = dependency relation. OCS words are transliterated using the ISO 9 system.

Contrary to standard practice in computer science, we do not create separate training and test sets, and thus perform all operations, including optimization, on the whole dataset. The reason for this solution is that our goal is not to evaluate how accurately a given parser can analyze a given text, but how its performance is affected by certain changes in the annotation of the input data. As regards absolute measures of performance, we want them to be as high as possible, in order to approximate human performance and thus increase the validity of the model. Training and parsing on the same set allows us to reach a LAS (labelled attachment score) of 0.938,^e while parsing of unfamiliar test sets usually results in a LAS in the high seventies at best.

2.3. The ablation experiments

We perform two experiments. In the first one, we estimate the functionality of the morphological categories listed in Table 1 (column 1), in the second one, the

^c <http://www.maltparser.org/>

^d <http://nil.fdi.ucm.es/maltoptimizer/index.html>

^e This score is reached when the parser is trained and tested on the whole dataset prior to any changes. In the actual experiments, parsing is done on relevant subsets (see section 2.3) and reference LASes vary, but typically lie around 0.900.

redundancy of grammemes (column 2). Prior to both experiments, we remove all information about word *forms* from the input, replacing every form with the corresponding lemma (see below for the rationale).

In the first experiment, all information about the given category is deleted from the "Features" column (see Table 2). If, for instance, we are interested in the OCS NUMBer category (which includes three grammemes: singular, **d**ual and **p**lural), then a subset containing all sentences which have at least one token with number among its features is created. The parser is trained on the whole dataset and tested on this subset, providing the reference LAS (0.907 for number). After that, the strings *NUMBs*, *NUMBd* and *NUMBp* get deleted (in Table 2, that would affect rows 1, 3 and 4) in both sets. If the removal of the category leaves the "Features" column of a given token empty, the string *INFLn* is inserted (i.e. the token is marked as non-inflecting). After the deletion, the parser is trained again on the whole dataset and tested on the "number" subset. The difference between the reference LAS and the LAS after the deletion ($0.907 - 0.903 = 0.004$) serves as the measure of functionality. This measure is reported in Table 1 (column CF).

In the second experiment, we deal with grammemes. Simple deletion is not a suitable solution both for technical and theoretical reasons. The disappearance of a grammeme almost always means that this grammeme *merged* with another one (as the dual merged with the plural in most Slavic languages). We model this process in the following way: every grammeme within a category (say, *s* within number) is successively merged with every other grammeme within the same category (*d* and *p* in this case). Technically, it means that the string *NUMBs* is always replaced by *NUMBd* during the *s-d* merger (mergers are symmetric: *s-d* is equivalent to *d-s*) and *NUMBp* during the *s-p* merger. As with category deletion, a subset is created, which contains all relevant sentences (for the *s-d* merger, all sentences which have at least one token either in the singular or in the dual form), and both before and after the merger the parser is trained on the whole dataset and tested on the subset (for the *s-d* merger, the reference LAS is 0.9065). The differences in parser performance for each merger are summed and divided by the number of mergers (for *s*, that would mean summing across the *s-d* and *s-p* mergers and dividing by two: $((0.9065 - 0.905) + (0.9065 - 0.904))/2 = 0.002$), the result is considered a measure of grammeme functionality. This measure is reported in Table 1 (column GF).

Note that these changes affect only the "Features" column. When we merge *s* and *d*, we only change their morphological descriptions. We are, however, unable to merge the word forms in a reasonable way (partly due to large form variation within OCS, partly due to the absence of form-generating software).

Thus, if we leave the forms as they are, the merger would not be complete: the parser would still potentially be able to see that there are systematic differences between the singular and the dual forms and use this information. In order to enhance the merger's impact, we perform all experiments after deleting all information about word forms (see above).

3. Results and discussion

Results are presented in Table 1 (see also supplementary materials for more detailed data). Most of the zeroes are clustered in the lower part of the table, and it seems that GF (grammeme functionality) positively correlates with the grammeme's survivability, while for CF (category functionality) the correlation is negative. The former observation is expected, but the latter one is quite surprising. A possible explanation is that the most functional categories are also the largest ones in terms of grammeme number (case has 7, tense has 4). This means that the competition between the grammemes can be higher, or, to put it another way, the sheer probability of a merger is higher. In addition, these categories have more resources to sacrifice, both in terms of grammemes and functionality.

A notable exception from the general trend are the Bulgarian and Macedonian cases, located high in the table. One of the important reasons for their loss most likely is the intense long-term language contact within the Balkan Sprachbund (Wahlström, 2015).

A reviewer asked whether other Slavic languages had experienced less contact. The answer depends on what type of contact we have in mind. In a sprachbund, the contact is long-term, co-territorial and likely to involve child bilingualism. This specific type of contact can favor complexification through additive borrowing (Trudgill, 2011). While Bulgarian and Macedonian did lose the nominal cases, they are the only ones among Slavic languages that developed a definiteness category. It is likely that its development has contributed to the case loss and, in its turn, has been facilitated by the contact (Wahlström, 2015).

Going back to the question, Bulgarian and Macedonian have definitely experienced much more sprachbund-type contact than any other Slavic language. However, as regards shorter-time contacts that involve adult bilingualism and thus are likely to favor simplification (Trudgill, 2011), these two languages score relatively low. For example, Bentz & Winter (2013) estimate the proportion of non-native speakers as 21% for Bulgarian, 42% for Russian, 52% for Serbian (but 26% for Croatian, which is listed separately), 3% for Polish.

It is remarkable that the resultative tense, which was very infrequent in OCS, but survived in all Slavic languages (and became the only past tense in many of them), gets high CF and GF values.

We tested the correlations by means of the mixed effects logistic regression with survival (1 or 0) as the dependent variable, using R (R Core Team, 2015) and *lme4* (Bates et al., 2015). Due to the small size of the dataset and the nature of the dependent variable (categorical and not continuous) we face severe convergence problems, which makes it difficult to apply Barr et al.'s (2013) recommendation to "keep it maximal" and, in some cases, to apply likelihood ratio tests (Wald-z statistic is used instead). We try to keep the model as simple as possible, including only the most important predictors as fixed effects, viz. (centered) CF and GF. Since the maximal model does not converge, we exclude random slope for CF (which is presumably a less important predictor). The final model (henceforth Model 1) includes language and grammeme as random effects with by-language and by-grammeme random slopes for GF. More complex theoretically-justified models either do not improve the goodness of fit (AIC) or do not converge.

According to Model 1, GF increases the logit estimate by 10.37 ± 3.30 standard errors (Wald-z = 3.14, p = 0.0017), while CF decreases it by -2.21 ± 0.62 standard errors (Wald-z = -3.59, p = 0.0003).

The random slopes and intercepts are highly correlated both for grammemes and languages. The analysis of the coefficients does not reveal any interesting patterns for grammemes. As regards languages, Bulgarian and Macedonian get much lower coefficients than others both as intercept (2.66 for both languages vs. mean 8.41, sd 3.21) and as slope (0.61 vs. mean 9.34, sd 4.92). This means that the impact of GF on survivability is much lower in these languages, and the average survivability of grammemes is lower. The real effect behind this is the fate of the case system, which is poorly predicted by our measures. Going back to the discussion of contact types, it can be that the sprachbund influence somewhat shields Bulgarian and Macedonian from the pressure to shed redundant complexity, while imposing some other pressures such as the influence of the neighboring languages.

One might wonder if the random slope coefficients correlate positively with the population size or the share of L2 speakers. This correlation would imply not only that the impact of functionality is different across languages, but also lend support to the language complexity theories cited in Section 1. We find, however, no significant correlation between (modern) population sizes and random slopes. We did not collect data about the shares of L2 speakers or other

social parameters that might be important for language complexity, though this is a promising avenue for future studies.

It is an important question whether the same results can be achieved using simpler predictors, first of all frequency (Table 1, column “freq”). Frequency and GF are not collinear, but seem to have similar impact. Simply replacing the fixed effect of GF by that of frequency in Model 1 (henceforth Model 2) results in slightly worse goodness of fit (AIC 127 vs. 120), but the change is not significant. If, however, we try to abandon GF and CF altogether and build a model using frequency only, then the best model (frequency as a fixed effect, random intercepts for grammeme and language, by-language random slope for frequency, no random correlation for grammeme, henceforth Model 3) is significantly worse than Model 1 ($\chi^2(3)=53.05$, $p < 0.0001$). Somers' D_{xy} and the C index are also slightly worse, resp. 0.978 and 0.989 for Model 1, 0.974 and 0.987 for Model 2, 0.923 and 0.961 for Model 3.

4. Conclusion

We show that the functionality of a morphological feature, measured as the importance of this feature for an automatic parser, is a significant predictor of survivability of the feature. Higher functionality of a grammeme increases its chances to survive, while higher functionality of a category, unexpectedly, decreases the chances of the respective grammemes to survive, although the slope is not as steep as in the former correlation.

The ablation technique that we described has several limitations, and ignores some potentially important factors. These problems can probably be partly overcome, but even in its current form the method can explain some of the morphological variation observed in modern languages and is thus a useful tool to test theories about language change, evolution and diversity.

Acknowledgements

The authors were supported by the Norwegian Research Council grant “Birds and Beasts” (222506). We are grateful to the CLEAR research group for discussing earlier versions of the paper.

References

- Ballesteros, M., & Nivre, J. (2012). MaltOptimizer: A System for MaltParser Optimization. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC), Istanbul, Turkey, 23–27 May 2012*. European Language Resources Association.

- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package ver. 1.1-9.
- Barr, D., Levy, R., Scheepers, C. & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3), 255–278.
- Bentz, C. & Winter, B. (2013). Languages with More Second Language Learners Tend to Lose Nominal Case. *Language Dynamics and Change* 3, 1–27.
- Berdicevskis, A. (2015). Estimating Grammatical Redundancy by Measuring Their Importance for Syntactic Parser Performance. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, 65–73. Association for Computational Linguistics.
- Comrie, B. & Corbett, G. (eds.) (1993). *The Slavonic Languages*. London: Routledge.
- Haug, D., Jøhndal, M., Eckhoff, H. Welo, E., Hertenberg, M., & Müth, A. (2009). Computational and Linguistic Issues in Designing a Syntactically Annotated Parallel Corpus of Indo-European Languages. *Traitement Automatique des Langues* 50(2), 17–45.
- Kiparsky P. (1982). *Explanation in phonology*. Dordrecht: Foris.
- Lloyd, P. (1987). *From Latin to Spanish: Historical phonology and morphology of the Spanish language*. American Philosophical Society.
- Lupyan, G. & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE* 5(1):e8559.
- Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., Marinov, S. & Marsi, E. (2007). MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(2), 95–135.
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Sampson, G., Gil, D. & Trudgill, P. (eds.) (2009). *Language complexity as an evolving variable*. Oxford: Oxford University Press.
- Trudgill, P. (2011). *Sociolinguistic typology: social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Wahlström, M. (2015). *The loss of case inflection in Bulgarian and Macedonian*. Helsinki: Slavica Helsingiensia.

A SCIENTOMETRIC ANALYSIS OF EVOLANG: INTERSECTIONS AND AUTHORSHIPS

TILL BERGMANN¹ & RICK DALE¹

¹*Cognitive and Information Sciences
University of California, Merced
Merced, CA 95340, USA*

Research on the evolution of language has grown rapidly, and is now a large and diverse field. Because of this growing complexity as a scientific domain, seeking new methods for exploring the field itself may help synthesize knowledge, compare theories, and identify conceptual intersections. Using computational methods, we analyze the scientific content presented at EvoLang conferences. Drawing on 365 abstracts, publication patterns are quantified using Latent Dirichlet Allocation (LDA), which extracts a semantic summary from individual abstracts. We then cluster these semantic summaries to reveal the frameworks and different domains present at EvoLang. Of course, our results show that EvoLang is an interdisciplinary field, attracting research from various fields such as linguistics and animal studies. Furthermore, we show that the framework of iterated learning and cultural evolution is a hub topic at EvoLang.

1. Introduction

In this paper, we explore the conceptual structure of research on language evolution itself by analyzing the submissions to the EvoLang conference over the past 10 years. Our goal is to provide insight into the network of theories, concepts, and methods that populate this growing field. Since its inception in 1996, EvoLang has become a prominent and well-attended conference. It is now the premiere conference on language evolution, with more than 100 presentations at the last EvoLang in Vienna and over 300 delegates in attendance. This is a five-fold increase from the first EvoLang in 1996. How might we quantify this rapidly growing scientific content?

There are numerous reviews of language evolution which attempt to unpack and relate its various theories and debates (e.g. Christiansen & Kirby, 2003; Bickerton, 2007; Fitch, 2010). These provide impressive coverage, especially considering the diversity and complexity of language evolution research. Research at EvoLang tackles a wide range of these topics, spanning the many levels of language, from the evolution of flexible signalling strategies, to the social cognitive processes that may undergird human linguistic skills.

In what follows, we use topic modeling (Griffiths & Steyvers, 2004; Yau, Porter, Newman, & Suominen, 2014) to extract the set of latent conceptual topics

that make up EvoLang. We find that there are three distinct conceptual clusters that can be inferred from the abstracts, including the iterated learning framework and comparative studies. Second, we combine these topic clusters with a co-authorship network analysis to assess the relative influence of these typical topic clusters, finding that the iterated learning cluster in particular serves as a central hub in the broader EvoLang community. By analyzing the knowledge bases of EvoLang, it may be possible to attain a firmer grip on the state of the art in the field, and the relationships among its various theories.

2. Modeling the content of EvoLang submissions

We selected all abstracts from submissions between 2006 and 2014 with more than 500 characters.^a We then applied a Latent Dirichlet Allocation algorithm (Blei, Ng, & Jordan, 2003) on the resulting 375 abstracts, a method that is commonly used in scientific content analysis (Griffiths & Steyvers, 2004). In LDA, each document (here, abstract) is represented by a distribution over topics, and the topics themselves are represented by a distribution over words. That is, each topic consists of a distribution of semantically related words, and each abstract can then be represented as a combination of these topics, which make up the *gist* of the document. For example, one abstract at EvoLang may combine the topics of non-human communication and learning, while another may combine syntax and computation. Importantly, the algorithm first extracts numerically identified topics, which are then interpreted by the researcher – so the example topic combinations here are simply hypothetical. Researchers typically inspect numbered topics (topic 1 . . . topic k), and these topics are then interpreted from their associated words. As we show below, this can result in a compelling intuitive set of topics.

After running the algorithm with a various number of topics, we selected the model of best fit, which contained 20 topics. Example topics are shown in Table 1 with associated terms. Note in the table that we have used a stemmer algorithm to obtain roots (e.g., “compar”, “abil”), to decrease the type-token ratio, and facilitate topic extraction. To further analyze the content, a correlation matrix of the probability distributions for the topics was calculated and a network of positively related topics was generated. Then, a community detection algorithm (Pons & Latapy, 2005) was used to cluster these topics. We found that the algorithm clustered the content of EvoLang submissions broadly into three communities or clusters. The resulting network is shown in Fig. 1, with the different clusters marked by color.

But what do these clusters consist of? To get more insight into the topics associated with each cluster, we extract the most probable terms associated with the

^a Abstracts before 2006 were published in a different format and were thus omitted to keep the data consistent.

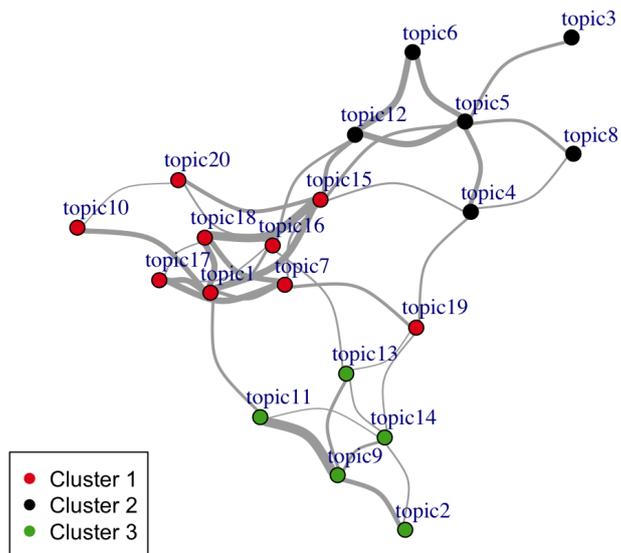


Figure 1. Network of positively correlated topics. The thicker an edge, the stronger the correlation. Topics belonging to the same cluster share a color. An interactive version of this plot is available on <http://shiny.tillbergmann.com/apps/evolang/>

topics in each cluster. The first cluster covers general terms covering linguistics and language evolution, but also more specific topics such as word order in topic 19 (see Table 1). The second cluster is more specific, including comparative studies involving primates and birds, as well as the study of gestures and music (Table 2). Lastly, papers in the third cluster approach language evolution through cultural evolution and iterated learning, as well as the emergence of structures in communication experiments (Table 3). Inspecting these terms and communities gives a good overview of different fields within EvoLang, and indeed, both the clustering and most probable terms make intuitive sense.

In general, these clusters show that EvoLang hosts a variety of sub-fields,

which approach the study of language evolution from varying angles. Not only does it include more theoretical linguistic work, but also comparative studies are well represented. Certainly this is well known intuitively by researchers within the community, but the analysis here suggests that there are crisp clusters that can be automatically extracted using the topic model. In the next section, we look at the author collaboration networks of EvoLang. This serves both as an illustration of the range of authorship patterns, as well as being the measure through which we further analyze the interconnectedness of these three topic clusters.

Table 1. Terms associated with cluster 1.

Topic 1	Topic 7	Topic 10	Topic 15	Topic 16
languag	semant	evolut	human	symbol
evolut	evolutionari	select	abil	evolutionari
evolv	grammar	extend	language	icon
evolution	syntax	behavior	research	language
language	structur	term	share	protolanguag
framework	approach	factor	compar	sound

Topic 17	Topic 18	Topic 19	Topic 20
system	process	word	signal
evolut	cognit	order	communic
paper	brain	inform	mechan
complex	evolut	divers	behaviour
increas	specif	cue	explain
stage	propos	speaker	provid

Table 2. Terms associated with cluster 2.

Topic 3	Topic 4	Topic 5	Topic 6	Topic 8	Topic 12
modern	question	vocal	speech	song	gestur
present	differ	human	origin	learn	communic
suggest	music	primat	function	development	ape
air	speech	call	involv	genet	studi
evid	pattern	product	action	finch	intent
homo	show	produc	area	complex	system

Table 3. Terms associated with cluster 3.

Topic 2	Topic 9	Topic 11	Topic 13	Topic 14
communic	learn	linguist	emerg	model
studi	mean	cultur	languag	social
game	categori	bias	develop	agent
refer	experi	evolut	form	popul
experiment	structur	languag	sign	network
strategi	iter	learn	languages	interact

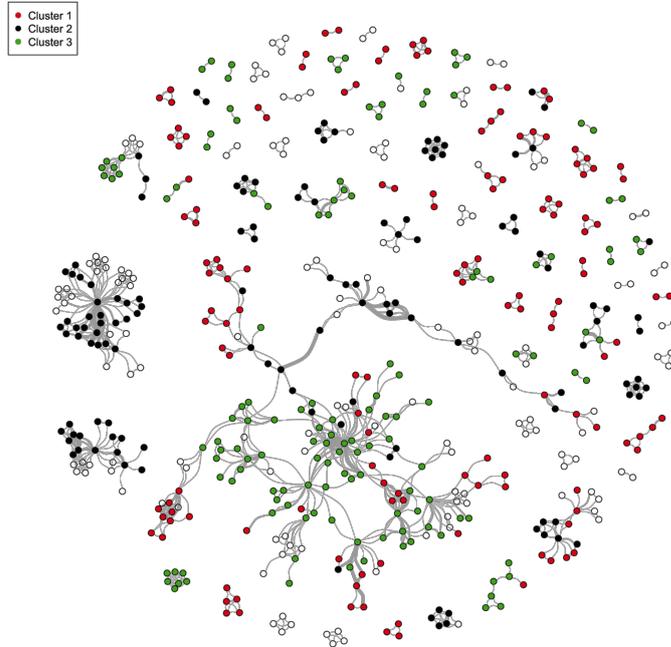


Figure 2. A network showing collaborations between authors. Nodes represent authors and are colored with respect to their dominant cluster. The thicker an edge, the more collaborations between the nodes. An interactive version of this plot is available on <http://shiny.tillbergmann.com/apps/evolang/>.

3. The interconnectedness of authors and clusters

By constructing an authorship network from co-authored abstracts, we can detect which authors have a high interconnectedness at EvoLang. Authors who publish and collaborate often are referred to as “central,” and by virtue of their centrality, we can also assess the contribution of their associated topics in their collaborations. In this network, each node is an author, and each edge between two nodes represents collaboration between these two nodes/authors. Edge weight (connection strength) is determined by the number of collaborations between these two authors. Using the topic clusters from the above analysis, we calculated the most prevalent cluster for each author, based on which cluster their respective papers were assigned. By plotting the author network (Fig. 2), we can see that there are some hubs in the middle of the network, as well as some collaborations outside these general hubs, not connected to the rest of the network. These smaller

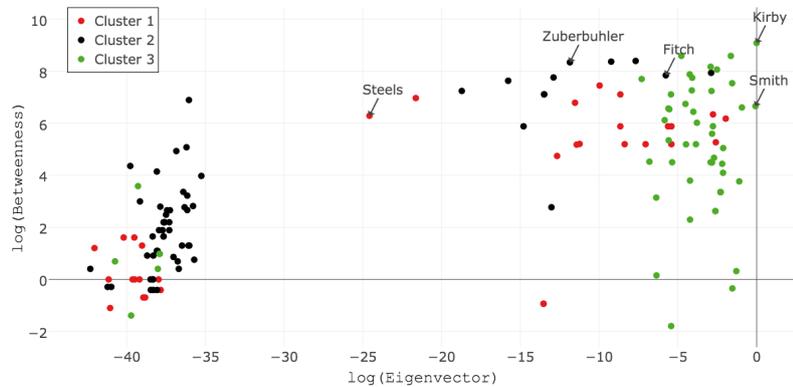


Figure 3. Betweenness and eigenvector centrality, on a log-scale. Each point represents an author, with the color representing their cluster. A few noteworthy authors are labeled. An interactive version of this plot is available on <http://shiny.tillbergmann.com/apps/evolang/>.

collaborations often consist of advisor-advisee relationships within the same lab or department. The color of the nodes represents the respective cluster an author has mainly published in. As not all papers were included in the content analysis due to abstract length, some nodes remain white because their cluster could not be determined. The bigger hubs in the center of the network mainly belong to cluster three, covering the iterated learning framework. Cluster 1 and 2 are more interspersed, and cluster 2 forms its own smaller hubs, showing a strong sense of collaboration in comparative studies.

After constructing the network, centrality measures were used to detect the most influential authors within this network. In network theory, there are multiple ways to measure the centrality of nodes (Freeman, 1978; Koschützki et al., 2005). Here, we look at two values: eigenvector centrality and betweenness centrality. Eigenvector centrality measures the influence of a node by assigning a score based on connections to high scoring nodes (here, nodes with a lot of collaborations and thus submitted papers). The score is bound between 0 and 1, with 1 representing highest centrality. Betweenness centrality assigns a score based on how often the node is part of the shortest path between two other nodes, and thus measures how well a node connects different parts of a network. These nodes are considered to be important in communication between other nodes and keeping the network connected. Fig. 3 shows the centrality measures of authors on a log-scale (purely for illustrating purposes): Authors with high eigenvector values but low betweenness have close contact to important people, while authors with low eigenvector values but high betweenness values serve as valuable connections between nodes.

In the plot, there is a division between authors with a high and low Eigenvector centrality. Authors with a high Eigenvector centrality tend to be in cluster 3, while authors in cluster 1 are more likely to have low Eigenvector centrality. Cluster 2 authors seem to be more interspersed.

Table 4. Summary statistics for each cluster of topics.

Cluster	M(Eigenvector)	SD(Eigenvector)	M(Betweenness)	SD(Betweenness)
1	0.003033	0.01517	63.29	227.6
2	0.002757	0.01482	220.42	782.2
3	0.037867	0.12790	336.23	1113.3

Table 5. Summary of multinomial logistic regression showing log-odds and standard errors.

	<i>Dependent variable:</i>	
	Cluster 1	Cluster 2
Betweenness	-0.001 (0.0003)	0.0001 (0.0002)
Eigenvalue	-21.989* (0.001)	-25.957* (0.001)
Constant	0.275* (0.128)	0.137 (0.132)
Akaike Inf. Crit.	866.008	866.008

By using the centrality measures calculated for each author, we were able to deduce the influence of each topic cluster. That is, to which cluster do the most widely collaborating individuals belong? Table 4 shows summary statistics for the author centrality measures in each cluster. Not surprisingly, cluster 3 has both the highest average eigenvector and betweenness centrality, however, it also has the highest deviations. While the deviations suggest that there is a lot of variation within clusters, it looks like cluster 3 is the most central set of topics within EvoLang.

To test whether this difference in centrality measures is significant, a multinomial logistic regression was run with the clusters as a dependent variable, and the two centrality measures as the independent measures. Cluster 3 was chosen as the baseline community, as we hypothesized that it had higher centrality than the other two clusters. The model output is summarized in Table 5 and was significant compared to a null model ($\chi^2(4) = 44.208, p < 0.0001$). Significance values were calculated using Wald tests. Coefficients for betweenness centrality were not significant (Cluster 1: $p = 0.08$, Cluster 2: $p = 0.59$). However, eigenvector centrality was a significant predictor for both cluster ($p < 0.0001$ for both

clusters). As the log odds are very high, any increase in eigenvector centrality increases the probability of that a paper is in cluster 3.

From this analysis, we conclude that cluster 3, which appears strongly related to iterated learning and cultural evolution, serves as a “hub cluster” within EvoLang. However, as the betweenness centrality was not a significant predictor of cluster/framework, authors within each cluster serve as an important connection between other authors, and clusters as a whole.

4. Summary

We analyzed the content of abstracts presented at EvoLang. Our analysis of latent topics shows that EvoLang is an interdisciplinary conference, and seems to draw from three major clusters of topics. Using a network analysis of author collaborations, we investigated these clusters with regard to their influence. Our results suggest that the iterated learning and cultural evolution framework is associated with a high centrality property within EvoLang. Comparative studies with primates are an important interconnector between authors and communities, while the cluster covering linguistic approaches is interspersed and well represented throughout the conference. The interconnectedness of the author network suggests that each cluster draws inspiration from each other, and that in fact no single framework – according to the LDA topic model – is isolated from any other.

Though these patterns may be intuitive to highly initiated attendees of the conference, the purpose of this paper is to demonstrate that scientometric techniques can be used to reveal these patterns quantitatively. With just under 400 abstracts, a number of natural authorship and conceptual patterns emerge. It may be useful and interesting to carry out similar analyses in subsequent years to discern how this field is changing, and how topic clusters may be converging or co-fertilizing.

References

- Bickerton, D. (2007). Language evolution: A brief guide for linguists. *Lingua*, 117(3), 510–526.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Christiansen, M. H., & Kirby, S. (2003). Language Evolution: The Hardest Problem in Science? In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 1–15). Oxford: Oxford University Press.
- Fitch, T. W. (2010). *The Evolution of Language*. Cambridge, MA: Cambridge University Press.
- Freeman, L. C. (1978). Centrality in Social Networks: Conceptual Clarification. *Social Networks*, 1(3), 215–239.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *PNAS*, 101, 5228–5235.

- Koschützki, D., Lehmann, K. A., Peeters, L., Richter, S., Tenfelde-Podehl, D., & Zlotowski, O. (2005). Centrality Indices. In U. Brandes & T. Erlebach (Eds.), *Network Analysis* (pp. 16–61). Berlin and Heidelberg: Springer.
- Pons, P., & Latapy, M. (2005). Computing Communities in Large Networks Using Random Walks. In P. Yolum, T. Güngör, F. Gürgeç, & C. Özturan (Eds.), *Computer and Information Sciences - ISCIS 2005* (Vol. 3733, p. 284-293). Berlin and Heidelberg: Springer.
- Yau, C. K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics*, *100*(3), 767–786.

USING CAUSAL INFERENCE TO DETECT DIRECTIONAL TENDENCIES IN SEMANTIC EVOLUTION

JOHANNES DELLERT

*Seminar für Sprachwissenschaft, Universität Tübingen
Tübingen, Germany
jdellert@sfs.uni-tuebingen.de*

This paper proposes a novel application of causal inference in the area of semantic language evolution, which attempts to infer unidirectional trends of lexical change exclusively from massively cross-linguistic dictionary data. First, we show how colexification between concepts can be modeled mathematically as mutual information between concept variables. Core notions of causal inference (most prominently, the unshielded collider criterion) are then applied to predict the dominant directionality in pathways of semantic change. The paper concludes by revisiting a few well-known examples of synchronic polysemies, and illustrating how the method succeeds in building hypotheses about their historical development.

1. Introduction

As a language evolves over time, all parts of the language system undergo constant and sometimes drastic change. Compared to phonological and syntactic change, semantic change is much more subject to historical developments and accidents, making it notoriously erratic and hard to predict (Hollmann, 2009). Still, the long search for patterns which allow to predict at least some kinds of semantic change has started to yield some promising results. Most prominently, Traugott and Dasher (2002) manage to establish a general tendency for words to evolve towards more speaker-oriented meanings.

If for a given concept, we want to predict how the words for that concept are likely to evolve, we have the problem that extralinguistic factors can strongly influence each individual case, and can easily invert the expected course of events. For instance, while it is quite common for words denoting large crops to be used colloquially as words for the concept HEAD:N (e.g. French *citron* “lemon”, German *Birne* “pear”, and Russian *repa* “turnip”), the reverse development has occurred in Thai, where the word *hua* “head” has become the count noun for large fruit.

Therefore, we can always only talk about tendencies, and a typical question we might hope to answer has the following form: Is a word denoting the concept X more likely to additionally develop the sense Y, than a word denoting Y is to develop the sense X? For the concepts SUN:N and DAY:N, it seems obvious that

semantic change will always occur in one direction. But are such unidirectional patterns exceptional, or do they occur along many pathways in conceptual space? Previous attempts to identify such tendencies have relied on intimate knowledge of historical developments within a small number of languages, e.g. English and Japanese in the case of Traugott and Dasher (2002). Such a small language sample detracts from the reliability of broad generalizations, a problem which concerns more recent computational approaches such as Sagi, Kaufmann, and Clark (2011) as well. If we want to come to more general conclusions, we will need to consider a wider range of languages, as did classical work in lexical typology like Viberg (1984) or Wilkins (1996) for small semantic domains.

To gather evidence for and against a postulated unidirectional change, we could extensively study all available historical sources and etymological literature for a large sample of languages. Unfortunately, only very few languages and even fewer language families are sufficiently well-documented over relevant timespans. Moreover, the semantic reconstructions presented in etymological dictionaries are often speculative, and therefore cannot be used as empirical evidence.

Evans and Wilkins (2000) propose synchronic polysemies as a more readily available source of evidence for semantic evolution, and argue that cross-linguistic polysemies should be used systematically to constrain the semantic reconstructions postulated in etymologies. Analyses of semantic change processes have established (Zalizniak et al., 2012) that a word undergoing semantic change will typically pass an intermediary polysemous stage where it denotes both the original and the newly developing sense. Synchronic polysemies thus provide us with a snapshot of semantic evolution in action, which we can exploit as observational data in order to retrieve hints about tendencies in their historical development.

With massively cross-linguistic lexical databases slowly becoming available, we can use the power of statistical methods to leverage large amounts of polysemy data for inferring new general tendencies in semantic evolution, or to gather additional evidence in favor of or against existing theories. The purpose of this work is to show that causal inference has the potential to become such a method.

2. Causal Inference

Causal inference (Pearl, 2009) is a relatively recent branch of statistics which has developed partial solutions to the classical problem that correlation between two statistical variables does not imply a causal relationship. By taking additional variables and their covariance patterns into account, the existence and the direction of immediate causation between statistical variables can often be inferred from observational data alone, given some reasonable background assumptions.

The central idea of causal inference is to exploit patterns of conditional independence. A correlation between two variables may vanish when conditioning on other variables, i.e. considering every combination of their possible values separately. The pattern in which some correlations disappear upon conditioning on

sets of other variables provides hints which help us to exclude some causal hypotheses, sometimes leaving only one possible direction of causation.

Causal inference depends on a stability condition which is equivalent to postulating that a true causal connection between two variables can safely be assumed to exist if the dependence between them cannot be explained away by observing any subset of the remaining variables. Applying this criterion in a principled way to a larger number of variables gives us what is called a **causal skeleton**, an undirected graph linking all pairs of variables whose dependence cannot be explained away by conditioning on other possibly intervening variables.

The key to turning such a skeleton into a partially directed causal graph is to consider **unshielded triples**, i.e. triples of variables of the form $A - B - C$. Consider the different conditional independence patterns one would expect for all possible causal patterns. If the true pattern is $A \leftarrow B \rightarrow C$, we would expect some correlation between A and C which disappears when conditioning on the common cause B . An analogous argument applies to the cases $A \rightarrow B \rightarrow C$ and $A \leftarrow B \leftarrow C$. By contraposition, we can thus infer the **unshielded collider** $A \rightarrow B \leftarrow C$ if conditioning on B was not necessary to explain away any possible correlation between A and C . This pattern of reasoning lies at the heart of causal inference algorithms such as the PC algorithm (Spirtes, Glymour, & Scheines, 2000), and informs the criterion we will use to infer the directionality of lexical change.

3. Measuring Conditional Independence between Concepts

In order to apply causal inference to the domain of semantic evolution, we need a model where language-independent concepts play the role of variables. These variables are represented by German (or English) glosses, and the observations are realizations of these concepts across a large number of languages. If we do this for two related concepts, there will be some languages where the two concepts are **colexified**, i.e. they can be denoted by the same polysemous lexeme. For instance, the concepts SUN:N and DAY:N are colexified because the Hungarian word *nap* denotes both senses (as do equivalents in many other languages).

Assuming a conceptual space which is given as a set of language-independent senses, the subset which can be expressed by a word (or lexeme) w in some language is called the **isolectic set** of that word (François, 2008). The isolectic set of Hungarian *nap* could be represented as {SUN:N, DAY:N, SOLAR:A}.

For ease of exposition, we assume a one-to-one-mapping from German (or English) glosses to language-independent concepts. In my data, each “concept” is defined by a single gloss in German. In what follows, “colexification” therefore means that two German glosses occur together on one side of a dictionary equation, and could more accurately be called co-translation into German. Because polysemies represent intermediary stages of semantic evolution which we do not expect to be stable over time, it is an acceptable simplification to count every colexification equally, without correcting for genealogical relatedness.

We now turn to measuring the connectedness of concepts based on overlaps of isolectic sets across many languages. Mathematically, we will model dependence between two concepts in terms of non-vanishing mutual information. Isolectic sets will be our observations, and the mutual information of variables for colexified concepts will be non-zero. The joint information measure R over sets of variables will be designed to be a **submodular information measure**, which means that it meets the following axioms:

1. $R(\emptyset) = 0$
2. $S \subseteq T \Rightarrow R(S) \leq R(T)$ for all sets of variables S and T
3. $R(S) + R(T) \geq R(S \cup T) + R(S \cap T)$ for all sets of variables S and T

Every submodular information measure gives rise to a measure of conditional mutual information which still has all the essential properties needed for causal inference (Steudel, Janzing, & Schölkopf, 2010).

To measure joint information in a set of concepts $\{c_1, \dots, c_n\}$, we use a very simple and trivially submodular measure R based on the sets $iso(c_i)$ of isolectic sets containing each concept c_i :

$$R(c_1, \dots, c_n) := \left| \bigcup_{i=1}^n iso(c_i) \right|$$

This is mathematically equivalent to a measure in Steudel et al. (2010, Section 5.4), which is defined analogously on sets of content words, and presented as a way to measure causal influences between texts.

Informally, the joint information content R of a set of concepts is thus the number of isolectic sets in which these concepts are involved. If the isolectic sets of two different lemmas have exactly the same elements, they are still counted separately. Whenever an isolectic set covers multiple concepts from the set $\{c_1, \dots, c_n\}$, $R(c_1, \dots, c_n)$ will be different from the sum $\sum_{i=1}^n iso(c_i)$. It is this difference that the resulting measure of mutual information will quantify. In the case of two concepts, the derived mutual information $i(c_i, c_j)$ simply counts the number of isolectic sets in which the two concepts c_i and c_j co-occur, i.e. the number of colexifications in the data:

$$\begin{aligned} i(c_i, c_j) &:= R(c_i) + R(c_j) - R(c_i, c_j) \\ &= |iso(c_i)| + |iso(c_j)| - |iso(c_i) \cup iso(c_j)| = |iso(c_i) \cap iso(c_j)| \end{aligned}$$

For instance, if our entire dataset consisted of three lemmas with the isolectic sets $\{\text{SUN:N, DAY:N}\}$, $\{\text{SUN:N, SOLAR:A}\}$, and $\{\text{SUN:N, DAY:N, SOLAR:A}\}$, we would have $i(\{\text{SUN:N, DAY:N}\}) = 3 + 2 - 3 = 2$.

Conditional mutual information between two concepts c_i and c_j given a set of concepts $S := \{s_1, \dots, s_n\}$ is then defined in the following way:

$$i(c_i, c_j; S) := R(c_i, s_1, \dots, s_n) + R(c_j, s_1, \dots, s_n) - R(c_i, c_j, s_1, \dots, s_n) - R(s_1, \dots, s_n)$$

The submodularity of the information measure R ensures that $i(c_i, c_j; S)$ is always nonnegative (Steudel et al., 2010, Lemma 1).

Intuitively, $i(c_i, c_j; S)$ counts the colexifications between c_i and c_j which cannot be explained away by colexification with any of the concepts in S . In our example, we have $i(\{\text{SUN:N, DAY:N; SOLAR:A}\}) = 2 + 2 - 2 - 1 = 1$, but $i(\{\text{DAY:N, SOLAR:A; SUN:N}\}) = 2 + 2 - 2 - 2 = 0$, which means that DAY:N and SOLAR:A are independent given SUN:N, and we get the unshielded triple DAY:N–SUN:N–SOLAR:A in the causal skeleton.

Turning to the question how to detect the directionality of the established causal links (or possible paths of semantic evolution), we reconsider the intuition behind the unshielded collider criterion. To infer a causal arrow $c_1 \rightarrow c_2$, we need a third concept c_3 which forms an unshielded collider $c_1 \rightarrow c_2 \leftarrow c_3$. This means that if c_1 and c_3 are colexified, none of the isolectic sets in question must extend only to c_2 , but there needs to be at least one other concept c_s on which we can condition to remove the link between c_1 and c_3 . Whenever we see such a configuration, it becomes more likely that words for c_2 were extended to cover the sense c_1 , because in every other scenario we would expect some of these words to also have been extended to c_3 .

For instance, if words for UNDERSTAND:N were likely to evolve into words for HEAR::V, the colexification between SEE::V and UNDERSTAND:N would cause some isolectic areas to cover all three concepts. The absence of such isolectic areas provides us with evidence that the true pattern is much more likely to be SEE::V \rightarrow UNDERSTAND::V \leftarrow HEAR::V.

If we check for many different concepts c_3 whether they form unshielded colliders together with c_1 and c_2 , the small sample size will often lead to conflicting evidence, i.e. some unshielded colliders which imply $c_1 \rightarrow c_2$ and others which imply $c_2 \rightarrow c_1$. In such a case, a scoring scheme can be used to decide whether one of the directions is more probable. The current version of my implementation simply weights each arrow $c_1 \rightarrow c_2$ for each concept c_3 where $R(c_1, c_3) > 1$ or $R(c_2, c_3) > 1$ by the factor $w(c_1 \rightarrow c_2; c_3) := \frac{R(c_1, c_2) \cdot R(c_3, c_2)}{R(c_2)}$, i.e. the number of colexifications between c_1 and c_3 which we would have expected if the causal pattern were $c_1 \leftarrow c_2 \rightarrow c_3$ or $c_1 \leftarrow c_2 \leftarrow c_3$. If the summed arrow score $sc(c_1 \rightarrow c_2) := \sum_{c_3} w(c_1 \rightarrow c_2; c_3)$ is more than 20% higher than $sc(c_2 \rightarrow c_1)$, the current implementation returns the arrow $c_1 \rightarrow c_2$.

4. Examples

The ideal way to evaluate the method’s potential would be to collect a large number of clear-cut example cases where the etymological literature across language families only reconstructs semantic evolution in one direction. Unfortunately, even the largest available database of semantic shifts (Zalizniak et al., 2012) only contains very few instances of multiply attested unidirectional semantic shifts, and many of these involve very specialized concepts which one cannot expect to find across many dictionaries (e.g. *amber*, *catnip*, or *woodlouse*).

Studies on cross-linguistic patterns of semantic change such as Koch (2008) do yield some more useful examples, but they typically specialize on very small semantic domains. It is thus extremely difficult to find enough examples of cross-linguistically attested unidirectional semantic shifts for a numerical evaluation in terms of precision and recall. Considering only some very prominent examples instead, it is at least possible to illustrate how the inference method works, and to highlight both the power and the inadequacies of causal inference in this application. To ensure reproducibility of these initial results, isolectic sets for all three examples, and log files allowing to trace the computations in detail, are available as supplementary materials on the author’s webpage.

4.1. Example 1: The Eye of a Needle

Consider the three concepts EYE, EAR, and EYEOFNEEDLE. Some languages use the same word for EYE and EYEOFNEEDLE, as English does. Other languages like Polish and Korean use the word for EAR to denote the same concept. In either case, it is clear that the words for body parts are used by metaphorical extension to describe an oblong hole in a needle, and any method for analysing semantic shifts should infer that semantic evolution will occur exclusively in the directions $\text{EAR} \rightarrow \text{EYEOFNEEDLE}$ and $\text{EYE} \rightarrow \text{EYEOFNEEDLE}$.

In this simple case, it turns out that both of the the desired causal arrows can be derived from a single unshielded collider $\text{EYE} \rightarrow \text{EYEOFNEEDLE} \leftarrow \text{EAR}$. Table 1 displays some of the relevant isolectic sets. The collider is trivial to infer because

Table 1. Isolectic sets spanning EYEOFNEEDLE:N and EYE:N or EAR:N.

Language	Lemma	Isolectic Set
Basque	<i>begi</i>	{EYE:N, KNAG:N, EYEOFNEEDLE:N, STITCH:N, DROPOFGREASE:N, CHEESEHOLE:N}
Dutch	<i>oog</i>	{EYE:N, LOOP:N, EYEOFNEEDLE:N}
Korean	<i>gwi</i>	{EAR:N, SPOUT:N, CORNER:N, EYEOFNEEDLE:N}
Livonian	<i>silma</i>	{EYE:N, LOOP:N, SHACKLE:N, EYEOFNEEDLE:N}
Nenets	<i>xa</i>	{EAR:N, HANDLE:N, EYEOFNEEDLE:N}
Polish	<i>ucho</i>	{EAR:N, HANDLE:N, EYEOFNEEDLE:N}

there is no lemma in any language which covers both EAR and EYE. We therefore

have an unshielded triple where EYEOFNEEDLE is not needed to separate EAR and EYE, because these two concepts are already independent unconditionally. To illustrate the causal argument, for any other causal pattern the expected number of colexifications $w(c_1 \rightarrow c_2; c_3)$ between $c_1 = \text{EYE}$ and $c_3 = \text{EAR}$ is 2.377, as opposed to zero colexifications in the data.

Overall, building on 130 isolectic sets from 77 languages across 19 families, the algorithm returns $\text{EYE} \rightarrow \text{EYEOFNEEDLE}$ with a score ratio of 1.269, and for $\text{EAR} \rightarrow \text{EYEOFNEEDLE}$, the score ratio based on 112 isolectic sets from 76 languages across 20 families is as high as 2.765. In both cases, we get a clear result in favor of the expected directionality, showcasing that the method is able to detect a causal signal representing widespread metaphorical extension.

4.2. Example 2: Counting and Calculating

Many languages use the same word for the concepts of counting and calculating. Instances of isolectic sets subsuming both concepts are given in Table 2. From a historical perspective, the counting sense arguably is the earlier one, calculation being a later cultural achievement. Is this fact mirrored by the causal signal we can detect from synchronic polysemies? Based on 134 isolectic sets from 68

Table 2. Isolectic sets spanning COUNT:V and CALCULATE:V.

Language	Lemma	Isolectic Set
Coptic	<i>op</i>	{COUNT:V, CALCULATE:V, ESTIMATE:V}
Czech	<i>počítat</i>	{COUNT:V, CALCULATE:V}
Indonesian	<i>membilang</i>	{COUNT:V, CALCULATE:V, NARRATE:V}
Udmurt	<i>lydjany</i>	{COUNT:V, CALCULATE:V}
Spanish	<i>contar</i>	{COUNT:V, CALCULATE:V, NARRATE:V}

languages covering 21 language families, we get a score ratio of 1.162 in favor of $\text{COUNT:V} \rightarrow \text{CALCULATE:V}$. The evidence is thus not strong enough to make a decision, although it does point into the expected direction. Investigation of the isolectic sets in question shows that the signal would be quite a bit stronger if we could remove the effects of polysemy in the German gloss *zählen* “to count”, which is also used in the senses “to be valid” and “to have (inhabitants)”.

4.3. Example 3: Hoping and Expecting

Finally, we take a look at a pair of verbal concepts whose semantic relationship is not clear on external grounds. Between the concepts of HOPE:V and EXPECT:V (which are frequently colexified), is there a unidirectional pattern we can retrieve from the data? Existing theories appear not to make any prediction here, since both concepts refer to mental states, meaning that they belong to roughly the same level of abstraction. Table 3 gives a number of polysemous verbs which synchronically denote both HOPE:V and EXPECT:V.

Table 3. Isolectic sets spanning HOPE:V and EXPECT:V.

Language	Lemma	Isolectic Set
Chinese	<i>xīwàng</i>	{HOPE:V, EXPECT:V, WISH:V}
Hebrew	<i>jixel</i>	{EXPECT:V, HOPE:V}
Japanese	<i>nozomu</i>	{EXPECT:V, HOPE:V, WISH:V}
Portuguese	<i>esperar</i>	{HOPE:V, EXPECT:V, WAIT:V}
Turkish	<i>ummak</i>	{HOPE:V, EXPECT:V, WAIT:V}

Perhaps surprisingly, the result is very strong. Based on 203 isolectic sets from 70 languages covering 22 families, the score ratio is 2.813 in favor of HOPE:V \rightarrow EXPECT:V. The prediction is that words for HOPE:V are more likely to develop the additional sense EXPECT:V than the reverse pattern. This result seems plausible in the light of examples known to the author, e.g. the attested development of Latin *spērāre* into Spanish *esperar*, or the Finnish verb *toivoa* “to hope”, which is also used e.g. for expectations from a person. The algorithm has provided us with a hypothesis based on massively cross-linguistic data, and we could now look into individual language histories to verify this claim, or to find counter-examples.

5. Conclusion

In this paper, we have established that causal inference can be applied to an information geometry defined by cross-linguistic polysemies in order to measure causal influences between concepts. The resulting causal structures can be taken to indicate probable vectors of semantic expansion. For three examples, we have seen that the method does manage to recover some hidden diachronic information from purely synchronic polysemies, and leads to plausible results.

Since much larger sample sizes are typically needed to guarantee correct results in causal inference, the approach must not be interpreted as producing objective proofs of historical events. However, the method does provide an unbiased summary of large amounts of easily available data which are too varied and extensive to be processed by a human expert. It allows us to quickly derive interesting hypotheses about possible directional patterns of semantic evolution, which can then be verified and further elucidated based on the documented history of various languages. Moreover, the new computational tool for quickly developing initial hypotheses about the directionality of semantic shifts will be helpful for researchers seeking to shed more light on this central aspect of language evolution.

Acknowledgments

This research has been supported by the ERC Advanced Grant 324246 EVOLAEMP, which is gratefully acknowledged. The author wishes to thank all members of the EVOLAEMP group for their valuable comments, and three anonymous reviewers for their helpful suggestions.

References

- Evans, N., & Wilkins, D. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, 76(3), 546–592.
- François, A. (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In M. Vanhove (Ed.), *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations* (pp. 163–215). Amsterdam: Benjamins.
- Hollmann, W. B. (2009). Semantic change. In J. Culpeper, F. Katamba, P. Kerwill, & T. McEnery (Eds.), *English language: description, variation and context* (p. 301–312). Basingstoke: Palgrave.
- Koch, P. (2008). Cognitive onomasiology and lexical change: Around the eye. In M. Vanhove (Ed.), *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations* (pp. 107–137). Amsterdam: Benjamins.
- Pearl, J. (2009). *Causality*. Cambridge University Press.
- Sagi, E., Kaufmann, S., & Clark, B. (2011). Tracing semantic change with Latent Semantic Analysis. In K. Allan & J. A. Robinson (Eds.), *Current Methods in Historical Semantics* (pp. 161–183). Berlin, Boston: De Gruyter.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search* (2nd ed.). MIT Press.
- Studel, B., Janzing, D., & Schölkopf, B. (2010). Causal markov condition for submodular information measures. In A. Kalai & M. Mohri (Eds.), *Proceedings of the 23rd annual conference on learning theory* (pp. 464–476). Madison, WI, USA: OmniPress.
- Traugott, E., & Dasher, R. (2002). *Regularity in semantic change*. Cambridge University Press.
- Viberg, A. (1984). The verbs of perception: a typological study. *Linguistics*, 21(1), 123–162.
- Wilkins, D. P. (1996). Natural tendencies of semantic change and the search for cognates. In M. Durie & M. Ross (Eds.), *The Comparative Method reviewed: Regularity and irregularity in language change* (pp. 264–304). New York: Oxford University Press.
- Zalizniak, A. A., Bulakh, M., Ganenkov, D., Gruntov, I., Maisak, T., & Russo, M. (2012). The catalogue of semantic shifts as a database for lexical semantic typology. *Linguistics*, 50(3), 633–669.

THE EMERGENCE OF THE PROGRESSIVE TO IMPERFECTIVE DIACHRONIC CYCLE IN REINFORCEMENT-LEARNING AGENTS

DANKMAR ENKE¹, ROLAND MÜHLENBERND², IGOR YANOVICH³

¹*Institute of German Philology, Ludwig-Maximilian-University, Munich, Germany*

^{2,3}*Department of Linguistics, Eberhard Karls University, Tübingen, Germany*

³*Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA, US*

¹*dankmar.enke@lmu.de, ²roland.muehlenbernd@uni-tuebingen.de,*

³igor.yanovich@uni-tuebingen.de

Deo (2015) offers a model within the framework of evolutionary game theory for the analysis of an attested phenomenon in semantic change: the progressive to imperfective cycle of shifts. While Deo studies the evolutionary dynamics of four preselected types of progressive-imperfective grammars, we investigate which types of grammars would emerge from the first principles in a population of agents under reinforcement learning. In our model, the actual progressive-to-imperfective cycle arises from such atomic interactions between learner agents after the addition of several simple assumptions to the basic game-theoretic model. The most important such addition concerns the problem of why the progressive but never the habitual generalizes to the broad imperfective. Deo (2015) conjectured that this might be due to children being more frequently exposed to progressive-type contexts than habitual-type ones. Our model vindicates Deo's conjecture: early asymmetrical exposure derives the asymmetry between the progressive and the habitual, wherein only the former gives rise to a diachronic cycle.

1. Introduction

It is a well-known typological observation that languages without a distinct progressive (PROG) morphology realize the communicative function of the PROG through the imperfective (IMP) aspect (if morphologically instantiated). This primarily motivates treating the PROG as a subdomain of the IMP (cf. Comrie, 1976). In Russian, the imperfective form licenses a PROG interpretation, while the same form refers to a habitual/generic (HAB/GEN) situation. In languages which have both the PROG and the IMP aspects, the IMP often does not license a PROG reading, such as in English. However, in languages with a less grammaticalized PROG marker, as in German, Dutch, or Shakespearean English, IMP still allows PROG interpretations. A crosslinguistically robust generalization is as follows: functional elements restricted to PROG reading semantically generalize to license IMP readings such as the HAB/GEN or the stative. This generalization has been attested according to data from, e.g., Turkish (Göksel & Kerslake, 2005,

- (1) a. saat ikide çalış-iyor-du-m
At two o'clock work-PROG-PST.COP-1.SG
'A two o'clock I was working.'
- b. genekkikle iki saat çalış-ir-di-m
Usually for two hours work-IMP-PST.COP-1.SG
'I would usually work for two hours.'
- (2) sen Ömer'i benden daha iyi tan-iyor-du-n
you Omer me better than know-PROG-PST.COP.2.SG
'You knew (lit. were knowing) Ömer better than me.'

p. 331), as shown in (1) and (2). The verb form with PROG *-(I)yor* in (1a) refers to an ongoing eventuality, while the inflected verb with IMP *-(I)r* in (1b) refers to a HAB reading. Recently, the PROG *-(I)yor* has begun to license a wider range of readings, notably in everyday language. (2) shows that *-(I)yor* occurs with a stative verb 'know'. Such data indicate that the Turkish PROG is expanding to semantically overlap with the domain of the IMP Aorist *-(I)r*, thus instantiating the PROG-to-IMP shift (Bybee, Perkins, & Pagliuca, 1994).

2. Progressive-to-Imperfective Cycle

Such typological data motivate positing a cyclic diachronic process (Table 1). This cycle starts with the language having only one broad imperfective form covering all imperfective meanings, (a). Then an optional progressive form is innovated, (b); it becomes obligatory for progressive meanings, (c); and at the last stage, (d), it generalizes and takes the semantic place of the old broad-imperfective form. Note that (a) and (d) are identical except for the formal exponents of IMP.

The four states (a-d) can be intuitively regarded as distinct *strategies* for communicating *phenomenal* (facts of local import, pertaining to specific times) and *structural* (stable facts that characterize the world as a whole) sub-meanings Goldsmith and Woisetschlaeger (1982) within the imperfective domain. In systems with two forms, namely emergent-PROG and categorical-PROG, the choice of form helps the hearer to correctly identify the speaker's intended sub-meaning. The zero-PROG and generalized-PROG strategies use a single form while relying on the hearer's understanding of contextual cues for successful communication.

Importantly, PROG induces a cycle through (a-d), but habitual HAB, though also being more specific than the broad imperfective IMP, does not eventually generalize to IMP, Deo (2015). In other words, there is no (d)-type stage for HAB, and therefore no HAB-to-IMP cycle.

Deo (2015) sets up an evolutionary game-theoretic model for studying the dynamics of these systems. She defines a simple game representing individual interactions of agents with different grammars, in the tradition of game-theoretic models of communication (cf. Benz, Jäger, and van Rooij (2006) for an overview.) Deo then lifts this game to an evolutionary game that represents the dynamics of

Table 1. The PROG-to-IMP historical cycle and sample languages.

	form(s)	strategy type	sample languages
(a)	X_{imp}	zero-PROG	Russian, Arabic
(b)	$(Y_{prog})X_{imp}$	emergent-PROG	German, Dutch
(c)	Y_{prog}, X_{imp}	categorical-PROG	English, Swahili
(d)	Y_{imp}	generalized-PROG	Turkish, Tigre

grammar retention and change over long temporal horizons. The goal of this is to model the cyclic semantic shift as in Table 1.

Both the basic and the evolutionary games are defined by Deo only for four strategies (i.e. grammars) corresponding to (a-d) above, without considering any other possibilities. In this paper, we investigate systems with PROG, IMP and HAB at a more granular level, without hardcoding the desired strategies. We consider populations of agents that build their own grammars through reinforcement learning on the basis of speaker-hearer interactions with other agents in the model. In particular, we investigate (i) which strategies actually arise, and what assumptions would rule out the ones that are not empirically observed in human languages, as well as (ii) how to induce the PROG-to-IMP, but crucially not the HAB-to-IMP cycle. Thus, we effectively provide microfoundations for Deo’s macro model of the progressive-imperfective cycle.

3. Deo’s Model

Deo’s original model consists of two parts. First, the *Basic Imperfective Game* defines communicative success in interactions between speakers of different IMP-PROG grammars. Second, the *Evolutionary Imperfective Game* models what happens in large populations of speakers who communicate according to the model of the Basic Imperfective Game, using a replicator-mutator rule for the evolution of aggregate shares of selected grammars in the population. We instead plug the Basic Imperfective Game into a reinforcement-learning framework, deriving evolutionary behavior directly from atomic interactions. The rest of this section describes the setup of the Basic Imperfective Game, without the restriction to particular strategies. For more details on the behavior of the Evolutionary Imperfective Game, we refer the reader to Deo (2015) and Yanovich (2015).

3.1. The Basic Imperfective Game

The *Basic Imperfective Game* is modeled as a *signaling game*^a Lewis (1969), a game-theoretic model that depicts the communication situation between a speaker S and a hearer H . The Basic Imperfective Game can be given as $BIG = \langle (\mathbf{S}, \mathbf{H}), C, T, F, P, U_S, U_H \rangle$. $T = \{t_s, t_p\}$ a set of two states to be signalled (s (*structural*) and p (*henomenal*)). Signals are $F = \{f_{pr}, f_{im}\}$ (note that indices,

^aThe signaling game model proved itself useful for the analysis of phenomena in language change (cf. Ahern, 2014; Jäger, 2007, 2008; Quinley & Mühlennend, 2012).

Table 2. All possible 16 speaker (left) and 16 hearer (right) strategies. Of those, Deo (2015) considers only 4 speaker and 3 hearer strategies (highlighted in gray).

	C_p	C_p	C_s	C_s		C_p	C_p	C_s	C_s
	t_p	t_s	t_p	t_s		f_{pr}	f_{im}	f_{pr}	f_{im}
S_0	<i>f_{im}</i>	<i>f_{im}</i>	<i>f_{im}</i>	<i>f_{im}</i>	H_0	t_p	t_p	t_p	t_p
S_1	<i>f_{im}</i>	<i>f_{im}</i>	<i>f_{im}</i>	f_{pr}	H_1	t_p	t_p	t_p	t_s
S_2	<i>f_{im}</i>	<i>f_{im}</i>	f_{pr}	<i>f_{im}</i>	H_2	t_p	t_p	t_s	t_p
S_3	<i>f_{im}</i>	<i>f_{im}</i>	f_{pr}	f_{pr}	H_3	t_p	t_p	t_s	t_s
S_4	<i>f_{im}</i>	f_{pr}	<i>f_{im}</i>	<i>f_{im}</i>	H_4	t_p	t_s	t_p	t_p
S_5	<i>f_{im}</i>	f_{pr}	<i>f_{im}</i>	f_{pr}	H_5	t_p	t_s	t_p	t_s
S_6	<i>f_{im}</i>	f_{pr}	f_{pr}	<i>f_{im}</i>	H_6	t_p	t_s	t_s	t_p
S_7	<i>f_{im}</i>	f_{pr}	f_{pr}	f_{pr}	H_7	t_p	t_s	t_s	t_s
S_8	f_{pr}	<i>f_{im}</i>	<i>f_{im}</i>	<i>f_{im}</i>	H_8	t_s	t_p	t_p	t_p
S_9	f_{pr}	<i>f_{im}</i>	<i>f_{im}</i>	f_{pr}	H_9	t_s	t_p	t_p	t_s
S_{10}	f_{pr}	<i>f_{im}</i>	f_{pr}	<i>f_{im}</i>	H_{10}	t_s	t_p	t_s	t_p
S_{11}	f_{pr}	<i>f_{im}</i>	f_{pr}	f_{pr}	H_{11}	t_s	t_p	t_s	t_s
S_{12}	f_{pr}	f_{pr}	<i>f_{im}</i>	<i>f_{im}</i>	H_{12}	t_s	t_s	t_p	t_p
S_{13}	f_{pr}	f_{pr}	<i>f_{im}</i>	f_{pr}	H_{13}	t_s	t_s	t_p	t_s
S_{14}	f_{pr}	f_{pr}	f_{pr}	<i>f_{im}</i>	H_{14}	t_s	t_s	t_s	t_p
S_{15}	f_{pr}	f_{pr}	f_{pr}	f_{pr}	H_{15}	t_s	t_s	t_s	t_s

while suggestive, by themselves do not have a meaning). Contexts $C = \{c_s, c_p\}$ model the fact that realistically, certain communication contexts highly favor phenomenal or structural meanings. Deo sets to $P(t_s|c_s) = P(t_p|c_p) = .9$. Speaker strategies $S_{\in S} : T \times C \rightarrow F$, and hearer strategies $H_{\in H} : F \times C \rightarrow T$ define mappings from states to signals, and from signals to states respectively. On both the speaker and hearer side, the strategy may take into account the current context drawn from C , hence it is also an argument. Finally, $U_S, U_H : T \times \mathbf{S} \times \mathbf{H} \rightarrow \mathbb{R}$ are the utility functions for the speaker and hearer. The hearer’s function U_H is the δ_t -function of Jäger (2007) that returns 1 if the hearer correctly guessed the state that the speaker signalled (communicative success), and 0 otherwise (failure). The speaker function U_S is similar, but also includes a penalty of k always applying to a speaker who uses a two-form system as opposed to a single-form one. (In our implementation in the learning setting, we drop cost k from the basic model, but manipulate it in more complex models of Experiments III and IV.)

Speaker strategies define which of the two signals f_{im}, f_{pr} is to be used for every one of the four pairs $\langle t, c \rangle$. Thus there are theoretically $2^4 = 16$ different speaker strategies. Similarly, there are 16 hearer strategies that define how to guess state t given a pair of signal f and context c . All possible S and H are shown in Table 2. We highlight the 4 speaker and 3 hearer strategies to which Deo (2015) restricts attention in her modeling. These strategies compose the ‘progressive \gg imperfective cycling path’ (PROG-path) that passes through the speaker strategies $S_0 \rightarrow S_2 \rightarrow S_{10} \rightarrow S_{15}$ and the hearer strategies $H_3 \rightarrow H_1 \rightarrow H_5 \rightarrow H_3$.

In contrast, we will use the full strategy space, and show that there exist conditions under which exactly the PROG-to-IMP diachronic cycle emerges. Pairs

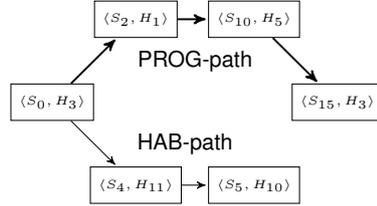


Figure 1. The ‘progressive \gg imperfective’ cycling path (*PROG-path*, thick arrows) and the alternative deadlock path (*HAB-path*, thin arrows).

of strategies which will also play an important role in our modeling also include linguistically unattested $\langle S_6, H_9 \rangle$ (“surprisal” strategy for signalling that the intended state is disfavored by the context), as well as HAB-related pairs $\langle S_4, H_{11} \rangle$ (emergent-HAB) and $\langle S_5, H_{10} \rangle$ (categorical-HAB). The question for us is: under which conditions do the trajectories of change in Fig. 1 emerge, with the cycling PROG-path and deadlocking HAB-path, and no other trajectories do?

4. Strategy Selection: a Game-Theoretic Analysis

In our simulation experiments we consider a population of 20 agents which communicate with each other via the Basic Imperfective Game over the full strategy space in Table 2. Agents have a learning memory and update their behavior via reinforcement learning (c.f. Roth & Erev, 1995)^b. Agents have a maximal age A_{max} which defines the number of rounds of play after which they are replaced by a new agent with an empty memory. We set $A_{max} = 5,000$. To have a heterogeneous ‘age structure’, we initialize the model by randomly assigning ages between 0 and 5,000 to each agent. At the beginning of the simulation, the agent have only one form f_{im} at their disposal. After 1,000 simulation steps^c the second form f_{pr} is introduced. Below, we described 4 lines of experiments we conducted, each subsequent one building more assumptions into the model. Each line consisted of 100 runs under identical conditions.

^bThe reinforcement learning model is implemented as an urn model. Each agent has (i) 4 speaker urns for each context-state combination, and (ii) 4 hearer urns for each context-form combination. Urns contain balls of two types corresponding to two signals. Those encode information about past successes, namely cumulative reward. When agents play a game with each other, they make a probabilistic choice (of form or of guessed state) in dependence of the appropriate urn’s current contents, and afterwards update their urns in dependence of the communicative success. Note that in this model agents (i) play probabilistic strategies, and (ii) do not learn pure strategies as such, but approximate them in the long run. The distance of a probabilistic to a pure strategy can be measured, e.g. by the Hellinger distance (Hellinger, 1909). For ease of exposition, we say that an agent ‘uses’ a particular pure strategy if it is the Hellinger-closest to her current probabilistic strategy.

^cFor each simulation step each agent chosen as a speaker plays the Basic Imperfective Game with a randomly chosen hearer. Context c_i is chosen randomly, and then state t is drawn randomly according to the probability distribution $P(t|c_i)$. After each interaction the agents’ urns are updated.

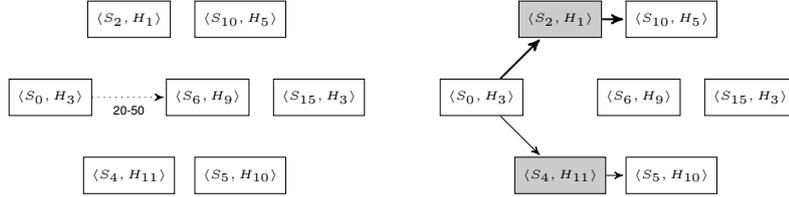


Figure 2. **Left:** Experiment I: after introducing the new form f_{pr} , agent switch directly (20-50 simulation steps) to the stable strategy pair $\langle S_6, H_9 \rangle$ (“surprisal” strategy). **Right:** Experiment II: Agents stabilize on $\langle S_{10}, H_5 \rangle$ (categorical-PROG state of the PROG-path) or $\langle S_5, H_{10} \rangle$ (categorical-HAB state of the HAB-path). Final strategy pair $\langle S_{15}, H_3 \rangle$ is never reached. (Gray: unstable states.)

Experiment I is our baseline, not containing any additional assumptions. Its results are depicted in Figure 2 (left). With only one message f_{im} available during the first 1,000 simulation steps, agents expectedly play $\langle S_0, H_3 \rangle$. After the introduction of the second form f_{pr} , all agents switch rapidly, in about 20-50 simulation steps, to the strategy pair $\langle S_6, H_9 \rangle$. This is the “surprisal” strategy: the new form f_{pr} signals that the intended state is context-unusual. Importantly, this strategy was never observed in human languages for the imperfective domain.

Though the emerging strategy pair $\langle S_6, H_9 \rangle$ is linguistically odd, Experiment II shows that we can rule it out by adding a simple and linguistically natural assumption. In real-life interactions, sometimes the hearer would not be able to observe the speaker’s context c . We model that by randomly withdrawing the context cue in 20% of the interactions. This small change has an tremendous effect, Fig. 2, right. Now “surprisal” $\langle S_6, H_9 \rangle$ never emerged, and instead the population stabilized either on strategy pair $\langle S_{10}, H_5 \rangle$ (the categorical-PROG state of the PROG-path), or on $\langle S_5, H_{10} \rangle$ (categorical-HAB state of the HAB-path). The empirical estimate of the probability for each path to emerge was .5.

There are two aspects that differ between the results of Experiment II and what is considered to be empirically observed diachronic trajectories in the imperfective domain Deo (2015). (1) The emergent-PROG $\langle S_2, H_1 \rangle$ and emergent-HAB $\langle S_4, H_{11} \rangle$ states are only short intermezzos in our model, while in reality they can be maintained for several centuries (e.g., both Shakespeare and Laurence Sterne used emergent-PROG). (2) Both PROG and HAB paths in Experiment II do not go towards the single-form state $\langle S_{15}, H_3 \rangle$. This is empirically correct for HAB, but incorrect for PROG. Here, we leave issue (1) aside, and concentrate on issue (2).^d We divide it into two sub-issues: (2a) how to achieve the simplification of

^dWe conjecture that their instability may be caused by the fact that we sometimes withdraw the contextual cue: unlike the categorical systems, which ignore the cue completely, emergent-PROG and emergent-HAB crucially rely on it. Thus it is not surprising that when the cue is withdrawn, these strategies have a hard time. The question is, what other property of the real-life imperfective communication makes those systems relatively stable?

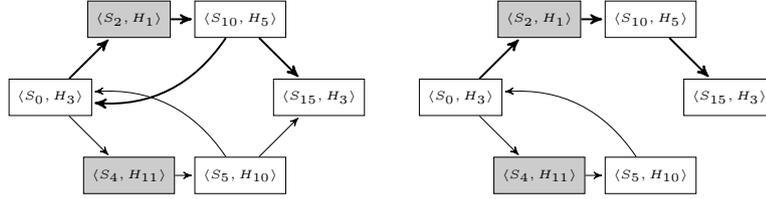


Figure 3. **Left:** Experiment III: The population switches finally to a one-message system, either $\langle S_0, H_3 \rangle$ or $\langle S_{15}, H_3 \rangle$, each equiprobable for both paths. **Right:** Experiment IV: for the HAB-path the population switches back to the initial situation, for the PROG-path the population completes the assumed cycle and switches to the final state $\langle S_{15}, H_3 \rangle$ (gray: unstable states).

a two-form system into a single-form system? (2b) how to derive the asymmetry between the PROG-path and HAB-path?

Consider (2a) first. A two-form system like categorical-PROG or categorical-HAB is perfectly efficient, always achieving communicative success. Why would it then be replaced by a less efficient single-form system? Intuitively, this would also happen if maintaining the efficient two-form system somehow becomes burdensome. Realistically, this could be a result of the “aging” of the old f_{im} , which can be a natural last stage of grammaticalization. If f_{im} becomes less and less suitable for use, there will be an incentive to generalize f_{pr} . Here, we do not test the effects of such a process, studying instead the following modeling alternative: in Experiment III, we gradually increase the cost k of maintaining the two-form system. The effects of this assumption are symmetric for both forms, so we do not hardwire the loss of f_{im} into the model.

Formally, we assign higher cost $c(f)$ to f with lower $fr(f)$, namely the number of interactions an agent encountered f : $c(f) = \alpha \times (1.0 - \frac{fr(f)}{\sum_{f' \in F} fr(f')})$. Note that $\forall f : 0 \leq c(f) \leq \alpha$, so higher α enables higher costs. Note also that a single form is not affected by costs: if $fr(f) = \sum_{f' \in F} fr(f')$ then $c(f) = 0$, no matter how great the α -value is. Thus by increasing the α -value over time, we put two-form systems at increased disadvantage. In Experiment III, we augmented the model of Experiment II with costs defined as above, and increased α by 0.01 after every 1,000 simulation steps. The resulting paths are depicted in Figure 3 (left). Like in Experiment II, the population first stabilizes on categorical-PROG $\langle S_{10}, H_5 \rangle$ or categorical-HAB $\langle S_5, H_{10} \rangle$. But after a while the costs to maintain a two-form system become too high, and the population switches to a one-form system, either $\langle S_0, H_3 \rangle$ (losing f_{pr}) or $\langle S_{15}, H_3 \rangle$ (losing f_{im}). The switch in both directions was equiprobable for both paths.

We turn to (2b). In Experiment III, both PROG and HAB may generalize to a new all-purpose imperfective, and both may be lost in favor of the old IMP form. But we want PROG to generalize, and HAB to never do that. What causes such asymmetry? Deo (2015) conjectures that it might be due to an asymmetry of input

during early language acquisition: “this asymmetry likely stems from the nature of the input to the child, specifically the relative prevalence of PROG forms vs. HAB forms in caregiver speech. [...] this asymmetry in the frequency of phenomenal vs. structural inquiries in child-directed speech would lead to learners generalizing the PROG form rather than any specialized HAB form since exposure to the latter is likely to be less frequent” (Deo (2015, p. 22)). This hypothesis is easy to test in our system: agents of a low age may be presented with state t_s much less often than with t_p . We start with 0 probability of t_s at age 0, and increase it uniformly towards .5 at each step until age C_{max} : $P(t_s|age = a) = \frac{0.5}{\max(C_{max}-a, 1)}$. In Experiment IV, we set $C_{max} = 200$.

This childhood input asymmetry leads to the emergence of exactly the desired trajectories, Fig. 3 (right). From the categorical-HAB state, the system reverts to the initial state after the costs for having two forms rise too high. But if the population enters the PROG-path, then it always generalizes PROG to a new all-purpose imperfective, $\langle S_{15}, H_3 \rangle$, where f_{pr} is the new generalized form. From here on, new emerging forms can realize another cycle of the same shape, and Deo’s conjecture leads to the desired result in our model.

5. Conclusion

We used experiments with reinforcement learning agents playing the Basic Imperfective Game of Deo (2015) with the full strategy space to investigate whether the empirically observed grammar changes involving the imperfective IMP, progressive PROG and habitual HAB would emerge in this setting. With relatively simple assumptions, we achieved the emergence of both the PROG-cycle and the deadlocking HAB-trajectory. The assumptions that we sequentially added to the basic model and their consequences were: (1) the emergence of linguistically odd “surprisal” strategy $\langle S_6, H_9 \rangle$ was prevented by sometimes withdrawing the context cue from the hearer; (2) the switch from a perfectly communicatively efficient two-form grammar calls for a cost function was forced by a gradually increasing symmetric cost for having two forms; and (3) when agents were mostly presented with phenomenal statements in the childhood, the empirically observed PROG-to-IMP cycle emerged, while the unattested hypothetical HAB-to-IMP cycle was ruled out — vindicating a conjecture by Deo (2015).

Acknowledgements

We gratefully acknowledge the support of the ERC under project EVOLAEMP, <http://www.evolaemp.uni-tuebingen.de> (RM), the department of general linguistics at Eberhard Karls University, Tübingen (RM), the Alexander von Humboldt Foundation (IY), Carnegie Mellon University, Philosophy (IY), and the Institute of German Philology of Ludwig-Maximilian-University, Munich (DE).

References

- Ahern, C. (2014). Mergers, migration, and signaling. *University of Pennsylvania working papers in linguistics*, 1(20), Article 2.
- Benz, A., Jäger, G., & van Rooij, R. (2006). An introduction to game theory for linguists. In *Game theory and pragmatics* (p. 1-82). Palgrave Macmillan.
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: tense, aspect, and modality in the languages of the world*. Chicago: Chicago University Press.
- Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.
- Deo, A. (2015). The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift. *Semantics and Pragmatics*.
- Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London: Routledge.
- Goldsmith, J., & Woisetschlaeger, E. (1982). The logic of the English Progressive. *Linguistic Inquiry*, 13(1), 79-89.
- Hellinger, E. (1909). Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik*(136), 210-271.
- Jäger, G. (2007). Evolutionary game theory and typology: A case study. *Language*, 1(83), 74-109.
- Jäger, G. (2008). Applications of game theory in linguistics. *Language and Linguistics Compass*(2/3), 408-421.
- Lewis, D. (1969). *Convention*. Cambridge: Harvard University Press.
- Quinley, J., & Mühlenernd, R. (2012). Conquest, contact, and convention: Simulating the norman invasion's impact on linguistic usage. *Proceedings of BRIMS 2012*, 110-115.
- Roth, A., & Erev, I. (1995). Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games and Economic Behaviour*, 8, 164-212.
- Yanovich, I. (2015). *Analysing imperfective games (the companion paper)*. (Technical report, Universität Tübingen and Carnegie Mellon University. Available at <http://semanticsarchive.net/Archive/jAxYjUzY/>)

KAUFFMAN'S ADJACENT POSSIBLE IN WORD ORDER EVOLUTION

RAMON FERRER-I-CANCHO

*Departament de Ciències de la Computació, Universitat Politècnica de Catalunya
Barcelona, Catalonia (Spain)
rferrericancho@cs.upc.edu*

Word order evolution has been hypothesized to be constrained by a word order permutation ring: transitions involving orders that are closer in the permutation ring are more likely. The hypothesis can be seen as a particular case of Kauffman's adjacent possible in word order evolution. Here we consider the problem of the association of the six possible orders of S, V and O to yield a couple of primary alternating orders as a window to word order evolution. We evaluate the suitability of various competing hypotheses to predict one member of the couple from the other with the help of information theoretic model selection. Our ensemble of models includes a six-way model that is based on the word order permutation ring (Kauffman's adjacent possible) and another model based on the dual two-way of standard typology, that reduces word order to basic orders preferences (e.g., a preference for SV over VS and another for SO over OS). Our analysis indicates that the permutation ring yields the best model when favoring parsimony strongly, providing support for Kauffman's general view and a six-way typology.

1. Introduction

It is widely agreed that cultural evolution is more important than cognitive factors in governing word order change (Dryer, 2011; Dunn, Greenhill, Levinson, & Gray, 2011; Gell-Mann & Ruhlen, 2011). For instance, Dryer wrote that "*Dunn et al. are right that cultural evolution is more important than cognitive factors in governing word order change*" while Gell-Mann & Ruhlen wrote that "*we know of no evidence that SOV, SVO, or any other word order confers any selective advantage in evolution*".

However, word order evolution exhibits some trends: change has been for the most part from SOV to SVO and, beyond that, from SVO to VSO/VOS with a subsequent reversion to SVO occurring occasionally (Gell-Mann & Ruhlen, 2011). Are those trends not reflecting any cognitive bias? Cultural evolution and cognitive biases could be both at play.

Let us consider the case of the principle of dependency length minimization, which predicts that the optimal placement of the verb is at the center, e.g. SVO or OVS (Ferrer-i-Cancho, 2008, 2015). A proponent of that cognitive bias in word order evolution has to solve various puzzles, for instance:

1. Why the transition from SOV to OVS (one of the two orders where the verb

is at the center), is rare (Gell-Mann & Ruhlen, 2011). If that principle is strong enough to pull SOV towards SVO it should be able to do the same job for SOV towards OVS.

2. If the principle is so strong to drive word order change towards SVO, why some languages have left SVO towards VSO/VOS (Gell-Mann & Ruhlen, 2011)?

The first puzzle has been solved with the help of the concept of word order permutation ring, a circular graph where two word orders are connected if one yields the other by a swap of adjacent constituents (Ferrer-i-Cancho, 2008, 2015). That ring allows one to define a distance between word orders and to hypothesize that the likelihood of a transition is a decreasing function of the distance between the orders, which is justified as a prediction of a generalized principle of Euclidean distance minimization (Ferrer-i-Cancho, 2015). According to this hypothesis, SOV \rightarrow SVO is more likely than SOV \rightarrow OVS (Ferrer-i-Cancho, 2015). See the supplementary materials for a further justification and some clarifications on the ring hypothesis.

The second puzzle can be solved combining two facts. The first fact is that word order is a multiconstraint satisfaction problem and VSO/VOS is not optimal from the point of view of dependency length minimization but optimal from the perspective of maximizing the predictability of the contents of the nominal phrases (S and O) (Ferrer-i-Cancho, 2014). The second fact is a predisposition of SVO order to VSO/VOS. From the perspective of dependency length minimization, SVO languages could put adjectives and other dependents of the nominal heads of S and O either before and after those heads. However, postnominal placements prevent regression to SOV because SOV prefers prenominal placements. To escape from SOV, it is convenient for a SVO language to place dependents of nominal head after the verb which, interestingly, is the preferred placement for VSO/VOS (Ferrer-i-Cancho, 2015). Thus, escaping from SOV towards SVO and stabilizing the change can preadapt a SVO order to become VSO/VOS. Again, the principle of dependency length minimization is a key factor.

The subtle predictions of dependency length minimization (and other word order principles) reviewed above suggest that the conclusion that cultural evolution is more important than cognitive factors comes from a hidden assumption: that cognitive factors apply absolutely, regardless of the current state of a language (e.g., its current dominant word order) and its nearest states (e.g., the nearest dominant word orders). In absolute terms (not conditioning on the current state), dependency length minimization predicts SVO/OVS. In local or relative terms (conditioning on the current state), dependency length minimization predicts that SVO is more likely than OVS when the current state is SOV, in full agreement with *“the current state of a linguistic system shaping and constraining future states”* (Dunn et al., 2011). These converging relativistic views can

be seen as examples of Kauffman's adjacent possible: the current configuration of a system can determine the next states, overriding prior unconditional biases (Kauffman, 2000).

In this article we consider the case of languages that do not exhibit a single dominant ordering of the triple but rather two primary alternating word orders (Table 1), e.g., SOV/SVO in case of German (Dryer, 2013b). Those languages are a subset of the languages lacking a dominant word order: there are 189 languages lacking a dominant word order but only 67 exhibit such a couple of primary alternating orders (Dryer, 2013b). Here we investigate why an order associates with another to form a couple of primary alternating word orders as a mirror of constraints on word order variation and evolution. In particular, we would like to know if those associations are constrained by the distance between the partners in the word order permutation ring, being closer associations more likely, in agreement with Kauffman's adjacent possible. If that was the case, that would mean that the restrictions imposed by the ring are pervasive and then fundamental for the development of a parsimonious theory of language and word order in particular. The relationship between the problem of primary alternating word orders and the evolution of language becomes evident under the assumption that word order change exploits constraints on word order variation. If word order evolution is actually constrained by the word order permutation ring diachronically as we have explained above, we should also observe the manifestation of this ring in word order variation synchronically.

We consider various hypotheses for the choice of an order y as a partner given that the other partner is order x . We begin with a general and intuitive presentation (full mathematical details are given in Section 2).

- Model 0: the null hypothesis of obtaining y by rolling a die (one side for every ordering of the triple) till $y \neq x$.
- Model 1: a model where the probability of choosing y depends on its distance to x in the permutation ring mentioned above.
- Models 2-4: a random choice proportional to *a priori* probabilities of each of the six orders.

Model 1 and all the arguments on word order transitions reviewed above are examples of a six-way approach (the six possible orderings of S , V and O) that is in stark contrast to the dual two-way view of linguistic typology, where it is believed that linguistic properties are a direct consequence of preferences over pairs of basic orders (Cysouw, 2008). For instance, according to the latter, the frequency of each of the 6 possible orderings of S , V and O is argued to originate from (Cysouw, 2008)

- A preference for SV over VS.

- A preference for SO over OS.
- No preference for OV over VO.

In Model 2, the *a priori* probabilities reflect the dual two-way view. The parameters of the model define the probability for every basic word order. Model 3 is a six-way model where the *a priori* probabilities of each triple are derived from the probability of association of the triples in the real dataset. In contrast, Model 4 is another six-way model where *a priori* probabilities are defined by the relative frequency of the six orders as dominant orders, a factor considered to be relevant when analyzing primary alternating words orders (Dryer, 2013b).

It is important to notice that all the models except Model 1 make absolute predictions (in the sense that the knowledge about x only imposes that $y \neq x$ since the primary alternating orders must be different by definition). Model 1 makes a truly relative prediction because it takes into account the distance between x and y in the word order permutation ring. Furthermore, Model 1 is the only model with a clear cost-cutting hypothesis behind. Our analysis based on information theoretic model selection and further concerns for parsimony will show that Model 1 is the best, calling for a revision of the view that cognitive factors are secondary in word order change (Dryer, 2011; Dunn et al., 2011; Gell-Mann & Ruhlen, 2011) and providing new support for the six-way approach in the debate between six-way and two-way typology (Newmeyer, 2005; Dryer, 2013a).

Table 1. Primary alternating orders of S , V and O according to the World Atlas of Language Structures (Dryer, 2013b). $m(x, y)$ is the number of languages where orders x and y are primary alternating orders. $d(x, y)$ is the distance between x and y in the permutation ring (Ferrer-i-Cancho, 2015).

x	y	$m(x, y)$	$d(x, y)$
SOV	SVO	29	1
VSO	VOS	14	1
SVO	VSO	13	1
SVO	VOS	8	2
SOV	OVS	3	2

2. The models

\mathcal{P}_1 is defined as the set of all the possible orderings of S , V and O , i.e.

$$\mathcal{P}_1 = \{SOV, SVO, VSO, VOS, OVS, OSV\}. \quad (1)$$

\mathcal{P}_2 is defined as the set of all the different unordered pairs of elements of \mathcal{P}_1 . The cardinality of \mathcal{P}_1 , $|\mathcal{P}_1|$, is 6, and $|\mathcal{P}_2| = \binom{6}{2} = 15$.

Suppose that $p(y|x)$ is the probability that y is chosen as a partner knowing that x is already one of the partners in the couple of primary alternating orders of a language. $p(y|x)$ is defined over \mathcal{P}_2 ($p(y|x) = 0$ if $(x, y) \notin \mathcal{P}_2$). We use π to refer to the parameters of a model (if any). We consider the following models for $p(y|x)$:

- *Model 0.*

A null model where $p(y|x) = 1/5$. This model has 0 parameters.

- *Model 1.*

Suppose that $k(d)$ indicates the number of neighbors of an order at distance d in the permutation ring. It is easy to see that $k(d) = 2$ if $d \in \{1, 2\}$ and $k(d) = 1$ if $d = 3$. Suppose that $d(x, y)$ is the distance between orders x and y in the permutation ring (Table 1). This model defines $p(y|x)$ as

$$p(y|x) = \frac{p(d(x, y))}{k(d(x, y))}, \quad (2)$$

where $p(d)$ is the probability that an association involves orders at distance d in the permutation ring. The $1/k(d(x, y))$ factor in Eq. 2 indicates that Model 1 chooses the partner at distance $d(x, y)$ uniformly at random. Model 1 defines $p(d)$ with two parameters $\pi(1)$ and $\pi(2)$ such that $p(1) = \pi(1)$, $p(2) = \pi(2)$ and $p(3) = 1 - \pi(1) - \pi(2)$.

- *Models 2-4.* These three models assume

$$p(y|x) \propto q(y), \quad (3)$$

where $q(y)$ is some *a priori* probability of y . Each of the three models defines $q(y)$ in a different way (details of each below).

In standard typology, the choice of a possible ordering of S , V and O is believed to originate from local word order preferences that can be defined by $p(\alpha\beta)$, the probability of placing α before β (not necessarily consecutively), with $p(\beta\alpha) = 1 - p(\alpha\beta)$. Model 2 has three parameters, i.e. $\pi(SV)$, $\pi(SO)$ and $\pi(OV)$, which define $p(SV)$, $p(SO)$ and $p(OV)$, respectively. In Model 2, $q(y) = cr(y)$, where c is a normalization constant and

$$\begin{aligned} r(SOV) &= c \cdot p(SV)p(SO)p(OV) \\ r(SVO) &= c \cdot p(SV)p(SO)(1 - p(OV)) \\ r(VSO) &= c \cdot (1 - p(SV))p(SO)(1 - p(OV)) \\ r(VOS) &= c \cdot (1 - p(SV))(1 - p(SO))(1 - p(OV)) \\ r(OVS) &= c \cdot (1 - p(SV))(1 - p(SO))p(OV) \\ r(OSV) &= c \cdot p(SV)(1 - p(SO))p(OV). \end{aligned} \quad (4)$$

Model 3 defines $q(y)$ with five parameters, e.g., $\pi(y)$ for $y \in \mathcal{P}_1 \setminus \{OSV\}$ since

$$q(SOV) = 1 - \sum_{y \in \mathcal{P}_1 \setminus \{OSV\}} \pi(y). \quad (5)$$

Model 4 is a particular case of Model 3 where $q(y)$ is the relative frequency of languages where y is dominant (Dryer, 2013b). Model 4 has no parameters. Model 0 is provided as a control for Models 1-4. Models 1-4 should be better than Model 0, the null hypothesis. See the supplementary materials for further mathematical details about the models.

3. Evaluation

$m(x, y)$ is the number of languages where x and y are primary alternating orders, with $m(x, y) = m(y, x)$ (Table 1). The total number of languages is defined as

$$m = \sum_{(x,y) \in \mathcal{P}_2} m(x, y) = \frac{1}{2} \sum_{(x,y) \in \mathcal{P}_1 \times \mathcal{P}_1, x \neq y} m(x, y), \quad (6)$$

where $\mathcal{P}_1 \times \mathcal{P}_1$ is the Cartesian product of \mathcal{P}_1 with itself. The actual sample size is $n = 2m$ since every language in the sample has two orders in partnership and our models predict each.

The log-likelihood of every partner in the primary alternating orders of a set of languages is defined as

$$\mathcal{L} = \sum_{(x,y) \in \mathcal{P}_1 \times \mathcal{P}_1, x \neq y} m(x, y) \log p(y|x). \quad (7)$$

We evaluate the quality of the fit of a model with k parameters over a sample of size n with two scores. One is the corrected Akaike information criterion (Burnham & Anderson, 2002), defined as

$$AIC_c = -2\mathcal{L} + 2 \frac{kn}{n - k - 1}. \quad (8)$$

The other is the Bayesian information criterion (Wagenmakers & Farrell, 2004), defined as

$$BIC = -2\mathcal{L} + k \log n. \quad (9)$$

The smaller the score, the better the model (Burnham & Anderson, 2002). We use two criteria and not simply one to show overtly the assumptions of our arguments and reflect about them. AIC_c and BIC offer complementary perspectives (Wagenmakers & Farrell, 2004):

- While BIC assumes that the true generation model is in the set of candidate models, AIC_c does not assume that any of the candidate models is necessarily true.

- When $n \geq \lceil e^2 \rceil = 8$, BIC introduces a stronger penalization for the number of parameters than AIC_c (see supplementary materials for a detailed argument). Here $n = 2m = 134$ since $m = 67$ according to Table 1 and thus BIC is going to penalize more than AIC_c . This is of special relevance here because in case that BIC and AIC_c do not agree on the best model, we will assume that reality is relatively low dimensional or that the best model has to be inserted in a compact general theory of language. As a result of these assumptions, we will take the best model according to BIC as the final best, as this criterion favors the model with the smallest number of parameters in our dataset.

The best parameters of every model are estimated replacing every probability parameter by the corresponding proportion in the dataset in Table 1 (see supplementary materials for further details about parameter estimation).

4. Results

Table 2 summarizes the results of the evaluation of the different models. Model 3 is the best model according to AIC_c whereas Model 1 is the best model according to BIC. These two models are in a kind of tie because the difference of AIC_c between Model 1 and 3, as well as that of BIC, are small. However, we favor the winner according to BIC for two reasons. On a local basis, Model 1 has less parameters than Model 3. On a general basis, Model 1 leads to a lighter general theory of language: with Model 1, we can explain both the diachronic patterns reviewed in Section 1 but also synchronic patterns such as the primary alternating orders. If we chose Model 3, the general theory becomes fatter because Model 3 has more parameters than model 1 and also because additional arguments to explain the origins of these diachronic patterns are required. Interestingly, Model 2, the model inspired by standard typology, is worse than Model 1 and 3 according to both AIC_c and BIC. Model 4 is even worse than Model 0, the null hypothesis.

The fact that Model 1 is the final best model does not imply that the word order permutation hypothesis is valid because this hypothesis means $p(1) > p(2) > p(3)$. Actual support for the word order permutation is provided by the maximum likelihood estimates, which are $\pi(1) = 0.84$, $\pi(2) = 0.16$. The latter finding suggests that word order associations for dominant word orders are constrained in the same way as word order transitions during evolution. As for Model 2, the maximum likelihood parameters are $\pi(SV) = 0.61$, $\pi(SO) = 0.81$ and $\pi(OV) = 0.26$.

5. Discussion

We have shown that Model 1 is the best model when favoring parsimony strongly on a local basis (using BIC) and also on a general basis (favoring a light general theory of language). Model 1 has beaten all the models that assume that word

Table 2. Summary of the evaluation of the models. \mathcal{L} is the log-likelihood, k is the number of parameters of the model, AIC_c is the corrected AIC and BIC is the Bayesian information criterion.

Model	\mathcal{L}	k	AIC_c	BIC
0	-215.7	0	431.3	431.3
1	-152.7	2	309.5	315.2
2	-161.5	3	329.3	337.8
3	-147.3	5	305.2	319.2
4	-285.3	0	570.6	570.6

order biases apply independently from the current state of the system. This is not very surprising since those models suffer from a strong bias for SVO or SOV that has difficulties to deal with the 14 languages where VSO and VOS are the primary dominant orders (Table 2). Some parameter reductions of Models 1 and 2 are discussed in the supplementary materials. That of model 1 is specially relevant because it could increase the difference in BIC between Model 1 and 3, which is currently rather small (Table 2).

We have presented an analysis that shows the potential of information theoretic model selection (Burnham & Anderson, 2002) to evaluate competing hypotheses about word order variation. The failure of Model 2 challenges the strong belief in the predictive power of a dual two-way approach of standard typology (Cysouw, 2008; Dryer, 2013a). Six-way approaches (Newmeyer, 2005) should be reconsidered. The success of Model 1 when favoring parsimony strongly suggests that the view that cognitive cost minimization is secondary to word order change (Dryer, 2011; Dunn et al., 2011; Gell-Mann & Ruhlen, 2011) should be revised.

All these considerations suggest that Kauffman’s notion of adjacent possible (Kauffman, 2000) is crucial for progress in typology and word order evolution research. Although our work has an important limitation, i.e. it assumes that languages are independent (Cysouw, 2008), we hope to have built “adjacent possibles” to apply, hopefully in a near future, the powerful phylogenetic methods that have already been used to investigate word order evolution with the traditional dual two-way approach (Dunn et al., 2011).

Acknowledgments

The feedback of three anonymous reviewers has been crucial. We are also grateful to B. Bickel, W. Croft, E. Santacreu-Vasut and S. Strogatz for helpful discussions. This research has been funded by the grants 2014SGR 890 (MACDA) from AGAUR (Generalitat de Catalunya) and the grant TIN2014-57226-P from MINECO (Ministerio de Economía y Competitividad).

References

- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference. A practical information-theoretic approach* (2nd ed.). New York: Springer.
- Cysouw, M. (2008). Linear order as a predictor of word order regularities. *Advances in Complex Systems*, 11(3), 415-420.
- Dryer, M. S. (2011). The evidence for word order correlations. *Linguistic Typology*, 15, 335 - 380.
- Dryer, M. S. (2013a). Against the six-way order typology, again. *Studies in Language*, 37, 267 - 301.
- Dryer, M. S. (2013b). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473, 79-82.
- Ferrer-i-Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E*, 70, 056135.
- Ferrer-i-Cancho, R. (2008). Some word order biases from limited brain resources. A mathematical approach. *Advances in Complex Systems*, 11(3), 393-414.
- Ferrer-i-Cancho, R. (2014). Why might SOV be initially preferred and then lost or recovered? A theoretical framework. In E. A. Cartmill, S. Roberts, H. Lyn, & H. Cornish (Eds.), *THE EVOLUTION OF LANGUAGE - Proceedings of the 10th International Conference (EVOLANG10)* (p. 66-73). Vienna, Austria: Wiley. (Evolution of Language Conference (Evolang 2014), April 14-17)
- Ferrer-i-Cancho, R. (2015). The placement of the head that minimizes online memory. A complex systems approach. *Language Dynamics and Change*, 5, 114-137.
- Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences USA*, 108(42), 17290-17295.
- Kauffman, S. A. (2000). *Investigations*. Oxford: Oxford University Press.
- Newmeyer, F. J. (2005). *Possible and probable languages. A generative perspective on language typology*. Oxford: Oxford University Press.
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11, 192-196.
- Zipf, G. K. (1949). *Human behaviour and the principle of least effort*. Cambridge (MA), USA: Addison-Wesley.

PHONEME INVENTORY SIZE DISTRIBUTIONS AND THE ORIGINS OF THE DUALITY OF PATTERNING

LUKE FLEMING

*Département d'anthropologie, l'Université de Montréal,
Montréal, Canada
luke.fleming@umontreal.ca*

Atkinson 2011 claims that phoneme inventories are largest in Africa and smaller elsewhere, and that this clinal distribution reflects a 'founder-effect' of human migrations 'out-of-Africa'. Because of the way in which velaric ingressive and pulmonic egressive airstream mechanisms combine to create extra-large consonant inventories, click languages have the largest phoneme inventories of all. Critics question why phoneme inventory size, but not other properties of language, should leave a trace of the origin and dispersal of natural language. This paper argues that large phoneme inventories would likely have been characteristic of the first fully modern languages if we assume, following Hockett 1960, that duality of patterning was the last 'design feature' of language to emerge. The diachronic trajectories of sign languages and writing systems illustrate that dually patterned phonologies—where minimal units of linguistic form (or phonemes) capable of distinguishing semantic units (or morphemes) are not meaningful in themselves—are often preceded by a stage in which minimal units of form map directly onto semantic functions. Click articulations would have been essential in elaborating large inventories, and thus large vocabularies, in spoken languages lacking duality of patterning. The contemporary distribution of phonemic clicks offers support for the hypothesis, as genetic studies increasingly point to an eastern or southern African origin for modern humans.

1. Introduction

Atkinson (2011) surveyed phonological data from over 500 languages to argue that “phonemic diversity” is greatest in Africa, and “declines with distance from Africa”, as with genetic and phenotypic diversity, and thus “supports an African origin of modern human languages” (Atkinson, 2011:346). Atkinson’s theory, drawing on Hay and Bauer (2007), presupposes a constant of phoneme inventory growth and reduction as a function of population size. However, the positive correlation between phoneme inventory size and population size, if valid (cf. Donohue & Nichols, 2011), is a proxy for complex social and linguistic processes (Trudgill, 2011). The population-size-to-phoneme-

inventory-size correlation seems to hold best for complex, agricultural, state-based societies of the kind found only since the Neolithic. Population size effects on phoneme inventories would likely not have been marked among the Paleolithic hunter-gatherer societies—of small populations, in absolute terms—responsible for human range expansion (Bowerman, 2011). Furthermore, evidence for a natural rate of increase of phoneme inventories (as assumed in Perreault and Mathew [2012]) is lacking (Ringe, 2011). In sum, the set of explanatory principles posited in Atkinson (2011) seem ill-suited to explain why the highest levels of genetic diversity and the largest phoneme inventories are both found in Africa.

An even larger global survey of more than 2,000 languages in Creanza et al. (2015) questions whether a “signal” of African origins can even be found in phoneme inventory distributions. Nevertheless, the authors do find a rough correlation between genetically polymorphic populations (the top fifth of which are all found in Africa) and languages with large phoneme inventories (the top 28 languages with the most phonemes in the 2,000+ sample are all from Africa) (Creanza et al., 2015:1267, Table S4). In this paper I suggest an alternative way to look at the intriguing idea that the distribution of phoneme inventories by size reflects human language origins and dispersal.

One way in which to get around the multiple methodological and empirical problems facing the founder effect model (albeit keeping Atkinson’s assumption of linguistic monogenesis) is to stipulate that the earliest natural languages had much larger phoneme inventories than contemporary languages. This possibility has likely not been pursued because of its seeming violation of the uniformitarian hypothesis. However, the uniformitarian hypothesis in historical linguistics should imply only that the processes of change affecting prehistoric languages were the same as those affecting contemporary languages, not that the kinds of languages upon which they operated had the same typological profiles as contemporary ones (Comrie, 2003; Newmeyer, 2002). The first modern human languages likely had features derived from the protolanguages out of which they emerged (Nichols, 2011). Since protolanguages were organized by different functional principles than modern languages we might expect the first human languages to be typologically anomalous with respect to contemporary languages, although we cannot know a priori along which axes they differed.

If the first natural languages had extra-large phoneme inventories a much modified version of Atkinson’s linguistic-founder effect could still apply, even without the Hay and Bauer population-size-to-phoneme-inventory-size correlations. There is a strong correlation between consonant inventory size and articulatory complexity; the larger a phonemic inventory the more articulatorily complex segments it possesses (Lindblom & Maddieson, 1988). If we assume that the first languages had large inventories we can also assume that they had complex articulations. In this scenario, as marked phonological features (e.g. ejectives, implosives, clicks, etc.) were lost in daughter languages there would have been low probability of the renewal of those features outside of contact

areas where they were present and could be renovated via horizontal cultural transmission.

This hypothesis does a decent job of accounting for the observed fit between largest phoneme inventories and most genetically diverse populations. The rule of thumb in historical linguistics is that the homeland of a language family is most likely “the area represented by the greatest diversity (largest number of subgroups) for which the minimum number of moves (migrations) would be required to bring the speakers of the diverse languages back to one place” (Campbell, 2004:430). On the assumption of extra-large phoneme inventories as the initial condition of spoken language, a good guess for the homeland of language would be a site where languages affiliated with distinct language families, all possessing extra-large phoneme inventories of high articulatory complexity, are found clustered together. Importantly, the largest phonemic inventories in the world are found in the areally linked, but genealogically separate, Kx’a, Tuu, and Khoe language families (cf. Greenberg’s “Khoisan”) (see section 4). And these “core click languages” (Güldemann & Stoneking, 2008) are spoken in southern Africa, in the same geographical area where some of the most genetically polymorphous populations are found (Knight, 2003; Tishkoff et al., 2007).

If this hypothesis offers a plausible cause for the correlation between largest phoneme inventories and most genetically diverse populations, it is still unclear what reason could motivate us to accept the counterintuitive conclusion that the earliest human languages had markedly larger and more complex phonological inventories than contemporary ones. In the following sections I argue that the evolution of the phonology-semantics interface may offer such a rationale.

2. Duality of Patterning and the Lexicon Design Problem

As Hockett ([1960] and Martinet [1960]) observed, languages are patterned at two distinct levels or planes. At one level of patterning meaningful units of language, words and other morphemes, combine with one another to form phrases and sentences. But those morphemes are themselves made up of another order of units, phonemes. The phonemes of spoken languages are sound segments capable of distinguishing meaningful units from one another, but not inherently meaningful in and of themselves. The organization of human language thus relies upon the combination of meaningless phonemic units to yield meaningful morphemic units, which themselves combine with one another to form sentences—two levels of patterning.

Duality of patterning is often seen as essential for the development of a large lexicon, and then most pressingly in the spoken modality of language (de Boer, Sandler, & Kirby, 2012). Since the need for duality of patterning only arises under conditions where a large number of meaningful speech signals are already employed, Hockett (1960:96) reasoned that it would be one of the last developments in the natural history of language. In a spoken language lacking

duality of patterning, minimal units of segmentable form would map directly onto semantic functions. Following Hjelmslev (1961:113), I call such linguistic systems MONOPLANAR. Since lexicons of languages of this type are limited by the inventory of representational forms which they employ, they tend towards the production a large number of perceptibly different form types (see section 3). However, to represent each morpheme in the lexicon, especially in the oral-aural modality, by means of a unique sign-form would be extremely burdensome. It may be true that visually based codes, like sign and writing, allow for indefinitely large monoplanar lexicons. For the oral-aural modality, however, there are rather stringent limits on the number of phonetic distinctions that humans can perceptually make. Duality of patterning solves this problem by adding a level of complexity in the form of a mediating level between representational-form and semantic-meaning. The emergence of duality of patterning in sign languages and writing systems offers good evidence for these claims.

3. Diachronies of Duality of Patterning Beyond the Spoken Modality

There are concrete obstacles to reconstructing the emergence of the duality of patterning in speech. The dominance of speech, among linguistic modalities, ironically means that we have perhaps the most limited knowledge about the emergence, *ex nihilo*, of spoken languages. Because of their polygenetic development, visually based codes offer the best empirical data for studying the emergence of the duality of patterning.

Sign languages, like spoken languages, typically exhibit duality of patterning in their phonological structures (Stokoe, 1960). Phonological categories typically organized into distinctive features in sign languages are Hand Configuration, Location, and Movement (Sandler et al., 2011). In their study of an emerging village sign language, Al-Sayyid Bedouin Sign Language (ABSL), Sandler et al. (2011) found that ABSL did not exhibit duality of patterning. When compared with established sign languages ABSL exhibits many more gesture types. For instance, ABSL has a much larger inventory of handshapes than occurs in established sign languages (see Sandler et al., 2011, Table 3). Many of the handshapes are, in typological phonological terms, highly marked, and would not be expected to occur in doubly articulated sign language phonologies. Sandler et al. (2011) suggests that the phonology-semantics interface is an emergent property of a growing sociocultural tradition of signing rather than the unmediated reflection of an innate cognitive-linguistic architecture. The takeaway from the ABSL data is straightforward: There appears to be a general tendency in the historical development of sign language phonology towards a progressive reduction of formal resources employed in signing concurrent with the development of duality of patterning (cf. Frishberg, 1975).

Writing systems also offer a fertile evidentiary source for conceptualizing what precedes the duality of patterning in the development of linguistic systems. In all three of the cases where writing is thought to have been independently invented—Mesoamerica, China, and Mesopotamia—writing systems began with primarily logographic orthographies (Boltz, 1994; Schmandt-Besserat, 1996). Syllabaries and alphabets, the writing systems which most closely parallel duality of patterning in speech, are always derived forms. Here again, change from a monoplane system to a dually patterned one involves a massive reduction in the number of minimal segments of form employed.

In both emergent sign languages and emergent writing traditions, the formal organization of the initial state of linguistic varieties does not exhibit duality of patterning. Rather, distinct morphemes are represented by means of a wide variety of unique and perceptually distinct forms, whether manual or graphic. Analogs from signed and written linguistic varieties suggest that monoplane speech was a historical precursor to duality of patterning in spoken natural language. Furthermore, in both the manual and graphic systems lacking duality of patterning, a much wider range of minimal units of representational form, whether handshapes or glyphs, are employed than in corresponding dually patterned systems. In terms of the evolution of speech, this would imply strong selection for complex articulations as components of very large inventories of monoplane spoken languages. In particular, as we will now see, given the way in which velaric ingressive and pulmonic egressive airstream mechanisms combine to generate large numbers of complex segments, the addition of click consonants would offer the possibility of greatly expanding the number of semantic sense distinctions that could be made in monoplane spoken language.

4. Multiplier Effect of Velaric Ingressives for Phonemic Inventory Size

The phoneme inventories of Khoesan languages are exceptionally large, hypertrophied in particular by their consonant inventories. Brugman (2009:28-29) cites the following numbers for phoneme inventories of select Khoesan languages: N!uu (Tuu) has 86 segments (73 of which are consonants); !Gui (Khoe) has 99 segments (89 of which are consonants); Jul'hoansi (Ju) has 123 (89 of which are consonants). Maddieson (2005:10) finds a mean consonant inventory of 22.7 for a sample of 562 languages, with 6 for Rotokas (Papuan) and 122 for !Xóǀ (Tuu) as the smallest and largest, respectively. Khoesan phonemic inventories are far and away the largest in the world, and this is due to their possessing click consonants.

Clicks are produced by means of a distinct airstream mechanisms than the one underlying the production of the vast majority of the most widespread phonetic articulations used in the world's languages. Phonation is typically achieved by modifying the outward flow of air. Phonemes produced in this manner are pulmonic egressives. Click consonants, however, are produced by lingual (or labial) suction. By making a closure with either the front of the

tongue or the lips, on the one hand, and the dorsum of the tongue, on the other, a low-pressure pocket of air can be created. When the front closure is released, high-pressure air from outside the mouth rushes in, creating the click sound. Phonemes produced in this manner are velaric (or lingual) ingressive.

Importantly, the addition of this other type of phonation does not just have an additive effect on phonemic inventory size—it has a multiplier effect. Because they rely on different airstream mechanisms, velaric ingressive consonants can co-occur synchronously with a range of pulmonic egressive phonations including voicing, nasalization, aspiration, glottalization, uvularization, and affrication (Clements, 2000:151). Take, for instance, the case of Ju|'hoansi. There are four velaric ingressive consonant types in Ju|'hoansi: dental, lateral, alveolar and palatal (l, ll, !, and †, respectively). Each of these types, however, can combine with 11 types of egressive phonation or 'efflux'. In !Xóõ, 5 click types (labial, dental, alveolar, lateral, and palatal) combine with 16 effluxes to make more than 80 different click types (Clements, 2000:151). It is the presence of clicks which makes these phoneme inventories so large. "Cross-linguistically no other sound class is subject to such an extensive series formation—inter alia by combining with a second, i.e., the pulmonic, air-stream mechanism" (Güldemann & Stoneking, 2008:105). Because of their multiplier effect on phoneme inventory size, in monoplanar spoken languages there would have been strong motivation for the elaboration of velaric ingressesives.

5. Conclusion

Hockett, Martinet (1960:15), and Hjelmslev (1961:46) all argued that duality of patterning is an adjustment not strictly necessary for language, as such, but for the creation of an indefinitely large and open-ended lexicon. In his foundational piece on the evolution of speech, Charles Hockett (1960:96) wrote that "[t]here is excellent reason to believe that duality of patterning was the last [design feature of speech] to be developed, because one can find little if any reason why a communicative system should have this property unless it is highly complicated." Hockett's historical sequencing accords with evidence that adaptations necessary to produce complex speech patterns came relatively late in human evolution (MacLarnon & Hewitt, 1999; Lieberman, 2007). If duality of patterning relies upon a specifically evolved genetic endowment, we would expect selection pressure to be the greatest in the transition to speech; a much larger variety of perceptibly distinct segments can be produced in the manual-visual modality than in the oral-aural one; manual-visual codes lacking duality of patterning, like the Al-Sayyid Bedouin Sign Language studied by Sandler et al. (2011), can have large lexicons notwithstanding the limitations of their phonological organization.

A linguistic stage lacking duality of patterning would have represented a significant evolutionary bottleneck in the development of speech. Comparative evidence from the diachronic development of written and signed linguistic

varieties suggests that dually patterned phonological systems would have been preceded by a monoplanar stage in which phonological segments were mapped directly onto semantic functions. During a phase of monoplanar spoken language, any expansion in the number of perceptibly distinct sound segments produced would enable a comparable expansion in the lexical or morphemic inventory of the language. The use of a different airstream mechanism would have been invaluable here; velaric ingressives form highly productive series in conjunction with pulmonic egressives, much more productive than either implosives or ejectives. Indeed, in the core click languages, languages with the largest phoneme inventories in the world, click consonants routinely make up more than half of the inventory segments. Velaric ingressive, thus, would have been invaluable in monoplanar spoken languages.

The evolutionary development of duality of patterning would have enabled a newfound flexibility for phonological inventories hitherto operating at capacity. Nevertheless, the sound systems of those first dually patterned spoken languages would have converted segments sourced from extant monoplanar languages in fashioning their phoneme inventories. Given the importance of velaric ingressives for generating large inventories, clicks would likely have been derived segments in the first languages possessing duality of patterning. Velaric ingressives would thus have been an areal linguistic feature where dually patterned languages first emerged. The contemporary distribution of phonemic clicks offers supporting evidence for the deduction; velaric ingressives have a unique typological profile, being found in a linguistic area—southern and eastern Africa—where modern humans are thought to have originated (Henn et al., 2011).

This hypothesis, if true, would explain why phoneme inventories, but not other features of language, evince an (albeit noisy) “out-of-Africa” signal. Large phoneme inventories reflect, on this account, the influence of a historically prior, monoplanar phonological system. Large phonemic inventories are prevalent in southern African languages not because these have a more direct genealogical relationship with such monoplanar protolanguages, but because of sustained linguistic contact between language groups employing large and articulatorily complex phonological inventories in this area. The highly marked velaric ingressive airstream mechanism is best conceptualized as an areal feature kept alive as much through horizontal, as by vertical, transmission, and too marked, when compared to the unmarked pulmonic egressive airstream mechanism, to be independently developed outside of a sub-Saharan linguistic contact zone.

References

- Atkinson, Q. (2011). Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science*, 332, 346-349.
- de Boer, B., Sandler, W. & Kirby, S. (2012). New perspectives on duality of patterning: Introduction to the special issue. *Language & Cognition*, 4(4), 251-259.
- Boltz, W. (1994). *The Origin and Early Development of the Chinese Writing System*. New Haven: American Oriental Society.
- Bowern, C. (2011). Out of Africa? The logic of phoneme inventories and founder effects. *Linguistic Typology*, 15(2), 207-216.
- Brugman, J. (2009). *Segments, Tones and Distribution in Khoekhoe Prosody*. Ph.D. Thesis. Cornell University.
- Campbell, L. (2004). *Historical Linguistics*. Cambridge, MA: MIT Press.
- Clements, G.N. (2000). Phonology. In B. Heine and D. Nurse (Eds.), *African Languages: An Introduction* (pp. 123-160). Camb.: Cambridge Univ. Press.
- Creanza, N. et al. (2015). A comparison of worldwide phonemic and genetic variation in human populations. *PNAS*, 112(5), 1265-1272.
- Donohue, M. & Nichols, J. (2011). Does phoneme inventory size correlate with population size? *Linguistic Typology*, 15, 161-170.
- Frishberg, N. (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language*, 51(3), 696-719.
- Güldemann, T. & Stoneking M. (2008). A historical appraisal of clicks: A linguistic and genetic population perspective. *Annual Review of Anthropology*, 37, 93-109.
- Hay, J. & Bauer L. (2007). Short report: Phoneme inventory size and population size. *Language*, 83, 388-400.
- Henn, B., et al. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *PNAS*, 108, 5154-5162.
- Hjelmslev, L. (1961). *Prolegomena to a Theory of Language*. F. Whitfield (Trans.). Madison: University of Wisconsin Press.
- Hockett, C. (1960). The origin of speech. *Scientific American*, 203, 88-111.
- Knight, A., et al. (2003). African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Current Biology*, 13, 464-473.
- Lieberman, P. (2007). The evolution of human speech: Its anatomical and neural bases. *Current Anthropology*, 48(1), 39-66.
- Lindblom, B. & Maddieson, I. (1988). Phonetic universals in consonant systems. In C. Li & L. Hyman (Eds.), *Language, Speech and Mind* (pp. 62-78). London: Routledge.
- MacLarnon, A. & Hewitt, G. (1999). The evolution of human speech: The role of enhanced breathing control. *American Journal of Physical Anthropology*, 109(2), 341-363.

- Maddieson, I. (2005). Consonant inventories. In M. Haspelmath et al. (Eds.), *The World Atlas of Language Structures* (pp. 10-13). Oxford: Oxford University Press.
- Martinet, A. (1960). *Éléments de linguistique générale*. Paris: Armand Colin.
- Newmeyer, F. (2002). Uniformitarian assumptions and language evolution research. In A. Wray (Ed.), *The Transition to Language* (pp. 359-375). Oxford: Oxford University Press.
- Nichols, J. (2012). Monogenesis or polygenesis: A single ancestral language for all humanity? In M. Tallerman and K. Gibson (Eds.), *The Oxford Handbook of Language Evolution* (pp. 558-572). Oxford: Oxford University Press.
- Perreault, C. & Mathew S. (2012). Dating the origin of language using phonemic diversity. *PLoS ONE*, 7(4), e35289.
- Ringe, D. (2011). A pilot study for an investigation into Atkinson's hypothesis. *Linguistic Typology*, 15(2), 223-231.
- Sandler, W. et al. (2011). The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory*, 29, 502-543.
- Schmandt-Besserat, D. (1996). *How Writing Came About*. Austin: University of Texas Press.
- Tishkoff, S. et al. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Molecular Biology and Evolution*, 24(10), 2180-2195.
- Trudgill, P. (2011). Social structure and phoneme inventories. *Linguistic Typology*, 15, 155-160.

Migration as a Window into the Coevolution between Language and Behavior

VICTOR GAY

*Department of Economics, University of Chicago,
Chicago, USA
victorgay@uchicago.edu*

DANIEL L. HICKS

*Department of Economics, University of Oklahoma
Norman, USA
hicksd@ou.edu*

ESTEFANIA SANTACREU-VASUT

*Department of Economics, ESSEC Business School
Cergy-Pontoise, France
santacreuvasut@essec.edu*

Understanding the causes and consequences of language evolution in relation to social factors is challenging as we generally lack a clear picture of how languages coevolve with historical social processes. Research analyzing the relation between language and socio-economic factors relies on contemporaneous data. Because of this, such analysis may be plagued by spurious correlation concerns coming from the historical co-evolution and dependency of the relationship between language and behavior to the institutional environment. To solve this problem, we propose migrations to the same country as a microevolutionary step that may uncover constraints on behavior. We detail strategies available to other researchers by applying the epidemiological approach to study the correlation between sex-based gender distinctions and female labor force participation. Our main finding is that language must have evolved partly as a result of cultural change, but also that it may have directly constrained the evolution of norms. We conclude by discussing implications for the coevolution of language and behavior, and by comparing different methodological approaches.

1. Introduction

1.1. *The Methodological Challenge*

Disentangling whether language influences the evolution of society, whether social factors impact language evolution, or whether they are independent of each other, is a daunting challenge. Indeed, it requires ruling out spurious

correlations in cross-cultural linguistic analysis and addressing a fundamental problem of identification. As Roberts and Winters (2013) highlight, it may be inappropriate to simply treat languages as independent data points because they are related by both vertical and horizontal transmission mechanisms.

For instance, sharing a common ancestor (language families) or spillovers via contact with neighboring languages in the past (linguistic areas) may generate spurious correlations between language and behavior. More concretely, it hinders our understanding of whether linguistic characteristics reflect changes in socio-economic relations and culture, whether they evolve independently, or even if they constraint and influence directly behavior. Roberts et al. (2015) demonstrate that cross-cultural correlations involving languages may be spurious once these language dependencies are accounted for and propose a series of empirical tests to help address these features of language.

In this paper, we consider a further methodological complication, which arises in when studying relationships between language structure and socioeconomic behavior: the potential for these associations to depend on the surrounding environment. That is, that language may co-evolve with institutional constraints.

An example is illustrative. Consider the correlation between future time reference (FTR) in language and the propensity to save as examined in Chen (2013) and Roberts et al (2015). Assume that a correlation between the two exists. That is, speakers of languages that exhibit a stronger FTR have a higher propensity to save. A task such as saving does not occur in vacuum. Rather, observed saving behaviors are dependent on the existence of a liquid and stable financial system regardless of individual preferences. Should such a system not exist (or should it be highly inefficient), a higher propensity to save may translate into higher investment in non-financial assets such as cattle – which may not be observable to the researcher. An empirical analysis of the relationship between languages' FTR and (financial) savings behavior could then falsely conclude that there is no relationship. Hence, it is possible for the estimated magnitude and significance of observed correlations between linguistic and socioeconomic behaviors to depend on the institutional environment within which individuals operate.

1.2. *Our Proposal*

We propose a new methodology to address this component of the identification problem: the application of the epidemiological approach. This approach takes its origin from epidemiologists who compare immigrants to natives in order to isolate the contribution of genetic factors from the influence of correlated environmental factors. The idea is to use immigrant populations to study the relationship between linguistic features and non-linguistic choices or individual outcomes that may evolve under a common institutional environment.

As an example, we study the labor market decisions of immigrants in the US. These migrants speak languages that exhibit varying levels of grammatical gender distinction. Theory suggests that we should expect women speaking languages that contain genders based on biological sex to participate less intensively in formal labor markets and instead to adopt more traditional gender roles such as work within the home (Hicks et al. 2015).

The empirical strategy we propose allows researchers to control for linguistic co-evolution, the institutional set up of the host country, and for unobservable cultural influences obtained in the origin country. This strategy draws its identification from migrants originating from the same country, but speaking languages with varying structure.

We empirically test this hypothesis on a sample of 675,000 immigrants in the U.S. from 156 countries and speaking 47 languages. We show that this approach is compatible to that of Roberts et al. (2015), which controls for language relatedness. In particular, allowing the intercept as well as the slope of the relationship to vary, as a function of language structure and behavior, is feasible. The rest of the paper is organized as follows. Section 2 presents the epidemiological approach. Section 3 presents an application. Section 4 concludes.

2. The Epidemiological Approach

Epidemiologists rely on the comparison of immigrant and native populations in order to isolate the contribution of genetic factors from the influence of environmental factors. This approach has been extensively applied within the economics research (Fernandez, 2007). Fundamentally, this approach implies studying variations across first and second-generation migrants to investigate the impact of their culture and disentangle its effect from the institutional and political environment of the host country. We propose that extending this approach to study language correlations with cultural and socio-economic outcomes is a fruitful avenue for future research.

Studying the behavior of migrants allows the researcher to compare individuals that evolve in a common institutional environment. As a result of the shared environment, incentives regarding their socioeconomic behavior are held constant across individuals. For linguistics specifically, it is possible to undertake a comparison of individuals who share the same country of origin, but speak different languages. Exploiting this source of heterogeneity allows researchers to control for a wide range of unobservable factors from both the home and the host country.

We provide an example to illustrate the set of strategies available to researchers when using this methodology. In particular, the next section presents an analysis of female labor participation among immigrants to the U.S and its correlation with sex-based grammatical distinctions in language. This

application also highlights the richness of available census data concerning linguistic diversity both across and within countries of origin.

3. Application: Gender Marking and Female Labor Participation

3.1. Data

Our sample comes from the US in the American Community Survey 2007-2011 (ACS, 5% sample) and consists of migrants who report speaking a language other than English in their own home. This provides 675,000 observations from 156 countries and speaking 47 different languages. For each migrant, we have information about their labor market status, country of origin, language spoken in the home, and various other socioeconomic indicators such as income, education, marital status, level of English proficiency, and time since migration. Our outcome variable is a dummy variable equal to 1 if the individual is in the labor force and 0 otherwise. To quantify the presence of gender distinctions in language, we assign a dummy variable equal to 1 if the language has a gender system based on biological sex, and 0 if not. We obtain this information from the World Atlas of Language Structures (Dryer & Haspelmath 2013). While most languages around the world have a sex-based gender system, migrants to the US are from sufficiently diverse countries that the sample offers a wide variation in language structure. In particular, the average value of our linguistic dummy is 0.81 with a standard deviation of 0.39.

3.2. Empirical Strategies available in the Epidemiological Approach

A further key advantage of the epidemiological approach is that it allows the researcher to employ fixed effects strategies, which we illustrate in the following example. As a benchmark, we start by assessing the simple correlation between labor participation and sex-based grammatical distinctions in language. Because we are interested in the gap in participation between women speaking languages with different grammatical structure, we include an indicator variable equal to 1 if the individual is a woman, and an interaction between that indicator variable and our language variable.

The coefficient of interest is this interaction term: it measures the additional impact on labor participation of being a female migrant speaking a language with a sex-based gender system compared to being a female migrant speaking a language without a sex-based gender system. This effect is in addition to the estimated impact of being a female compared to being a male (captured by the female coefficient alone), and in addition to the direct impact of speaking sex-based language alone (regardless of gender).

Additionally, we control throughout the analysis for the individual's income and education levels, English proficiency, marital status and state of residence, as these factors may influence economic participation rates. In this setting, the

interaction term compares women who have the same socioeconomic profile and live in the same state, but who speak a language that has a different grammatical structure. We use a simple OLS regression model. This simplifies the interpretation of the results, which are virtually the same as with a logit regression model. Column (1) in Table 1 presents these results.

A first strategy when using the epidemiological approach is to use country of origin fixed effects. This allows us to capture the role of norms of behavior related such as gender roles acquired prior to migration that are specific to an immigrant's country of origin. These fixed effects capture unobservable cultural influence on the migrants' behavior. Such a strategy allows us to effectively compare labor participation of women with similar socioeconomic background, living in the same US state, and coming from the same country, but speaking a language with a different grammatical structure. The results are presented in column (2) of Table 1.

Second, the epidemiological approach permits the use a set of fixed effects to address language relatedness. Indeed, languages may be related in two ways: a common ancestor (vertical dependence) and language contact (horizontal dependence) as discussed by Roberts et al. (2015). To account for the impact of language relatedness, we include a set of fixed effects for each language's family and linguistic area (Nichols et al. 2013). This allows the correlation between gender in language and labor market participation to have a different intercept across languages that pertain to a different language family and linguistic area. Column (3) of Table 1 includes language family and language area fixed effects.

Third, Roberts et al. (2015) argue that the strength of the correlation between a linguistic trait and a non-linguistic variable may itself be dependent on language relatedness. We can control for this dependence by including a set of interactions between each language's family and linguistic area, and the linguistic feature of interest itself. This allows the correlation to have a different slope across languages that pertain to a different language family and linguistic area. Column (4) of Table 1 presents the results.

A final strategy is to include fixed effects of the country of origin interacted by the subpopulation that the linguistic trait is supposed to affect. This approach depends on the particular nature of such a trait. In our example, the main assumption is that women speaking a language with a sex-based gender system are less likely to participate in the labor market, due to gender roles embedded in and/or caused by the language structure. If so, it should also be the case that these women behave differently than man in the country of origin. Therefore, an even more stringent strategy is to control for country of origin interacted with female fixed effects. With this strategy, we can control for characteristics of the origin country that are specific to women, thereby encapsulating the origin country characteristics that are most relevant to the question at hand. Column (5) of Table 1 presents the results.

3.3. Results

Table 1: Correlations between female labor participation and sex-based gender system

	Dependent variable: Female Labor Participation				
	(1)	(2)	(3)	(4)	(5)
Sex-Based	0.060*** [0.002]	0.037*** [0.005]	0.058*** [0.013]	0.086** [0.042]	0.063 [0.043]
Female	-0.173*** [0.002]	-0.167*** [0.002]	-0.166*** [0.002]	-0.166*** [0.002]	0.026 [0.060]
Female x Sex-Based	-0.063*** [0.003]	-0.069*** [0.003]	-0.069*** [0.003]	-0.069*** [0.003]	-0.029*** [0.010]
Socioeconomic Controls	Yes	Yes	Yes	Yes	Yes
Country of Origin FE	No	Yes	Yes	Yes	Yes
Language Fam. FE	No	No	Yes	Yes	Yes
Language Area FE	No	No	Yes	Yes	Yes
Language Fam FE x SB	No	No	No	Yes	Yes
Language Area FE x SB	No	No	No	Yes	Yes
Country x Female FE	No	No	No	No	Yes
Observations	674,476	669,739	669,720	669,720	669,720
R-squared	0.296	0.304	0.304	0.304	0.312

Notes: Estimates are survey weighted. Sample includes all immigrants aged 16 and above who report speaking a language other than English in the home. Additional controls include time since immigration, household income, household size, age, age squared, number of children, log wages, and indicators for survey wave, level of English language proficiency, marital status, student status, race and ethnicity, education level, and state of residence. Robust standard errors are in brackets. Source: Results calculated using the 2007-2011 ACS. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The results in column (1) show that compared to male migrants speaking a language lacking a sex-based gender system, male migrants speaking a language that has a sex-based gender system are 6.0 percentage points more likely to be in the labor force. In comparison, similar female migrants are 6.3 percentage points less likely to be in the labor force. This discrepancy is in addition to the average gap in labor force participation between male and female migrants of 17.3 percentage points.

Controlling for the country of origin (column (2)) and language relatedness (columns (3) and (4)) alters the magnitude slightly but does not remove the significance of the results, suggesting that there is not much heterogeneity in the relationship between labor participation and language across origin countries, linguistic families and linguistic areas in this context. Finally, controlling for the interaction between country of origin and female reduces the magnitude of the coefficient of interest. Women speaking a language with a sex-based gender system are 2.9 percentage points less likely to be in the labor force than similar women speaking a language without a sex-based gender system. The coefficient on the interaction term is still significant at the 1% level.

4. Discussion

4.1. *Implications for the coevolution of language and behavior*

Our application and analysis has centered on presenting a set of simple yet powerful strategies that the epidemiological approach makes possible. Studying migrant populations has several additional advantages that researchers interested in the study of language evolution and its relation to non-linguistic phenomena may find useful. Our example demonstrates that the correlation between gender in language and female labor force participation is robust to controlling for country of origin and for language relatedness. Yet, the magnitude of the coefficient is substantially reduced when controlling for female specific country fixed effects. This implies that language must have evolved partly as a result of cultural change, but also that it may have directly constrained the evolution of norms, even if to a smaller extent.

4.2. *External versus Internal Validity of Different Approaches*

While they propose a series of series of empirical tests to be applied to cross-cultural data, Roberts et al. (2015) conclude that “experiments or case-studies would be more fruitful avenues for future research on this specific topic, rather than further large-scale cross-cultural correlational studies.” We agree that there is much promise in experimental research. At the same time, while laboratory experiments arguably have a strong internal validity, they may not perform well in terms of external validity. The non-generalization of the results from lab experiments has been the subject of intensive research in economics (e.g., Stoop et al, 2012, Abeler & Nosenzo, 2014).

At the other extreme, cross-cultural studies perform well in terms of external validity by nature, but they are more likely to suffer from internal validity problems, as Roberts et al. (2015) makes clear. We thus place the epidemiological approach in the middle ground in terms of both external and internal validity. While the environment is not perfectly controlled by the researcher, migrants speaking different languages are observed within the same institutional environment. On the other hand, while findings are more generalizable than for lab or even framed field or natural experiments, migrants are a selected pool that may differ from the native populations.

While all approaches have advantages and disadvantages, the epidemiological approach provides researchers with an opportunity that should not be neglected. This is because (1) it provides a middle ground between cross-cultural correlations and experiments in terms of validity and (2) because it provides a rich new setting with which to test the relation between language evolution and non-linguistic phenomena.

Acknowledgments

All authors equally contributed. The feedback of two anonymous reviewers and of Sean Roberts has been crucial to improve the quality of the manuscript significantly. We are grateful to Ramon Ferrer-i-Cancho for helpful comments. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

References

- Abeler, J., & Nosenzo D. (2014). Self-selection into laboratory experiments: pro-social motives versus monetary incentives. *Experimental Economics*, 18(2), 195-214.
- Chen, M.K. (2013). The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets. *American Economic Review*, 103(2), 690-731.
- Dryer, M. S., & Haspelmath, M. (2013). WALS Online, Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Fernández, R. (2007). Alfred Marshall lecture women, work, and culture. *Journal of the European Economic Association*, 5(2-3), 305–332.
- Hicks, D. L., Santacreu-Vasut, E. & Shoham, A. (2015). Does mother tongue make for women's work? Linguistics, household labor, and gender identity. *Journal of Economic Behavior & Organization*, 110, 19–44.
- Nichóls, J., Witzlack-Makarevich, A. & Bickel, B. (2013). The autotyp genealogy and geography database: 2013 release.
- Roberts, S. G., Winters, J. & Chen, K. (2015), Future tense and economic decisions: controlling for cultural evolution. *PloS one*, 10(7).
- Roberts, S. & Winters, J. (2013), Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PloS one*, 8(8), e70902.
- Stoop, J., Noussair, C. & van Soest, D. (2012). From the Lab to the Field : Cooperation among Fishermen. *Journal of Political Economy*, 120(6), 1027-1056.

THE IMPACT OF COMMUNICATIVE NETWORK STRUCTURE ON THE CONVENTIONALIZATION OF REFERRING EXPRESSIONS IN GESTURE

MATTHEW HALL^{1,2}, RUSSELL RICHIE^{2,3,4}, MARIE COPPOLA^{1,2,3}

¹*Department of Linguistics*

²*IGERT in Language Plasticity*

³*Department of Psychological Sciences*

⁴*The Connecticut Institute for the Brain and Cognitive Sciences*

University of Connecticut,

Storrs, CT, USA

matthew.l.2.hall@uconn.edu, russell.richie@uconn.edu, marie.coppola@uconn.edu

The emergence of referring expressions is a critical component of the evolution of any linguistic system. Building on evidence from naturally emerging sign languages as well as computational simulations, we use a behavioral experiment to investigate how the structure of a communicative network influences the processes of conventionalization of referring expressions. We asked hearing individuals who do not have experience with a sign language to engage in a gestural communication task, and randomly assigned them to either a sparsely connected or richly connected network. Pairwise conventionalization was consistent in both conditions, but network-wide conventionalization was greater in the richly connected network. This is the first time this effect has been demonstrated in a controlled experiment in which humans communicate in a natural linguistic modality (i.e. gesture). Differences in the number of communicative interactions may account for the network effect in the present data; results in the literature are mixed on this point.

1. Introduction

1.1. *Conventionalization*

The role of social convention in creating referring expressions (i.e. names for things) has been a question of interest since as early as Plato. Within modern scholarship, there have been observational studies of emerging languages, computational simulations, and controlled experiments in various non-linguistic modalities (for reviews, see Meir et al., 2010; Steels, 2011, and Galantucci et al., 2012, respectively). However, there is a curious lack of experimentation on how referring expressions emerge in communities and modalities in which new languages are known to arise: that is, small numbers of human beings

communicating in the manual modality (silent gesture). By linking perceptual form to conceptual meaning, referring expressions are arguably the initial entry point into language structure. Only once they are in place (even if not fully conventionalized) can we begin to observe the emergence of other kinds of structure (e.g. syntax, phonology).

1.2. Referring Expressions

Imagine that you wish to communicate the notion of an avocado, but neither you nor your interlocutor has a word for it, despite having knowledge about it. One natural step would be for you to communicate the concept by referencing your knowledge about it as a string of semantic features: for example, its physical properties and affordances, customary uses, emotional valence, etc. Such a string would be the beginning of a referring expression. Objects in the world have a great variety of features that could be listed, of which any given speaker will choose only a finite number (e.g. “it’s kinda oblong, it’s green, and you eat it”). Moreover, it is likely that your interlocutor might initially choose different features (e.g. “you slice it, take out the pit, and eat it”). This type of communication can be effective to a degree, but is a far cry from the lexicalized symbol “avocado”. By what processes do these initially idiosyncratic referring expressions come to be shared, not only by a pair of interlocutors, but by a broader community?

As different speakers interact, they might each retain their own preferred expression, which would minimize the cognitive burden on the producer, but risks being communicatively ineffective, or at least inefficient, if the perceiver’s preferred expression is different. Alternatively, if a hypothetical producer wants to communicate with a familiar perceiver, she could retrieve that individual’s preferred expression from memory. This might increase communicative success but would (presumably) be representationally costly, especially as the number of potential interlocutors and potential referents grows. Another possibility is for our producer to allow her own mental proto-lexicon to be updated by the various interactions she has with multiple interlocutors. She may observe an innovation that is particularly effective, or that some features are becoming more common than others, and then update her own preferred referring expression accordingly. A population of such agents will eventually converge on a stable mapping of forms, in terms of which semantic features are represented (Richie et al., 2014). Along the way, reduction may also happen, in which excess verbiage is pruned away and the iconic links between form and meaning give way to perceptual/motor ease, increasing apparent arbitrariness in the signal and perhaps enabling the emergence of phonology in the traditional sense.

1.3. Extant Data

Processes resembling the above have been reported in at least two emerging sign languages: Al-Sayyid Bedouin Sign Language (ABSL; Sandler et al., 2005) and Kenyan Sign Language (KSL; Morgan, 2015). Referring expressions exist (and have existed for decades) in both languages, but they have not completely conventionalized throughout the community, at the level of either the semantic features represented or in their phonological form. Interestingly, Nicaraguan Sign Language (NSL), which is even younger than ABSL and KSL, appears to have fully conventionalized its repertoire of basic referring expressions (Richie et al., 2014). What might account for the varying speed at which these naturally-emerging systems conventionalize?

1.4. Communicative Network Structure

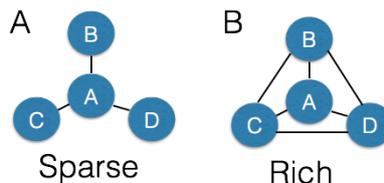


Figure 1. Schematic illustration of sparsely- and richly-connected networks.

Richie et al. (2014) proposed that one important factor might be the structure of the communicative network. They compared users of homesign, in which a single deaf individual (Fig. 1A, center) uses gestural referring expressions to communicate with their hearing friends and family (who use spoken language, rather than homesign, when communicating with one another; Fig. 1A, periphery) to users of NSL, who all use NSL to communicate with each other (Fig. 1B). Using computational simulations, they showed that richly connected networks (NSL-like) do in fact conventionalize faster than sparsely connected networks (homesign-like). However, such naturalistic and computational results are largely suggestive. The naturalistic data contain too many confounding variables to single out network structure as a causal factor, and the computational model only shows what obtains given certain (well-motivated) assumptions put into the model. Thus, it remains to be seen whether actual human beings, in a carefully controlled experimental setting, will show a similar pattern when randomly assigned to a richly- or sparsely connected network. This experimental approach also allows us to control for other relevant sociocommunicative variables that differ between homesign and NSL.

Table 1. Sequence of dyadic interactions in the two conditions.

<i>Round</i>	<i>Sparse Network</i>	<i>Rich Network</i>	
1	A-B	A-B	C-D
	A-C	A-C	B-D
	A-D	A-D	B-C
2	A-B	A-B	C-D
	A-C	A-C	B-D
	A-D	A-D	B-C
3	A-B	A-B	C-D
	A-C	A-C	B-D
	A-D	A-D	B-C
4	A-B	A-B	C-D
	A-C	A-C	B-D
	A-D	A-D	B-C

1.4 Present Study

We asked hearing undergraduates who had no experience with sign language to engage in a dyadic gestural communication task. Fourteen groups of four naïve participants each were randomly assigned to either a sparsely- or richly-connected condition; each participant was also assigned a “letter” within that condition, indicating their position in the network: A, B, C, or D. Dyads then proceeded as shown in Table 1. Participants took turns producing and comprehending gestured descriptions of real-world objects. Each participant had a booklet displaying a target stimulus to describe, as well as an array of 25 images corresponding to the possible items that their interlocutor might describe. The 25 images were identical for both partners, but ordered differently.

After a dyad had described all 25 images to each other, they switched partners. The first “round” was completed once each participant had communicated with all assigned interlocutors. Participants completed two rounds on day 1, followed by two additional rounds approximately one week later, for a total of four interactions with the same interlocutors. Table 1 shows the sequence of dyads in both conditions. Note that the interactions involving “A” (in bold) are identical in both conditions. These form the core of our analysis.

For each description, we coded the semantic features that were produced between trial onset and when the interlocutor (correctly or incorrectly) selected an image from the target array. We measured conventionalization by computing the similarity between strings of semantic features. The similarity between two unordered strings was quantified by the Jaccard index: the ratio of their intersection to their union. For example, the strings {a, b, c, d, a} and {c, d, x} have a Jaccard index of 2/5 (0.4), because their intersection contains 2 unique elements (c, d) and their union contains 5 unique elements (a, b, c, d, x). Thus, a

Jaccard index of 0 reflects complete divergence, while 1 reflects complete convergence, regardless of length.

2. Results

2.1. Comprehension accuracy.

Despite having no previous experience with sign language or pantomime, participants communicated successfully most of the time. Accuracy increased over time, as expected, from means of 83% and 78% after Round 1 to means of 96% and 98% after Round 4 (for rich and sparse conditions, respectively). There was no main effect of condition [$F(1,12) = .15, p = .70$], but there was a marginally significant interaction between condition and round [$F(1,12) = 4.51, p = .06$], with the sparsely connected network showing greater improvement from Round 1 to Round 4.

2.2. Direct comparisons: Pairwise conventionalization

Table 2. Direct comparisons measure dyadic conventionalization. Indirect comparisons measure network-wide conventionalization.

<i>Direct</i>	<i>Indirect</i>
A-to-B vs. B-to-A	B-to-A vs. C-to-A
A-to-C vs. C-to-A	B-to-A vs. D-to-A
A-to-D vs. D-to-A	C-to-A vs. D-to-A

To measure conventionalization over time within any given dyad, we plot the average Jaccard index for that dyad from Round 1 to 4 (Table 2 & Figure 2, left). Unsurprisingly, conventionalization increases substantially [$F(1,12) = 247.41, p < .001$]. We also found no significant difference between the richly- and sparsely-connected networks [$F(1,12) = .33, p = .57$]. Note, however, that under this analysis, high values could reflect dyads establishing a convention that was unique to them as a pair, rather than a network-wide convention.

2.3 Indirect comparisons: Network-wide conventionalization

To assess the extent to which conventionalization increased in the network *as a whole*, we measure the similarity of strings that were *not* produced in direct communication, but rather in indirect communication (Table 2 & Figure 2, right). Specifically, we compare the similarity of the strings produced to participant A by participants B, C, and D. Increases in this measure cannot be due to partner-specific effects, but must reflect increased conventionalization throughout the network. Here, we again find a general increase across rounds in both conditions [$F(1,12) = 197.79, p < .001$]. However, we now find that the

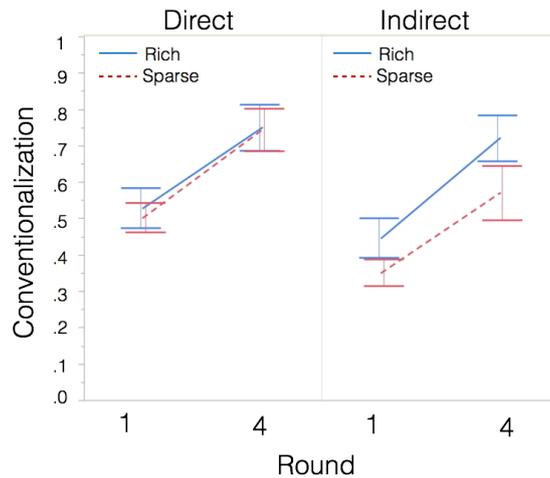


Figure 1. Conventionalization at the beginning (Round 1) and end (Round 4) of the experiment. Direct conventionalization (left) measures similarity between members of a given interacting pair (A & B, A & C, A & D). Indirect conventionalization (right) measures similarity among utterances that were produced by different participants (B, C, D) to a common interlocutor (A). Error bars represent 95% confidence intervals.

fully-connected network shows more conventionalization than the sparsely connected network [$F(1,12) = 17.82, p < .01$]. This is the central result.

2.3. Number of interactions

The advantage for the full over sparse network was observed even in Round 1, and there was no difference in the rate of increase from Round 1 to Round 4 [$F(1,12) = 2.48, p = .14$]. This was unexpected. However, one consequence of the network structure manipulation was that the fully connected network entailed, by design, *twice as many communicative interactions per round as the sparsely connected network* (see Table 1). To address whether this simple property could in fact be the mechanism underlying the group difference, we compare the end-state of the sparsely connected network (Round 4, after 12 interactions) against the midpoint of the richly connected network (Round 2, after 12 interactions). There is no difference in conventionalization between the two groups once number of interactions is controlled [$F(1,12) = .21, p = .66$], suggesting that the greater number of interactions in the richly connected network may be responsible for the bulk of the effect. To reiterate, it appears that the fully- and sparsely-connected networks achieve similar levels of group-wide conventionalization after comparable numbers of interactions.

3. Discussion

The emergence of referring expressions is an important milestone in the development of any language-like system. Early on, different members of a community might use expressions that differ in both their semantic content and their form; however, conventionalization is likely to happen over time. We used actual human behavior in a natural communicative modality to show that the structure of the communicative network can influence the process of semantic conventionalization. Richly-connected communities (where agents communicated with all other agents using the emerging system) were more fully conventionalized (at least semantically) than sparsely-connected communities (where one agent used the emerging system with all others, but those others did not communicate amongst themselves in the emerging system).

In our data, this effect appears to be driven primarily by an epiphenomenon of the network structure manipulation: the fact that, all else being equal, a richly connected network will involve more communicative interactions than a sparsely connected network. This finding contrasts with some previous computational and behavioral work: Judd et al. (2010) found that humans reached consensus on an arbitrary choice of a color fastest when connected in a richly connected network, an effect unexplained by number of interactions. Similarly, simulations by Richie et al. (2014) showed an effect of network structure even after controlling for the number of interactions.

These contrasting effects reveal the subtleties of network effects, and illustrate the insufficiency of our intuitions about what networks might facilitate collective behavior. This point is even further driven home by consideration of simulation results by Gong, Baronchelli, Puglisi, and Loreto (2012), also a simulation study into the effects of social network structure on the emergence of language. In their model, agents must carve a perceptual continuum (color) into perceptual categories, and then agree upon labels that refer to one or more perceptual categories. In contrast to our model, where the communicating agents know the referent (e.g., avocado) but adjust the probabilities of corresponding gestures, Gong et al.'s agents must infer the referent through perception, and they found that sparsely- and richly-connected networks offered comparable convergence properties. Clearly, even though all the foregoing work involves conventionalization or consensus-building of some kind, network effects seem to be highly sensitive to assumptions/simplifications made by models, and/or the precise nature of the task set before participants.

The present results generate a hypothesis about why some naturally-emerging sign languages appear to conventionalize their referring expressions (at the semantic level) more quickly than others: they may simply be communicating more. However, this remains to be empirically verified, and other alternatives remain to be explored. For example, it could also be that conventionalization within a local community (e.g. a single family) could impede larger-scale conventionalization among members of different families,

schools, villages, etc. Because the present study tested only a single community in which all members interacted, the results cannot yet speak to this issue.

3.1. *Future directions*

To date, we have focused on characterizing conventionalization at the semantic level. Future work will explore conventionalization at the level of form. We are particularly interested in testing the hypothesis that the emergence of syntax may not require that referring expressions be conventionalized, but the emergence of phonology might. This would be consistent with reports of emerging sign languages in which syntactic structure is evident, but phonological structure is not. In addition, simulating the impact of local conventionalization on global conventionalization is a major target for both computational and behavioral experiments.

References

- Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental semiotics. *Language and Linguistics Compass*, 6(8), 477-493.
- Gong, T., Baronchelli, A., Puglisi, A., & Loreto, V. (2012). Exploring the roles of complex networks in linguistic categorization. *Artificial Life*, 18, 107-121.
- Judd, S., Kearns, M., & Vorobeychik, Y. (2010). Behavioral dynamics and influence in networked coloring and consensus. *PNAS*, 107(34), 14978-14982.
- Meir, I., Sandler, W., Padden, C., & Aronoff, M. (2010). Emerging sign languages. *Oxford Handbook of Deaf Studies, Language, and Education*, 2, 267-280.
- Morgan, H. (2015). When does a word emerge? Lexical and phonological variation in a young sign language. *San Diego Linguistic Papers*, 5, 2-17.
- Richie, R., Yang, C., & Coppola, M. (2014). Modeling the emergence of lexicons in homesign systems. *Topics in Cognitive Science*, 6, 183-195.
- Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory*, 29(2), 503-543.
- Steels, L. (2011). Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4), 339-356.

DEICTIC TOOLS CAN LIMIT THE EMERGENCE OF REFERENTIAL SYMBOL SYSTEMS

ELIZABETH IRVINE

*School of English, Communication and Philosophy, Cardiff University
Cardiff, UK
irvinee@cardiff.ac.uk*

SEÁN G. ROBERTS

*Language and Cognition department, Max Planck Institute for Psycholinguistics
Nijmegen, Netherlands
sean.roberts@mpi.nl*

Previous experiments and models show that the pressure to communicate can lead to the emergence of symbols in specific tasks. The experiment presented here suggests that the ability to use deictic gestures can reduce the pressure for symbols to emerge in co-operative tasks. In the *gesture-only* condition, pairs built a structure together in *Minecraft*, and could only communicate using a small range of gestures. In the *gesture-plus* condition, pairs could also use sound to develop a symbol system if they wished. All pairs were taught a pointing convention. None of the pairs we tested developed a symbol system, and performance was no different across the two conditions. We therefore suggest that deictic gestures, and non-referential means of organising activity sequences, are often sufficient for communication. This suggests that the emergence of linguistic symbols in early hominids may have been late and patchy with symbols only emerging in contexts where they could significantly improve task success or efficiency. Given the communicative power of pointing however, these contexts may be fewer than usually supposed. An approach for identifying these situations is outlined.

1. Introduction

Gesture, and pointing in particular, is often heralded as a crucial step in the evolution of language (Arbib, Liebal, & Pika, 2008; Tomasello, Carpenter, & Liszkowski, 2007; Tomasello, 2010; Liszkowski, 2010). Pointing by itself is versatile enough to communicate about activities, objects, people, events, places and more, and can be made sophisticated when combined with theory of mind, advanced pragmatic inference, and abilities to engage in displaced reference. Used with attention-getters, pointing is more powerful still. One might wonder then, under what conditions symbols are necessary to complete complex collaborative tasks, given a pre-existing deictic system.

Experiments and models using a director-matcher tasks suggest that referential symbol systems routinely emerge (Steels, Belpaeme, et al., 2005; Verhoef,

Roberts, & Dingemans, 2015; Skyrms, 2010). In these tasks, a director requests an object with a signal and a matcher must guess which object they want. They are then given feedback indicating the correct answer. Stable symbols for objects usually emerge over time through interaction. However, the availability of communicative tools in many of these experiments is managed such that the only way to communicate (and the only way to reliably complete the task) is with referential symbols. In this case, it is not surprising that they emerge.

In addition, setting up a novel symbol system within the context of a task is costly, as it takes time and energy that could be spent completing the task. In this case, if a symbol system does not offer a sufficient pay-off in terms of increasing task efficiency that can offset this cost, then players may be biased against setting up such a symbol system. In director-matcher tasks, the costs and pay-offs are fairly straightforward: symbols are the only way to reliably complete the matching task (high pay-off), but they take time to stabilise so many early trials are unsuccessful (high cost). In situations with more freedom, the costs and pay-offs can be more complex, and not favour the emergence of symbols in such a directed way.

Accordingly, in the current experiment we tested whether novel symbol systems emerged, and whether they aided task performance, in a co-operative building task set in an embodied, 3D environment where it was possible to point and gesture. The task was complex enough such that having a symbol system would allow participants to complete the task more efficiently, but symbols were not the only means of communicating. Under these conditions, and in contrast to the experiments above, we found that no symbol systems developed. This suggests that deictic tools, and features of specific tasks (complexity, degree of collaboration required, etc) can affect the likelihood of the emergence of a symbol system. This type of experiment may then also help address debates about the extent to which linguistic abilities can be inferred from complex material artefacts and evidence of cooperation (Davidson & Noble, 1992; Cuthbertson & McCrohon, 2012).

2. Experiment

The experiment tested whether partners would develop a symbol system while performing a reasonably complex co-operative task in Minecraft (Mojang, 2015), a computer game set in a three dimensional virtual world. The experiment had two conditions which manipulated the ways in which partners could communicate. In the *gesture-only* condition partners were taught a pointing convention (gaze at something and jump repeatedly), and were told that they could only communicate using their avatar in the game (so they could use pointing and other gestures). In the *gesture-plus* condition partners were also taught the pointing convention, but were additionally allowed to use a second communication channel: knocking on the table (we do not assume that auditory communication is special; this was just an easy way to implement an additional ostensive channel). This provided a

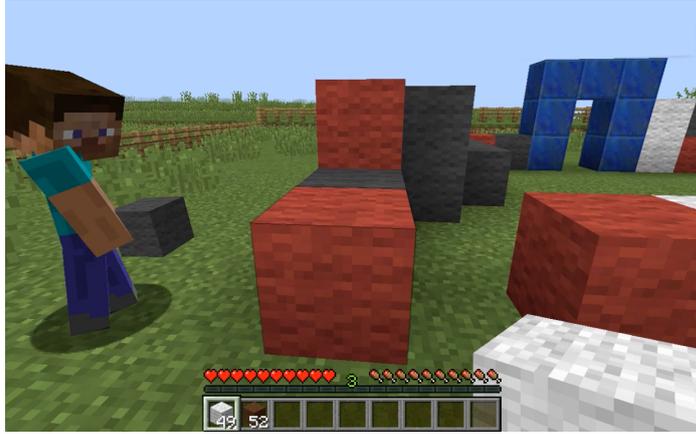


Figure 1. A screenshot from the experiment. The view is from the eyes of a virtual avatar. In the foreground are some of the placed blocks of the structure and one of the doorways. On the left is the other player's avatar, whose head orientation indicates they are looking at the red block in the centre foreground.

discrete signal space which was suited to labelling the discrete objects in the task.

The main hypothesis was that participants in both conditions (*gesture-only*, *gesture-plus*) would perform the task equally well, measured in terms of time taken to complete task, and number of errors made. In addition, we predicted that participants in the *gesture-plus* condition would not use their extra communication channel to develop a referential symbol system (though they might use it in other ways). These predictions stem from the ideas that 1) pointing and other deictic gestures are sufficient for many communicative needs and 2) it is costly to set up novel symbol systems, so they will only emerge if there is a direct pressure to do so.

2.1. Method

Participants built a construction together in Minecraft. Each participant can move around in and manipulate the world via a humanoid virtual avatar and can see their partner's avatar, including their gaze direction. The world is made up of a regular matrix of crate-sized blocks of different kinds (rock, dirt, stone etc.) that can be placed or removed (like life-sized lego, see figure 1).

First, in a training exercise, participants were taught how to move and look around the 3D environment. Then they were taught how to place and remove blocks, and given a plan of a small practice structure to build independently. The next task was to learn a pointing convention, which consisted of gazing directly

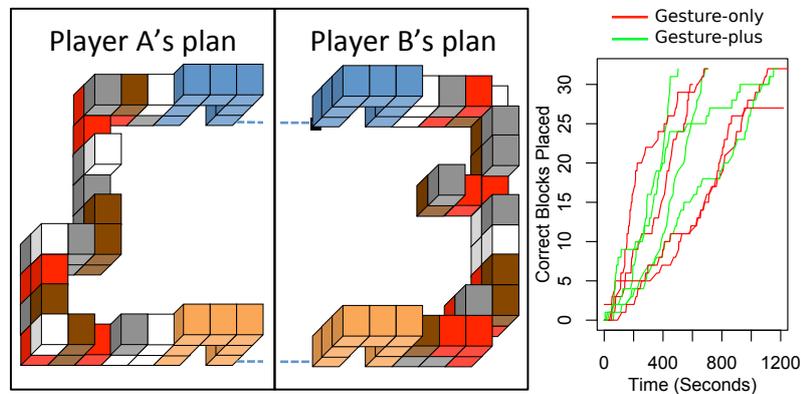


Figure 2. Left: The plan of the structure which player A and player B were given. The two doorways at the top and the bottom were already placed in the world. Right: The rate for correctly placing communicated blocks. Lines are coloured by condition.

at a block and then jumping. Participants practiced ‘pointing’ to a block for their partner. Experimenters checked participants’ progress and answered questions.

The main task involved building a structure from a plan. Each player had only half of the plan for the whole structure (figure 2). Each side of the plan included the same number of blocks, but the plan was not symmetrical. The plan included coloured doorways which were already placed in the world so that the participants had obvious anchor points to use. One of each type of block was also placed to one side of the doorways (figure 3), though participants’ attention was not directed to these. Each side of the plan was composed of four different coloured blocks (white, brown, grey and red - made of Minecraft ‘dyed wool’ which is easy to destroy). One player was given only white and brown blocks, while the other was given only grey and red blocks.

Participants were told that they could not place block types belonging to their partner, but they could destroy any blocks. Participants were then told that they would have to communicate with their partner to get them to place their coloured blocks in the appropriate places. Participants were told to complete the task in 20 minutes (but were given as much time as they needed).

The key manipulation was in the ways that participants could communicate. Participants in the *gesture-only* condition could only communicate via the actions of their avatar in the game. In the *gesture-plus* condition, participants were additionally allowed to knock on the desk with their knuckles in order to communicate. After the experiment, participants were given a questionnaire on their strategies and previous experience with Minecraft. A symbol system was considered to have



Figure 3. A completed structure viewed from the air, with some of the surrounding environment.

developed if participants reported that they could refer to at least one block type using a non-deictic method (either auditory or gestural). Participants played on two MacBook laptops which were connected by Wi-Fi. A CraftBukkit Minecraft server with customised plugins was used to manage the world and record player activity. Video and audio were also recorded for later coding.

2.2. Results

Results for 4 pairs in each condition were analysed. Participants' performance was rated according to the accuracy with which they matched the plan, the amount of time they took and the rate of successfully communicated blocks (blocks correctly placed by a player on their partner's side of the plan, for which communication was necessary).

All participants were essentially at ceiling in terms of accuracy. 5 trials had no errors, while the rest had 1, 2 and 10 errors respectively (less than 5% of possible errors in the bounding box of the plan), the latter two from the *gesture-only* condition. The majority of the last case can be attributed to duplicate insertions which amplified the errors (i.e. the errors are lower in terms of edit distance), and this may have more to do with map reading than difficulty communicating. All participants stated that they were satisfied that their final structure matched the plan. Pairs took between 8.3 minutes and 21.9 minutes. While the sample size is not big enough to pick up small differences, the two conditions did not differ significantly in time taken. Figure 2 shows the rate for correct placement of communicated blocks, demonstrating that the two conditions overlap. Participants varied in their

previous experience with Minecraft, but this did not predict success at the task.

Finally, there was minimal use of the extra communication channel (knocking). One pair did not use the channel at all. Two pairs used knocks as attention-getters at crucial times in the game (though used only once by one pair, and twice by another). One pair made more regular use of knocking, but knocks were not used referentially: one partner used it as an attention getter and to mark progress (e.g. roughly 'that's correct, now move on'), and the other partner similarly used it to mark communicative success or lack of it (roughly 'I understand' and 'I don't understand'). There was no overlap in the 'meanings' of knocks across this pair, and they were only used when Player A was directing the placement of Player B's blocks. Interestingly, the pair who did not use the channel at all had the shortest completion time. These participants reported that "I didn't think we needed to [use sound]" and "we found other ways to communicate what we wanted to say".

As reported in questionnaires, at least two participants attempted to develop a symbol system for colour in the *gesture-only* condition. One tried to code for colour via number of jumps, and another tried to code for the colour of their partner's block via the colour of the block they were holding (white means red, brown means grey), but their partners either did not notice, or did not understand, and these symbols were not adopted.

In general however, it is hard to identify trends. There was massive variation in the way communicative conventions were established (or not). Sometimes conventions were shared across a pair, such that the signs used by the first 'director' (communicating where they wanted differently coloured blocks placed) were copied by the second 'director'. Sometimes these signs were streamlined by the second director (e.g. only one sign each for location and colour, or a smaller meaning space), and sometimes they were added to (more ways to convey the same meaning, and a bigger meaning space). Some partners had signs for 'correct' and 'wrong' but some did not.

One clear result is that the same gestures were often used to convey multiple meanings within a pair (even to one sign meaning both 'that's right' and 'no, stop!'). The meaning intended was usually obvious from context. Deictics aimed at coloured blocks, used to convey both colour and location were usually easily disambiguated via context, and via the universally adopted system of referring to colour first, and location second. See the supplementary materials for video examples of various phenomena above.

Also, the fastest pairs all adopted a common strategy of completing their own side of the building as far as they could by themselves, then working with a partner, completing one colour at a time. This was efficient both in terms of time, and in terms of communicative effort: if only one colour is being placed at a time, then the director only needs to communicate about location.

3. Discussion

Our hypotheses regarding the use of pointing in reasonably complex co-operative tasks were confirmed: no pair developed a symbol system, and providing an extra communication channel did not aid performance. Instead, pairs used both the pointing convention they had been taught, and developed other deictic and non-referential signs (13/16 participants, or 7/8 pairs, used other deictic gestures), which were often fairly fluid in their meanings.

Digging deeper into questionnaires and debriefings to see why no symbol systems were developed in either condition, something like a paradox emerges. The less skilled or practised a participant is at the task, the less easily they can move in a precise way, the less efficiently they perform the task (so e.g. fail to use the one-colour-at-a-time strategy), and so the more they need symbols to minimise communicative effort. But given their lack of control over precise movements, setting up a symbol system is hard and costly. As one participant noted: “not being able to very accurately control [my avatar]... the risks of miscommunication are far too great”. Even if it had been possible to set up a simple symbol system, being physically unable to reliably repeat it would have hampered its use, and probably led to it being abandoned. Another participant explained that a symbol system “would have taken much longer to develop”, highlighting the cost involved. Instead, a range of fairly direct deictic gestures (hitting and jumping on blocks) were more easily performed and interpreted, and their precise form (e.g. number of jumps) does not matter.

Conversely, the more skilled or practised a participant is, the more easily they can move precisely, the more efficiently they do the task (e.g. use the one-colour-at-a-time strategy), and so the less they need symbols. They are more physically capable of setting up a symbol system, but use game strategies that minimise the need to communicate. In this case, deictics and non-referential signs, combined with highly efficient task strategies, are entirely sufficient for their needs.

Combinations of deictic and non-referential gestures (e.g. affirmatives) may then function as a local fitness peak in the context of some tasks. Those at the top of the local fitness peak can use deictics and other non-referential gestures to successfully and efficiently complete these tasks, and would not gain much by paying the costs of setting up a symbol system (switching to the global fitness peak). Those at the bottom of the fitness landscape would benefit most by developing a symbol system, but doing so would require developing their deictic skills, which would trap them in the local peak of the first group. In neither case then is a symbol system likely to emerge within some kinds of tasks, where deictic and other non-referential gestures can form a kind of ‘**deictic fitness trap**’.

Intriguingly, there was also evidence of the emergence of ways to organise action sequences. All but 3 participants used signals that helped to segment the task structure (‘yes, now move on’ or ‘no, that’s wrong’), even when these were

not necessary. One pair also developed something like a repair sequence, entirely mediated through pointing (see the supplementary materials). However, they are clearly helpful when participants must self-organise to complete sequences of sub-tasks. Previous models and experiments rule out the need for sequence organisation by assigning roles to participants (e.g. in director-matcher tasks), but this experiment suggests that signals to aid with sequence organisation emerge before symbols.

This can be used to suggest that symbol systems may have been late and patchy in emerging, and tied to features of specific tasks, though preceded by and existing along with deictic and non-referential communication conventions, and backed by well-developed theory of mind and abilities for pragmatic inference. More experimental research is needed on exactly what sort of tasks, relevant to hominid lifeways, would have absolutely required symbolic communication, and which would not. Classic signal guessing games are a poor framework in this regard, since if pointing was available, it would often make the development of symbols redundant. Embodied, cooperative and (somewhat) ecologically valid tasks such as the one in this experiment force researchers to specify which concrete tasks or social goals demand the use of symbol systems.

An alternative approach comes from (Sterelny, 2014, 2015), who identifies major new pressures on communication that would likely have acted on fission-fusion hominid groups present between 500 and 100kya (moving into sapiens and Neanderthals). These hominids would have needed to track economic exchanges covering a large range of resources over extended periods of time (e.g. how much is a leg of deer from last week worth now in tubers), and track others' reputation so as to plan future exchanges.

In this case a combination of pressures may lead to the emergence of symbols: 1) the need to refer to the distant elsewhere and elsewhere including 2) a wide range of resources that must be differentiated, among 3) individuals who do not share a huge amount of common ground (because of the fission-fusion structure). In contrast to tasks carried out in the here-and-now, in which deictics embedded in routines would likely have sufficed for communication, symbols would make communication far more efficient in these economic exchanges, and come with a start up cost worth paying. Identifying such social and communicative constraints on hominid lifeways is likely to be a productive way of framing specific hypotheses about symbol emergence, situated as they are within the activities and social contexts in which hominids had to function.

Acknowledgements

SR is supported by an ERC Advanced Grant No. 269484 INTERACT to Stephen Levinson. We would like to thank the Philosophy Department and the AHR Centre of Excellence for the Dynamics of Language at ANU for hosting the authors while these experiments were carried out.

References

- Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, 49(6), 1053–1076.
- Cuthbertson, C., & McCrohon, L. (2012). Sea crossings are an unreliable indicator of language ability in hominids. In T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, & J. R. Hurford (Eds.), *The Evolution of Language: Proceedings of the 9th International Conference (EVOLANG9)* (pp. 88–95). World Scientific.
- Davidson, I., & Noble, W. (1992). Why the first colonisation of the Australian region is the earliest evidence of modern human behaviour. *Archaeology in Oceania*, 27(3), 135–142.
- Liszkowski, U. (2010). Deictic and other gestures in infancy. *Acción psicológica*, 7(2), 21–33.
- Mojang. (2015). Minecraft 1.8.8. Retrieved from <https://minecraft.net/10/08/2015>.
- Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford University Press.
- Steels, L., Belpaeme, T., et al.. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and brain sciences*, 28(4), 469–488.
- Sterelny, K. (2014). A paleolithic reciprocation crisis: symbols, signals, and norms. *Biological Theory*, 9(1), 65–77.
- Sterelny, K. (2015). Deacon's challenge: From calls to words. *Topoi*, 1–12.
- Tomasello, M. (2010). *Origins of human communication*. MIT press.
- Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child development*, 78(3), 705–722.
- Verhoef, T., Roberts, S. G., & Dingemans, M. (2015). Emergence of systematic iconicity: Transmission, interaction and analogy. In D. Noelle, et al. (Ed.), *The 37th annual meeting of the Cognitive Science Society (CogSci 2015)*.

INFERRING THE WORLD TREE OF LANGUAGES FROM WORD LISTS

GERHARD JÄGER

*Institute of Linguistics, Tübingen University
Tübingen, Germany
gerhard.jaeger@uni-tuebingen.de*

SØREN WICHMANN

*Leiden University Center of Linguistics, Leiden University
Leiden, The Netherlands
and
Laboratory of Quantitative Linguistics, Kazan Federal University
Kazan, Russia
wichmannsoeren@gmail.com*

Since its launch in 2007, the *Automated Similarity Judgment Program* has collected basic vocabulary lists from more than 6,000 languages and dialects, covering close to two thirds of the world's languages. Using these data and phylogenetic techniques from computational biology, such as weighted sequence alignment and distance-based phylogenetic inference, we computed a phylogenetic language tree covering all continents and language families. Our method relies on word lists in phonetic transcription only, i.e. it does not rely on expert cognacy judgments. This decision enabled us to perform inference across the boundaries of language families. The world tree of languages thus obtained largely recaptures the established classification of languages into families and their sub-groupings. Additionally it reveals intriguing large-scale patterns pointing at a statistical signal from deep time.

1. Introduction

Hardly any element of human culture is as highly susceptible to vertical transmission and preserved as faithfully over time as language (Holman, Wichmann, Brown, & Eff, 2015). For these reasons language relations offer a unique framework for the study of cultural processes: by projecting such processes onto linguistic phylogenies ancestral states and horizontal transmission events can be inferred (Mace & Holden, 2005). Existing linguistic phylogenies, however, impose a limitation on such exercises. Both the traditional comparative method of historical linguistics and character-based methods from the modern biological toolkit offer tools for classifying languages, but they are only applicable to groups of languages that have already been demonstrated to be related. In this paper we

present cutting-edge distance-based methods which do not rely on this assumption and therefore allow for determining a relationship between any pair of languages drawn from the pool of the entire global linguistic diversity. More specifically, using the so-called ASJP database (Wichmann et al., 2013), we employ pairwise sequence alignment and distance-based phylogenetic inference to infer a tree of c. 6,000 languages and dialects, covering all continents and language families. This tree correctly identifies most established language families to a very good approximation and recovers their assumed internal structure with high accuracy. Additionally, it reveals a signal of common descent or contact beyond the level of established families. A quantification of linguistic distances such as the one which is tested here through the inference of a world language tree, promises to bridge gaps between historical linguistics and other disciplines within the social sciences.^a

2. The Automated Similarity Judgment Program

Since its launch in 2007, the collaborative project known as the Automated Similarity Judgment Program (ASJP) has achieved the compilation of a database of close to two thirds of the world's languages consisting of 40-item lists of universally stable lexical concepts (Holman et al., 2008). Publications drawing upon these data have mostly employed a modification of the Levenshtein distance called LDND (Wichmann, Holman, Bakker, & Brown, 2010) in order to compute a linguistic distance between the doculects of the database. The distance measure employed in the present paper demonstrably represents an improvement over LDND. Among the resources published on the ASJP site is (different versions of) an 'ASJP World Language Tree of Lexical Similarity' (see <http://asjp.clld.org/download>), similar in spirit to the tree discussed in the present paper, but based on LDND and vanilla Neighbor-Joining. The trees made available on the ASJP site were never intended as real publications, only as specimens providing some potentially useful insights into the data. In contrast, the tree that we present here is the result of extensive research towards developing an optimal distance measure and finding the most adequate algorithm for inferring a phylogeny based on the distances computed.

^aIn recent years, linguistic distances computed from the ASJP database have become a tool widely used by economists for studying how linguistic differences influence investment, trade, tourism, migration preferences and the L2 proficiency and general success of migrants. A paper by Isphording and Otten (2011) seems to have initiated this trend; cf. (Melitz & Toubal, 2014) for one of many recent examples. For comparative anthropology ASJP distances have also proven useful (Walker, Wichmann, Mailund, & Atkisson, 2014). Additionally, current research to which the present authors have contributed suggests that paleoanthropology and genetics can also profit by introducing an ASJP-derived distance measure for the purpose of correlational studies.

3. Distance measures

We defined two pairwise distance measures between doculects. The first one (taken from Jäger, 2013; see also Jäger, 2015) — called *PMI distance* as it is built on the notion of *Pointwise Mutual Information* between sound strings —, quantifies the lexical similarity between lists using sequence alignment. It aggregates information both about sound changes and the gain/loss of cognate classes. PMI distances are determined via sequence alignment, using differential weights for different symbol pairings. These weights are determined in a data-oriented way via unsupervised learning from the ASJP data.

To estimate the likelihood of sound correspondences, a corpus of *probable cognate pairs* was compiled from the ASJP data using two heuristics. First, a similarity measure between word lists related to the above-mentioned LDND distances was defined and the 1% of all ASJP doculect pairs with highest similarity were kept as *probably related*. (This notion is rather strict; English, for instance, turns out to be “probably related” to all and only the other Germanic doculects. In total, 99.9% of all doculect pairs defined that way belong to the same language family.) Second, the normalized Levenshtein distance was computed for all translation pairs from probably related doculects. Those with a distance below a certain threshold were considered as *probably cognate*. These probable cognate pairs were used to estimate PMI scores. Subsequently, all translation pairs were aligned using the PMI scores from the previous step as weights. This resulted in a measure of string similarity, and all pairs above a certain similarity threshold were treated as probable cognates in the next step. This procedure was repeated ten times. In the last step, approximately 1.3 million probable cognate pairs were used to estimate the final PMI scores.

Again, the similarity threshold being used is rather strict. For illustration, the only probable cognates pair between English and German that were kept during the last iteration are *fiS/fiS* ‘fish’, *laus/laus* ‘louse’, *b13d/blut* ‘blood’, *horn/horn* ‘horn’, *brest/brust* ‘breast’, *liv3r/leb3r* ‘liver’, *star/StErn* ‘star’, *wat3r/vas3r* ‘water’, and *ful/fol* ‘full’. To determine the distance between two word lists, all string similarities in the Cartesian product of the two lists are calculated. The distance between the word lists is a measure of how much the similarities between synonymous words (which are candidates for cognate pairs) exceed the similarity of non-synonymous pairs (i.e. random pairs of words). For more details, see (Jäger, 2013). The full PMI distance matrix is available online at <http://www.evolaemp.uni-tuebingen.de/details.html>.

To calculate the secondary distance measure we represented each doculect as a binary vector representing the presence/absence of bigrams of the 41 ASJP sound classes in the corresponding word lists. The *bigram inventory distance* between two doculects is then defined as the Jaccard distance between the corresponding vectors.

4. Phylogenetic inference

Based on the PMI distance and the bigram inventory distance, a phylogenetic tree was inferred using the *Minimum Variance Reduction* (MVR) algorithm (Gascuel, 2000) as implemented in the R package *ape* (Paradis, Claude, & Strimmer, 2004). Phylogenetic inference proceeds in two steps. First the two distances matrices (PMI distances and bigram inventory distances) are aggregated into a Consensus Distance Matrix using the *super distance matrix* (SDM) method (implemented in *ape*) from (Criscuolo, Berry, Douzery, & Gascuel, 2006). The relative weight of lexical to bigram inventory distances was, somewhat arbitrarily, set to 10:1. In this way it was assured that phylogenetic inference is dominated by the information in the PMI distances, and bigram inventory distances only act as a kind of tie breaker in situations where lexical distances do not provide a detectable signal.

SDM computes an aggregated distance matrix and a variance matrix associated to that distance matrix, which in turn serve as input for MVR. MVR is a modification of the well-known *Neighbor-Joining* algorithm which uses both distances and their estimated variances to compute a tree.

5. The world tree

The full tree is made available online.^b It is summarized in Fig. 1. All clades comprising doculects from the same family (according to the WALS classification, cf. Haspelmath, Dryer, Gil, & Comrie, 2008), with maximally one outlier, are collapsed into a triangle.

Generally, the automatically induced tree captures the established expert classification of languages into families fairly well. Of the 52 WALS families for which ASJP contains at least 10 doculects, 40 families correspond to a clade in the tree with an F-score^c ≥ 0.95 (meaning: the binary classification of taxa induced by that clade has an F-score ≥ 0.95 when evaluated against the extension of that family according to WALS). The 12 poorly recognized families are Trans-New Guinea (maximal F-score 0.61), Sko (0.67), Macro-Ge (0.68), Marind (0.71), Penutian (0.80), Otomanguean (0.81), Torricelli (0.82), Nilo-Saharan (0.85), West Papuan (0.89), Sepik (0.92), Sino-Tibetan (0.92), and Hokan (0.93). Most of these families are controversial.

^bIt can be inspected at <http://www.sfs.uni-tuebingen.de/~gjaeger/ODljNT/worldTree.svg>, using any standard web browser.

^cThe *F-score* is a statistic measuring the goodness of fit of a binary classification. It is defined as the harmonic mean between *precision* and *recall*, where

$$\begin{aligned} \text{precision} &\doteq \frac{\# \text{true positives}}{\# \text{true positives} + \# \text{false positives}} \\ \text{recall} &\doteq \frac{\# \text{true positives}}{\# \text{true positives} + \# \text{false negatives}} \end{aligned}$$

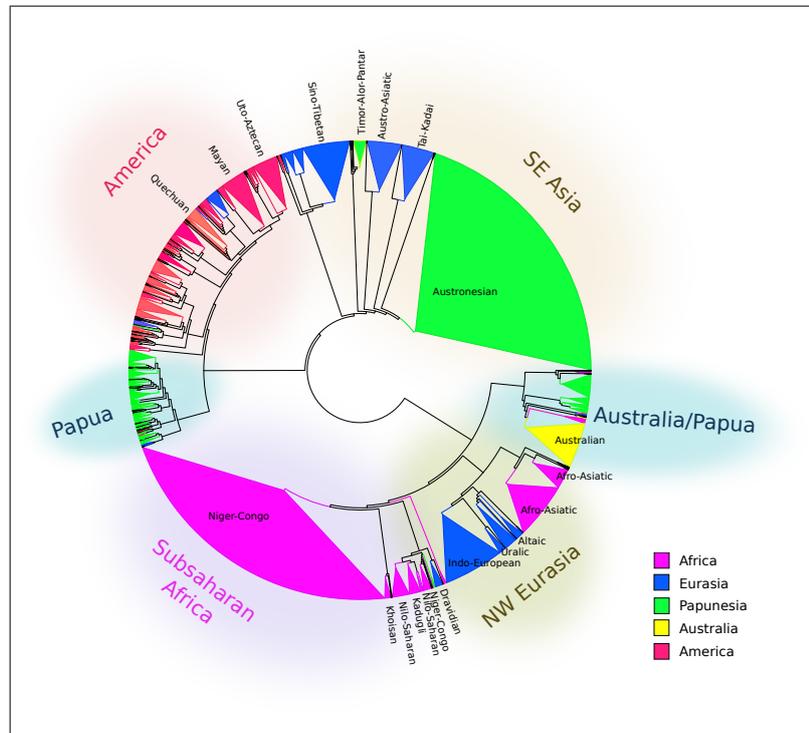


Figure 1. The world tree of languages

For the more conservative Glottolog classification (Hammarström, Forkel, Haspelmath, & Bank, 2015), only 9 of the 59 families with at least 10 members have an F-score below 0.95: Nuclear Trans-New Guinea (0.61), Sko (0.67), Otomanguean (0.81), Nuclear Torricelli (0.82), Nuclear Macro-Je (0.83), Sepik (0.85), Pama-Nyungan (0.88), Afro-Asiatic (0.93), and Sino-Tibetan (0.93).

The internal classification of language families, as assumed by experts, is also recaptured with high accuracy. The *Generalized Quartet Distance* (Pompei, Loreto, & Tria, 2011) between the automatically induced tree and the WALS classification is as low as 0.033 (0.066 for the Ethnologue Lewis, Simons, & Fennig, 2015 and 0.046 for the Glottolog classification).

6. Final remarks

Comparing extant languages in order to infer the evolution of the presently observed linguistic diversity only allows us to see the top of the iceberg in some

detail: we can reconstruct ancestral languages with relatively high precision down to around 5000 years before present. From around 5000 BP to around 10,000 BP the signal becomes increasingly more noisy and eventually gets lost. In this situation, the best we can do in order to reach further back in time is to simultaneously compare the bulk of the world's languages reaching the kind of result shown in Fig. 1. This result indicates the existence of four distinct areas: Africa + Western and Northern Eurasia, SE Asia (including Island SE Asia occupied by Austronesians), the Americas, and Sahul. The Papuan languages of New Guinea are distributed in three separate clusters. One of these is adjacent to Australia and may represent the earliest stratum of Papuan languages; we regard the Australian languages and these Papuan languages as belonging to one Sahul cluster. The two other Papuan clusters resist meaningful interpretations in terms of the regions they occupy in the tree.

Many aspects of the topology in the upper regions of the tree — above the level of established families — evidently reflect sustained contact rather than vertical transmission. For instance, since Vajda (2010) it is widely believed that the Yeniseian languages of Central Siberia are genealogically related to the Athapaskan-Eyak-Tlingit languages of North America, but in the tree they are located within the NW Eurasian cluster. The Eskimo-Aleut languages form a clade with the Chukotko-Kamchatkan languages (which could be interpreted as indicative of common descent by proponents of the Nostratic/Eurasiatic macro-family), but this clade is part of a larger clade comprising languages from the North-American Pacific Northwest, including Wakashan and the Salish. While the geographic proximity points to possible contact, there are no good reasons to assume that the involvement of Eskimo-Aleut in this larger clade should be of a genealogical nature. This list of examples could be increased.

On the other hand, several instances of known intense contact in shallow time (up to ca. 5000 BP) are not detectable in the tree. For instance, we observe neither an affinity between the Papuan languages and Austronesian nor between Dravidian and Indo-European or between the SE Asian families and Indic or Mongolic or Tungusic languages. Based on these, admittedly preliminary, considerations we tentatively conclude that language contact is reflected in the world tree mostly if it has been sustained since deep time.

In a few cases — surprisingly few, we dare to say —, the observed patterns can only be interpreted as the result of the accumulation of chance similarities. For instance, Siouan (North America) and Alor-Pantar (Papunesia) are embedded in the SE Asia part of the tree, Ainu (East Asia) in the American part, and Dravidian (South Asia) in (or neighboring to) the Subsaharan African part. Finally, as mentioned earlier, the loci in the tree of two large clusters containing Papuan languages are not meaningful, so they should probably be regarded as random.

To sum up, with some caveats we can say that the four distinct areas identified by the tree represent ancient zones of diffusion and interaction, and a more

tentative hypothesis, subject to further testing, is that there are also deep genealogical relations among some of the languages within these four regions, reaching far beyond conventionally established families. The fact that our data and methods produce clear geographical clusters shows that the deep branchings in the world tree, at least for a large part, are not due to chance. Thus, through a single, consistent, and novel method of comparative linguistics we have obtained a framework for tracing the evolution of language back to minimally four intermediate geo-genealogical aggregates. More trivially, the method also produces fairly accurate results for more recent phylogenetic evolution. Apart from its value as a contribution to historical linguistics, the tree also represents a potentially useful framework for studying cultural evolution at both large and small scales, and the distance measure on which it is based can be employed in cross-disciplinary correlational studies of many different kinds.

Acknowledgements

This work was supported by the ERC Advanced Grant 324246 *EVOLAEMP* (GJ), the ERC Advanced Grant 295918 *MESANDLIN(G)K* and support from the Russian government to further the competitive development of Kazan Federal University (SW), as well as the DFG-Center for Advanced Studies in the Humanities 2237 *Words, Bones, Genes, Tools* (both authors), which is gratefully acknowledged.

References

- Criscuolo, A., Berry, V., Douzery, E. J. P., & Gascuel, O. (2006). SDM: a fast distance-based approach for (super) tree building in phylogenomics. *Systematic Biology*, 55(5), 740-755.
- Gascuel, O. (2000). Data model and classification by trees: the minimum variance reduction (MVR) method. *Journal of Classification*, 17(1), 67-99.
- Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2015). *Glottolog 2.5*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (available online at <http://glottolog.org>, Accessed on 2015-09-18.)
- Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (2008). *The World Atlas of Language Structures online*. Max Planck Digital Library, Munich. (<http://wals.info>)
- Holman, E. W., Wichmann, S., Brown, C. H., & Eff, E. A. (2015). Inheritance and diffusion of language and culture: A comparative perspective. *Social Evolution & History*, 14(1), 49-64.
- Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A., & Bakker, D. (2008). Explorations in automated language classification. *Folia Linguistica*, 42(2), 331-354.
- Isphording, I. E., & Otten, S. (2011). Linguistic distance and the language fluency of immigrants. *Ruhr Economic Papers*, 274.

- Jäger, G. (2013). Phylogenetic inference from word lists using weighted alignment with empirically determined weights. *Language Dynamics and Change*, 3(2), 245-291.
- Jäger, G. (2015). Support for linguistic macrofamilies from weighted sequence alignment. *Proceedings of the National Academy of Sciences*, 112(41). (doi: 10.1073/pnas.1500331112)
- Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.). (2015). *Ethnologue: Languages of the world* (Eighteenth ed.). Dallas, Texas: SIL International. (Online version: <http://www.ethnologue.com>)
- Mace, R., & Holden, C. J. (2005). A phylogenetic approach to cultural evolution. *Trends in Ecology and Evolution*, 20(3), 116-121.
- Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, 93, 351-363.
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289-290.
- Pompei, S., Loreto, V., & Tria, F. (2011). On the accuracy of language trees. *PLoS ONE*, 6(6), e20109.
- Vajda, E. (2010). A Siberian link with Na-Dene languages. *Archaeological papers of the University of Alaska*, 5, 75-156.
- Walker, R. S., Wichmann, S., Mailund, T., & Atkisson, C. J. (2014). Cultural phylogenetics of the Tupi language family in Lowland South America. *PLoS ONE*, 7(4), e35025.
- Wichmann, S., Holman, E. W., Bakker, D., & Brown, C. H. (2010). Evaluating linguistic distance measures. *Physica A: Statistical Mechanics and its Applications*, 389(17), 3632-3639. (doi:10.1016/j.physa.2010.05.011)
- Wichmann, S., Müller, A., Wett, A., Velupillai, V., Bischoffberger, J., Brown, C. H., Holman, E. W., Sauppe, S., Molochieva, Z., Brown, P., Hammarström, H., Belyaev, O., List, J.-M., Bakker, D., Egorov, D., Urban, M., Mailhammer, R., Carrizo, A., Dryer, M. S., Korovina, E., Beck, D., Geyer, H., Epps, P., Grant, A., & Valenzuela, P. (2013). *The ASJP Database (version 16)*. <http://asjp.clld.org>.

PROTOLANGUAGE POSSIBILITIES IN A CONSTRUCTION GRAMMAR FRAMEWORK

SVERKER JOHANSSON

*Dalarna University,
Falun, Sweden
sja@du.se*

Identifying possible stages of protolanguage critically depends on the underlying nature of language. Theories of language differ in evolvability, and in whether they permit protolanguage stages. In this presentation, I will study the protolanguage potential and evolvability of Construction Grammar. Postulating that CG is a biologically real description of language, its evolvability through a sequence of intermediate protolanguages is investigated.

1. Introduction

Protolanguages are postulated intermediate stages in language evolution. Unless one postulates a saltational origin for language, against which there are solid arguments (e.g. Tallerman 2014), one or more intermediate protolanguage stages are required. In order to be evolvable in a Darwinian framework, a protolanguage must be an improvement on earlier communication systems^a, and in order to say something interesting about language evolution it must lack enough language features to be distinct from full human language. Identifying possible protolanguage stages in language evolution becomes an exercise in decomposing language into separable features that can be added sequentially (cf. Johansson 2006, Jackendoff & Wittenberg 2014). This exercise critically depends on the underlying nature of language, on which there is little consensus among linguists (Jackendoff 2010). Language theories vary widely in evolvability, with e.g. Minimalism (Chomsky 1995) being effectively unevolvable (Kinsella 2009), rendering it irrelevant in the context of language evolution.

^a It is assumed here, as argued by e.g. Johansson (2014), that language evolved for communicative purposes.

2. Construction Grammar

I will here investigate the protolanguage potential and evolvability of Construction Grammar (e.g. Goldberg 2006, Langacker 2005). Construction Grammar is a general label for theories of language in which knowledge of language is based on knowledge of *constructions*, form-function pairings that can be more abstract and general language templates than the basic form-function pairings commonly found in the lexicon of most theories of language. Also syntactic knowledge is carried by the constructions, thus removing the need for a separate set of syntactic rules.

Various flavors of Construction Grammar exist, that are structurally similar as grammars, but differ in their underlying assumptions about the nature and structure of our knowledge of language (Langacker 2005). One notable difference is whether it is postulated that grammatical categories exist independent of (and prior to) constructions, which has some impact on evolvability. Some examples of Construction Grammar include Radical Construction Grammar (Croft 2009), Usage-based Construction Grammar (Tomasello 2003), and Fluid Construction Grammar (Steels 2011). For now, I will leave aside these differences, as my primary goal here is to investigate whether Construction Grammar as a generic framework can be suitable for the study of language evolution.

Construction Grammar is used in evolutionary linguistics mainly for modelling the cultural evolution of grammar, with Fluid Construction Grammar extensively employed as it is, by design, amenable to computational modelling of modifiable grammars (e.g. Steels 2011). Biological evolution of language in a Construction Grammar context has received less attention, with Hurford (2012) a notable exception.

Postulating that Construction Grammar is a biologically real description of language, its evolvability through a sequence of intermediate protolanguages can be investigated. By a description being “biologically real”, I mean that its theoretical concepts have a direct correspondance to structures and neural events in the language faculty of people.^b

^b It is perhaps more common to use “psychologically real” in this sense. But in the context of the biological evolution of the language faculty, I think “biologically” is more appropriate.

In a generic Construction Grammar, constructions can be of several different types. Goldberg (2009) provides some examples: “partially or fully filled words, idioms, and general linguistic patterns”, all the way up to the “basic sentence patterns of a language” (p . 94). The cognitive and computational requirements of these different types of constructions vary considerably, and they can be ordered in a hierarchy of increasingly demanding types:

1. The simplest constructions are those that are nothing but a static one-to-one mapping between form and meaning, with no empty slots or other complications. Ordinary words belong here, as do filled idioms. These constructions are basically the same as the lexicon entries in many other theories of language.
2. Constructions with gaps, to be filled with words (unspecified, or pragmatically/semantically specified). Partial idioms belong here, and likewise some morphology.
3. Constructions that are categorized into constituent classes, and with gaps labelled with the constituent class(es) that may fill it. Whether this is really separate from the preceding point depends in part on the status grammatical categories are assigned, which differs between Construction Grammar varieties.
4. Generalizations across similar constructions, leading to a hierarchy of constructions, with inheritance of features between levels in the hierarchy.
5. Fully generalized abstract constructions, with no filled slots (passive, ditransitive, etc.).

Computationally, a central point is whether (and which) constructions allow recursive filling of gaps, so that a hierarchical sentence structure can be built. Type 1 and 2 above typically don't allow this, whereas type 3 and upwards do.

3. Evolving a Construction Grammar

A Construction Grammar does not necessarily need all the types of constructions listed above. This opens up the possibility of evolving a Construction Grammar-based language faculty in several steps, making one type after another available. The order is constrained to be the one in the list above, as each type is a subset of the next type in the list.

Whether this is a possible evolutionary sequence for the human language faculty depends on whether each step corresponds to a viable protolanguage.

In the list below I go through the various steps (with recursive gap-filling added as a step of its own), and consider protolanguage possibilities during each step.

1. Extensible lexicon with static one-to-one mappings between form and meaning, with no empty slots. Already constructions at this level enables a range of possible protolanguages, roughly corresponding to the first group of protolanguages in Jackendoff & Wittenberg (2014).
 - a. One-word stage.
 - b. Pragmatically juxtaposed words with no grammatical connection, similar to the protolanguage of Bickerton (2014).
 - c. Holistic learning of common juxtapositions – filled idioms.
 - i. Links from juxtaposition entry to its components?
 - d. A holophrastic protolanguage would also fit here, but for other reasons I do not regard this as a likely option (cf. Tallerman 2009).
2. Constructions with gaps.
 - a. Generalizing over juxtapositions with one component in common would naturally lead to constructions with gaps – partially filled idioms. Similar to pivot grammar in language ontogeny. Permits a more memory-efficient way of storing a given repertoire of expressions, and likely facilitates language acquisition as well.
3. Categorized constructions and gaps.
 - a. Similar to previous step, but grammar partially replaces pragmatics. Computational improvement.
4. Recursive gap-filling.
 - a. Sentences with hierarchical structure. Opens up new expressive possibilities with more complex sentences that can still be parsed.
5. Inheritance hierarchy.
 - a. Similar to previous step, but more open-ended and powerful. Computational improvement.
6. Fully generalized abstract constructions, with no filled slots (passive, ditransitive, etc.).
 - a. Full modern language.

The details of the sequence are speculative, but it is clear that Construction Grammar naturally provides numerous possibilities for protolanguages. Most of the steps are likely to be evolutionarily small and arguably provide language improvements, either communicatively or computationally, that may be adaptively relevant.

The second step, going from “constructions” that are filled and static, to constructions with gaps, may be non-trivial, and is likely to be a key evolutionary step. Non-human primates do have some general capacity for generalizations (Hurford 2007), but even language-trained apes show little if any evidence of the type of linguistic generalizations required here, despite appropriate patterned input.

The Construction Grammar itself does not specify the language even at the abstract level of Chomskyan UG. Within the broad limits of what can be stored as constructions of different types, at different stages, and what can be handled computationally in construction parsing, languages may evolve culturally quite freely. This provides scope for biological/cultural coevolution, and grammaticalization processes (cf. Heine & Kuteva 2007) come naturally at the later stages.

In summary, Construction Grammar appears to be a very promising framework for further evolutionary studies.

References

- Bickerton, Derek (2014). *More than nature needs*. Cambridge, MA: Harvard University Press
- Chomsky, Noam. (1995). *The Minimalist Program*. Cambridge, Mass.: The MIT Press.
- Croft, William (2009) *Radical construction grammar*. Oxford: Oxford University Press.
- Goldberg, Adele E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele E. (2009). The Nature of Generalization in Language. [Target Article] *Cognitive Linguistics* 20 1: 93-127
- Heine, Bernd & Kuteva, Tania (2007). *The genesis of grammar: a reconstruction*. Oxford: Oxford University Press
- Hurford, James R. (2007). *The origins of meaning*. Oxford: Oxford University Press
- Hurford, James R. (2012). *The origins of grammar*. Oxford: Oxford University Press
- Jackendoff, Ray. (2010). Your theory of language evolution depends on your theory of language. In Richard Larson, Viviane Déprez & Hiroko Yamakido (eds.), *The Evolution of Human Language: Bilingual Perspectives*, 63–72. Cambridge: Cambridge University Press

- Jackendoff R, Wittenberg E. (2014). What can you say without syntax? A hierarchy of grammatical complexity. In *Measuring Grammatical Complexity*, ed. FJ Newmeyer, LB Preston, pp. Oxford: Oxford Univ. Press, pp. 65-82
- Johansson, Sverker. (2006). Working backwards from modern language to proto-grammar. In Cangelosi et al. (Eds.). *Evolution of Language — Proceedings of Evolang6*. Singapore: World Scientific Publishing.
- Johansson, Sverker. (2014). Did language evolve incommunicado. In Cartmill et al. (Eds.). *Evolution of Language — Proceedings of Evolang10*. Singapore: World Scientific Publishing.
- Kinsella, Anna R. (2009). *Language Evolution and Syntactic Theory*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. (2005) Construction Grammars: Cognitive, radical, and less so. In M. Sandra Peña Cervel & Francisco J. Ruiz de Mendoza Ibáñez (Eds) *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*. Berlin: Mouton de Gruyter.
- Steels, Luc (ed.) (2011). *Design Patterns in Fluid Construction Grammar*. John Benjamins, Amsterdam.
- Tallerman Maggie. (2009) Holophrastic protolanguage: Planning, processing, storage, and retrieval. *Interaction Studies* 9(1), 84-99.
- Tallerman Maggie. (2014) No syntax saltation in language evolution. *Language Sciences* 46(Part B), 207-219.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Harvard University Press.

EVOLUTION OF THE LANGUAGE-READY BRAIN: WARFARE OR ‘MOTHER TONGUES’?

CHRIS KNIGHT,
DEPT OF ANTHROPOLOGY, UNIVERSITY COLLEGE LONDON
chris.knight@live.com

CAMILLA POWER,
ANTHROPOLOGY PROGRAMME, UNIVERSITY OF EAST LONDON
c.c.power@uel.ac.uk

For language to evolve, group-level normativity, cooperation and mutual understanding must have intensified beyond the range of variation permitted by non-human primate social life. Although this general assumption is broadly shared, recent theoretical models to explain the necessary group-level cooperation have clustered around two poles. At one extreme, theorists have traditionally invoked inter-group conflict including warfare. At the other, scholars have invoked grandmothering and coalitionary alliances between females to share the burdens of childcare. These competing approaches make divergent predictions testable in the light of recently available evidence from population genetics.

1. Introduction: two competing models

This paper is not directly about the evolutionary emergence of language, but follows a strand of thinking in EVOLANG over two decades addressing social factors constraining the process. Precedents in this respect include Terrence Deacon's (1997) concept of pair-bonding through ritual contracts; Jean-Louis Dessalles' (1998) focus on relevance in connection with hominin politics; Robin Dunbar's (2003) 'gossip and grooming' hypothesis; Tecumseh Fitch's ideas about kin selection and mother tongues (2004) and Klaus Zuberbühler's (2012) suggestion of a link between human infant babbling and cooperative breeding.

Behind these debates, evolutionary psychologists Michael Tomasello and Sarah Hrdy have been influential background figures. In her *Mothers and Others* (2009), Hrdy draws on Tomasello's careful work differentiating great ape mind-reading capacities from the joint attention, egocentric perspective reversal and intersubjectivity characteristic of human social cognition from an early age. But whereas Hrdy views alloparenting as decisive in generating mutual understanding and a language-ready brain, Tomasello et al. (2012) give priority to (1) foraging and (2) intergroup conflict or warfare. In their two-stage model for the emergence of language, these authors marginalise Hrdy's cooperative

breeding scenario as relevant only their to Stage One – foraging imperatives giving rise to shared intentionality with rudimentary symbols. Neither foraging nor reproductive imperatives, they say, suffice to explain the crucial progression to Stage Two – the emergence of community-wide norms with collective intentionality. For syntactically complex language to evolve, according to these authors (2012: 674), ‘the entire social group needed to work together interdependently in order to compete with other social groups, leading to such things as group-created conventions, norms, and institutions.’ The only factor powerful enough to generate the required interdependence, mutual understanding and within-group solidarity, they insist, would have been group selection driven by territorial conflict and associated external threat.

This divergence between Hrdy and Tomasello fits into a broader picture of two theoretical camps, one stressing childcare, the other stressing male alliances for hunting, warfare and/or defence. While EVOLANG has had a historic tendency to go against the grain in focusing on childcare and reproductive strategies (e.g. Fitch 2004; Power 2014), among evolutionary theorists this has been a minority position.

2. The warfare model

With few exceptions, primatologists speculating on human origins have argued that because humans are great apes, it follows that hunter-gatherers, like their great ape relatives, must always have been male-philopatric. Palaeolithic archaeologists and evolutionary theorists have frequently followed this tendency among primatologists to emphasise the centrality of male kin-bonding to key developments in human evolution. Linked with this have been assumptions about paternity certainty, fights over women, leadership hierarchies, aggression, territorialism – and warfare between neighbouring bands. Key figures here (broadly converging but without necessarily agreeing on all points) have been Alexander (1987), Bowles (2009), Chapais (2008), Foley and Gamble (2009), Gavrilets (2012) and Wrangham and Peterson (1997).

3. Mothers and others

According to the strikingly different (and arguably more gender-aware) perspective of Hrdy (2009; see also Burkart et al. 2014), the transition from the psychology of bipedal apes to ‘emotional modernity’ in genus *Homo* involved significantly raised reproductive costs which could be met only by sharing burdens, progressively collectivising childcare. Mothers enlisted the help of their mothers and also their own children together with sisters and other relatives (including eventually males) in a process which profoundly transformed all

aspects of human psychology, social structure and communication. Mothers, their babies and other carers would interact, probing and testing to see whether sufficient trust could be established between them, these interactions fostering the development of two-way mind-reading, joint attention and intersubjectivity. The evolution of our distinctively human ‘cooperative eyes’ and language-ready brain begins here.

The alloparenting model provides supportive context for Fitch’s ‘mother tongues’ hypothesis, particularly where pathways of allocare by juveniles or older siblings are stressed (Kramer and Otárola-Castillo 2015). Mothers and their older offspring come under selection pressure for more intense cooperation over longer periods of life-history, leading to increasingly stable extended sibling networks. If we take it that in the period before farming, males tended to be nomadic rather than attached to place, a picture emerges in which, having found a bride, a young man might temporarily visit her among affines while regularly returning ‘home’, movement between nearby camps enabling each to retain childhood links with mother, sisters and other natal kin. Where this pattern prevails, bonds of trust between siblings of opposite sex may be retained and regularly serviced throughout adult life, enhancing the stability of kin-based strategies of co-operative communication along lines consistent with Fitch’s (2004) ‘mother tongues’ language evolution model.

4. New evidence from genetics

Ground-breaking new research by population geneticists may help us to decide between these competing models. The work depends on the familiar fact that mitochondrial DNA is transmitted through females, Y chromosome DNA through males. Matrilocality, where traditionally practiced over successive generations, yields a tight local clustering of mtDNA lineages while Y chromosome sequences are dispersed. Conversely, patrilocality yields localised Y chromosome lineages while dispersing corresponding sequences of mtDNA. The significant finding is that the former pattern characterises African hunter-gatherers, the latter pastoralists and farmers (see references below). This constitutes suggestive evidence that over the millennia, females in the major hunter-gatherer regions of sub-Saharan Africa have opted to reside postmaritally with maternal kin, male kin moving from their birthplace to reside with in-laws. With the transition to farming the reverse pattern comes to prevail: matrilocality is overridden and patrilocality becomes the norm.

To scholars interested in narrowing down the specific social conditions likely to have been responsible for the emergence of language, these findings are significant. We now have several robust studies by independent scholars

cumulatively confirming an African hunter-gatherer matrilocal bias. For Central African (Pygmy) populations, see Verdu and Austerlitz (2015) and also Destro-Bisol et al. (2004). For Southern African (Khoisan) hunter-gatherers, see Schlebusch (2010). Wood et al. (2005) suggest that the pattern noted is pan-African. An earlier global study by Hammer et al. (2001) suggests that the pattern may apply cross-culturally.

Although these findings are in one sense new, they confirm on the basis of twenty-first century methods a pattern long familiar to anthropologists working among immediate return hunter-gatherers. Among the Hadza of Tanzania, it is unusual for an adult woman to be living apart from her mother during the early years of her married life. Woodburn (1968), Blurton-Jones et al. (2005) and Wood and Marlowe (2011) all concur in stating that postmarital residence is flexible but with a matrilocal bias. Much the same applies to the once-numerous southern African Khoisan hunter-gatherer populations (Lee 1979; Marshall 1959, 1976). Studies by hunter-gatherer specialists across the world indicate that the traditional concept of a patrilocal band, once a staple of social evolutionary theorising, has in fact no applicability to the nomadic hunter-gatherer lifestyle (Alvarez 2004; Marlowe 2004).

The new findings from African hunter-gatherer population genetics suggest that the matrilocal bias found among extant populations is not an outlier or product of recent marginalisation but represents a default for humans across the continent prior to the emergence of farming. Residence among these assertively egalitarian immediate-return hunter-gatherers is almost always matrilocal during the crucial early years of a first child's life, when the mother is most in need of help. Even where a woman subsequently chooses to reside patrilocally, she is likely to retain close links with mother, either because distances are not great or because an older woman will choose to live with a child-burdened daughter in order to help in a grandmothing role.

5. What is the relevance to EVOLANG?

All this has implications for the study of human social and cognitive evolution and therefore for the origins of language. Although the time-depth of the genetic studies remains relatively shallow – reaching back at most a few thousand years – the results cast doubt on the assumption that the intense cooperation necessary for the evolution of language must have emerged as an adaptation for warfare. We may agree with Wrangham and others that territorial conflict is practised by chimpanzees, frequently in the form of a male gang from one group cornering and severely assaulting a lone individual from another. But since these primate relatives of ours lack either group-level normativity or language, we in

EVOLANG surely need to assume not primatological continuity but the emergence of something new. In light of the new genetic evidence, a shift from male to female philopatry (in anthropological terms, from patrilocal to broadly matriloca residence) must have been part of the novel package. The picture shifts from territorialism and warfare to something much closer to what is actually found among extant immediate-return hunter-gatherers – high levels of trust between neighbouring bands, in-laws constantly visiting one another across permeable or non-existent boundaries, substantial gender equality, an absence of leaders – and community relations in which playful, good-humoured forms of social pressure limit the disruptive consequences of interpersonal conflict.

These findings help explain why a language-ready brain evolved among our own ancestors but in no other lineage of great apes. Male philopatry means that a female chimpanzee must leave her natal group on reaching sexual maturity and move to where she has no female relatives to support her. Infanticide risks being high, no chimpanzee mother would dare to leave her baby alone with another carer – she would fear for its safety. This restriction actively hinders mothers from producing offspring with large, slow-maturing and correspondingly costly brains. Note also that it is not possible for an older female to support her adult daughter's reproduction if that daughter has been forced to relocate and raise her offspring in a neighbouring group's territory. The greater the geographical distance, the harder it is for a grandmother to forage for her daughter or protect her grandchildren from hunger or harassment. Because great apes are condemned to being single mothers, selection pressures have in their case discriminated against burdensome offspring with larger brains; in this sense, male philopatry imposes what has been termed a 'gray ceiling' (Isler and van Schaik 2012) – preventing brain size from exceeding a certain limit.

Clearly, evolving human mothers found a way to break through those great ape encephalisation constraints. Brain size in genus *Homo* reached double that of previous hominins early in the Pleistocene, smashing that grey ceiling – indicating that cooperative breeding was already underway with *H. erectus*. This laid the foundation for the subsequent tripling of brain size in *H. heidelbergensis* and descendants. The secret of producing and caring for burdensome offspring with increasingly large (and by implication 'language-ready') brains was for mothers and daughters to stay together, if not always through strict matriloca residence, at least through sufficient residential flexibility for a mother to move to wherever her daughter chose to live.

6. Brideservice as the alternative to warfare

Wherever hunter-gatherers still follow traditional ways, brideservice is the fundamental economic institution (Collier and Rosaldo 1981). Brideservice means that males do not attempt to kill rivals in neighbouring bands in order to carry off females. Instead, the strategy which proves evolutionarily stable is one in which groups of related males successfully defend their female kin, meanwhile compelling neighbouring males to earn any ongoing marital rights by visiting periodically and bringing home provisions. 'Women expect meat from lovers', as Collier and Rosaldo (1981: 314) put it, referring to 'brideservice societies' throughout the world. Groups of brothers, while ready at all times to return home to defend sisters, look elsewhere to supply meat to and thereby earn ongoing marital rights with unrelated brides in nearby locations where they are visitors and must therefore adopt correspondingly reversed roles.

Effective brideservice correlates broadly with matrilocal residence because the more distant a daughter is from her natal home, the more difficult it becomes for her kin to demand continuing services from their son-in-law. When a bride resides with the family of her spouse, her kin lack the control they might exert where a son-in-law is obliged to keep visiting them, providing game meat on pain of losing his sexual rights. In life-history terms, the longer the period of initial matrilocal residence, the greater the amount of brideservice which can be extracted and so the better-provisioned the mother and her baby. The connection with the evolution of large brains should by now be clear.

All this invites us to envisage an evolutionary scenario radically at odds with the currently favoured 'primitive warfare' model. Endlessly unresolved male-on-male conflict could not conceivably have increased provisioning support or safety for mothers and their vulnerable offspring. A dynamic in which male solidarity falls apart in the absence of external warfare provides no basis for the cultural 'ratchet effect' (Tomasello et al. 1993). Where warfare is endemic, cumulative cultural evolution is threatened because it depends on continuous victory and no foraging group can expect to win all the time. In this kind of situation, conflicts between neighbouring groups of males can find no resolution because neither side can afford to lose. It is quite different with gender conflict, because here, female coalitionary action can raise the costs of attempting dominance to the point where males have a genetic interest in 'losing' to females – that is, to successful mothers who are likely to be co-operatively nurturing their own and one another's offspring. This yields gender solidarity through essentially playful conflict between opposed gender camps – the pattern still characteristic of African immediate-return hunters and gatherers.

Gender strategies and issues of trust are all too often absent from current evolutionary models. By focusing on them when discussing the origins of language, we can begin to understand more precisely just how and why it was that large-brained *Homo sapiens* succeeded in making the breakthrough to an egalitarian political ethos of the kind necessary for linguistic cognition and communication to evolve.

6.1. References

- Alexander, R. (1987). *The Biology of Moral Systems*. Hawthorne, N.Y.: Aldine.
- Alvarez, H. (2004). Residence groups among hunter-gatherers: a view of the claims and evidence for patrilocal bands. In B. Chapais & C. Berman (Eds.) *Kinship and Behavior in Primates*. New York: OUP, pp. 420-442.
- Blurton-Jones, N., Hawkes, K., O'Connell, J. F. (2005). Older Hadza men and women as helpers: residence data. In: Hewlett, B.S., Lamb, M.E. (Eds.) *Hunter-gatherer childhoods*. New Brunswick, NJ: Aldine, pp. 214-236.
- Bowles, S. (2009). Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science*, Vol. 324 (5) June 1293-1298.
- Burkart, J. M., Allon, O., Amici, F. et al., (2014). The evolutionary origin of human hyper-cooperation. *Nat. Commun.* 5, 4747.
- Chapais, B. 2008. *Primeval Kinship: How Pair Bonding Gave Birth to Human Society*. Cambridge, MA: Harvard University Press.
- Collier, J. F. & M. Z. Rosaldo (1981). Politics and gender in simple societies. In S. B. Ortner and H. Whitehead (Eds.), *Sexual Meanings. The cultural construction of gender and sexuality* (pp. 275-32). Cambridge: CUP.
- Deacon, T. (1997). *The symbolic species: The co-evolution of language and the human brain*. London: Penguin.
- Dessalles, J.-L. (2000). Language and hominid politics. In C. Knight, M. Studdert-Kennedy & J. R. Hurford (Eds.), *The Evolutionary emergence of language* (pp. 62-80). Cambridge: CUP.
- Destro-Bisol G., Donati F., Coia V., Boschi I. et al., (2004). 'Variation of female and male lineages in sub-saharan populations: the importance of sociocultural factors.' *Mol. Biol. Evol.* 21: 1673-82.
- Dunbar, R. I. M. (1996). *Grooming, gossip and the evolution of language*. London: Faber & Faber.
- Fitch, W. T. (2004). Kin Selection and 'Mother Tongues': A Neglected Component in Language Evolution. In D. Kimbrough Oller & Ulrike Griebel (Eds.), *Evolution of Communication Systems: A Comparative Approach*, pp. 275-296. Cambridge, MA: MIT Press.
- Foley, R. & C. Gamble (2009). The ecology of social transitions in human evolution. *Philos Trans R Soc Lond B Biol Sci*. Nov 12, 2009; 364(1533): 3267-3279. doi: 10.1098/rstb.2009.0136.
- Gavrilets, S. (2012). Human origins and the transition from promiscuity to pair-bonding. *PNAS* June 19, vol. 109, no. 25, pp. 9923-9928.
- Hammer M.F., Karafet T.M., Redd A.J. et al. (2001). 'Hierarchical patterns of global human Y-chromosome diversity.' *Mol. Biol. Evol.* 18: 1189-203.
- Hrdy, S. B. (2009). *Mothers and others. The evolutionary origins of mutual understanding*. Cambridge, MA: Harvard University Press.

- Isler, K. & van Schaik, C.P. (2012). How our ancestors broke through the gray ceiling: Comparative evidence for cooperative breeding in early *Homo*. *Current Anthropology* 53, S6, Human biology and the origins of *Homo* (December): S453-S465
- Keeley, L. (1996). *War before civilization: the myth of the peaceful savage*. Oxford: OUP.
- Kramer, K. L. & E. Otárola-Castillo (2015). When mothers need others. Life history transitions associated with the evolution of cooperative breeding. *Journal of Human Evolution* 84:16-24.
- Lee, R. B. (1979). *The !Kung San. Men, women and work in a foraging society* (pp. 24-42). Cambridge: CUP.
- Marlowe, F. (2004). Marital residence among foragers. *Current Anthropology* 45: 277-283.
- Marshall, L. (1959). Marriage among !Kung Bushmen. *Africa* 29:335-365.
- Marshall, L. (1976). *The !Kung of Nyae Nyae*. Cambridge, MA: Harvard UP.
- Power, C. (2014). Female philopatry and egalitarianism as conditions for the emergence of intersubjectivity. In E. A. Cartmill, S. Roberts, H. Lyn and H. Cornish (Eds.), *The Evolution of Language. Proceedings of EVOLANG X*. World Scientific, pp. 252-259.
- Schlebusch, C. M. (2010). 'Genetic variation in Khoisan-speaking populations from southern Africa.' Dissertation, University of Witwatersrand.
- Tomasello, M., Kruger, A. & Ratner, H. 1993. Cultural learning. *Behavioural and Brain Sciences* 16 495-552.
- Tomasello, M., A. P. Melis, C. Tennie, E. Wyman and E. Herrman (2012). Two key steps in the evolution of human cooperation : The interdependence hypothesis. *Current Anthropology* 53(6): 673-692.
- Verdu, P. & F. Austerlitz, (2015). 'Post marital residence behaviours shape genetic variation in hunter-gatherer and agricultural populations from Central Africa.' *Hunter Gatherer Research*, 1:1.
- Wood, B. & F. Marlowe (2011) Dynamics of postmarital residence among the Hadza. *Human Nature* 22: 128-138.
- Wood, E.T., Stover D.A., Ehret C. et al., (2005). Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.* 13: 867-76.
- Woodburn, J. (1968) Stability and flexibility in Hadza residential groups. In: Lee, R.B. and I. DeVore, (Eds.) *Man the Hunter* (pp. 103-110). Chicago: Aldine.
- Wrangham, R. & D. Peterson (1997). *Demonic Males: Apes and the Origins of Human Violence*. New York: Houghton Mifflin.
- Zuberbühler, K. (2012). Cooperative breeding and the evolution of vocal flexibility. In Tallerman, M. & K. Gibson (Eds.), *The Oxford handbook of language evolution* (pp. 7-81). Oxford: OUP.

THE EMERGENCE OF ARGUMENT MARKING

SANDER LESTRADE

*Centre for Language Studies, Radboud University
Nijmegen, The Netherlands
s.lestrade@let.ru.nl*

The emergence of grammatical role marking and person indexing is modeled in a cognitively motivated, multi-agent computer simulation of language change. As the forms of frequently used words erode and their meanings desemanticize, they develop into maximally short forms with maximally general meanings, which eventually can no longer be used as referring expressions. Using an artificial language that initially does not have any grammatical argument-marking strategy whatsoever, it can be shown how lexical ad hoc solutions for event-role ambiguity develop into case marking, while referring expressions develop into verb indexes.

1. Introduction

Consider example (1) from Turkish. As in many other languages (cf. Siewierska & Bakker, 2009), the verb indexes features of its external argument and the internal argument is marked for its role by a case marker (by means of *-um* and *-i*, respectively).^a Unless word order were used, without these markers it would be unclear what the argument structure of the sentence is (i.e., who is loving whom).

- (1) *Ben sen-i seviyor-um*
I you-ACC love.PROG-1SG
'I love you.' Turkish (Derya Demircay, p.c.)

In this paper it will be shown how such argument marking can result from cultural evolution (Deacon, 1997; Smith & Kirby, 2008; Christiansen & Chater, 2008). That is, it is not part of whatever constitutes our inherent language capacity (cf. Pinker & Bloom, 1990; Müller, 2002), but emerges in language as a product of our behavior instead. Rather than reconstructing the developmental history of natural-language markers (cf. e.g. Heine & Kuteva, 2002, 2007), the grammaticalization process will be studied in a computer simulation of artificial-language change. The main hypothesis is that grammaticalization does not merely explain the development and renewal of constructions within an existing system, but can also account for the emergence of the argument-marking system *itself*.

^a Abbreviations: *1, 2, 3* first, second, third person, *A* actor, *ACC* accusative, *PROG* progressive, *SG* singular, *U* undergoer, *V* verb.

Two relevant studies that should be mentioned in this context are van Trijp (2012) and Beuls and Steels (2013). Both studies simulate the development of grammar in terms of cultural evolution and come to very similar conclusions. For reasons of space, only the important differences will be mentioned here. First, both of these studies are implemented in *fluid construction grammar* (Steels, 2011). This is comforting, in fact, as the different software implementations merely corroborate the mutual findings. With respect to content, Beuls and Steels (2013) are concerned with internal agreement, i.e. the marking of lexical items as belonging to the *same* constituent. The present study, instead, is concerned with marking the relationship *between* constituents (by means of case marking or *external* agreement; cf. Lehmann 1988 for discussion). Here too it will be important to combine elements into constituents, but for this a grouping principle is assumed that stays active throughout all generations (cf. Section 2.5). Van Trijp (2012) does simulate the evolution of argument marking. However, the grammaticalization process that is fundamental to the present proposal is simplified in his study. For example, whereas he assumes a *case strategy* from which eventually a conventional *case system* develops, this strategy is not provided in advance here. Instead, the recruitment and development of markers are an important part of the simulation.

In Section 2, the model will be introduced. Section 3 very briefly discusses two important mechanisms of grammaticalization and the way in which these are implemented. Section 4 shows the results, which are discussed in Section 5.

2. Modeling event communication

WDWTW (for *who does what to whom*) is a cognitively motivated multi-agent model developed by the author in which event communication and the emergence of grammar can be simulated.^b A population of agents starts out with a shared lexicon of referring expressions only and a “language-ready” brain, which basically means that the agents have a desire for communicative success and are capable of joint attention (Tomasello, 2003; Arbib, 2015).

2.1. General procedure

Agents communicate about automatically generated events in their virtual world. The speaker has to find an adequate wording for a target event that is sufficiently distinctive given the situational context in which other events are ongoing too (i.e., if there are similar distractor objects, referential expressions have to be more specific). This crucially involves making clear the distribution of predicate roles over the event participants in the communicated event (e.g., if there’s a hitting

^bA user-friendly version is still being developed and will be distributed via the CRAN archive (R Core Team, 2014). In the mean-time, the codes are available from the author upon request.

event, who does the hitting and who is being hit). If the hearer correctly identifies the event the speaker is talking about, the agents mark the successful usage of the words that constitute the utterance, remember the exact meaning for which the words were used, and next either switch turns to go on with their conversation or end it, after which two new agents are randomly selected for a new conversation (cf. Steels' 1997 *language games*).

After some time, agents procreate, at which point their offspring inherit their lexicon with minor modifications to the meanings of those words that have not been used until then, and without the usage history. This they develop themselves, partly by interacting with their parents, which die a bit later. As the development and maintenance of a conventional lexicon are not of primary concern here and have been successfully modeled elsewhere moreover (cf. e.g. Hurford, 1989; Hutchins & Hazlehurst, 1995; Steels, 1997, and Kirby, 2000), the present simplifications seem warranted.

2.2. Representation of meaning

The agents of WDWTW live in a very abstract virtual world to which none of our concepts apply. For example 'hitting', which was used for illustration above, does not mean anything to them. Still, it is possible to provide the agents with a mental lexicon that is similarly organized as ours, for present purposes at least.

According to Wierzbicka (1996), all natural-language concepts can be decomposed into meaning primitives such as CONCRETE, HUMAN, MALE, etc. (cf. e.g. also Guiraud, 1968). Similarly, in a way, Gärdenfors (2000) argues that concepts are sets of values on different meaning dimensions. Thus, we can think of a cat as something that is time-stable, concrete, alive, four-legged, tailed, etc. Abstracting away from the quality of the dimensions that organize our mental lexicon, concepts can be modeled as vectors specifying values on a number of numerical meaning dimensions: For example, instead of our animacy dimension with the values animate and inanimate, agents may have a binary dimension with values 0 and 1. By default, the object lexicon of the agents is organized along nine such abstract dimensions, the first five of which are binary, the next four make a nine-way distinction (but this and virtually all other settings of the model can be manipulated). The increasing distinctiveness is loosely motivated by the fact that nodes higher-up in a taxonomy bisect the world in major types (e.g. actions vs things, concrete vs abstract; i.e. mostly the distinctions grammar cares about), whereas nodes further down use a more fine-grained classification (e.g., types of animal).

Verbs are similarly specified, with the addition of one or two *perspectival* roles, viz. the *external* and, in case of a two-place predicate, *internal argument* role.^c These are characterized using vector representations too. And as for nouns, one could think of each meaning dimension as one that is grammatically relevant

^cThe external argument is the participant whose perspective on the event is taken by the corre-

in natural language (\pm instigating, \pm intentional, \pm affected, etc.), although such notions have no meaning in the model. Values on external dimensions are on average higher than values on internal ones to implement prominence preferences known from natural language (subjects are preferably human, volitional, in control, etc.; cf. Dowty, 1991).

The forms and meanings of 999 nouns and 499 verbs are randomly generated. The two nouns that have the highest values across all dimensions are used for first and second person reference.

2.3. Event generation

Agents find themselves in situations in which a differing number of automatically generated events is going on. One of the events is the target event, which is the one the hearer should single out on the basis of the speaker's description. The other events provide the distractor objects and actions.

Events are combinations of two or three vectors: one for the event action, one for the more active event participant (the *actor*), and, in case of a transitive event, one for the less active participant (the *undergoer*; after Van Valin, 1999). The participants that figure in the situation are dependent on the *common ground* of the speech participants, which they develop while they talk. The common ground consists of the things that have been discussed in the current conversation. Initially, it consists of the speech participants themselves and (the meanings of) three randomly selected lexical entries. Other objects have a small probability of entering the scene. Elements from the common ground are assigned a number of external roles (in which the same element may figure in multiple events). Next, on the basis of these elements' propensities and affordances, action predicates are sampled from the mental lexicon (using a real-world example, dogs more often found running than flying). Finally, for each two-place predicate that is selected, a second argument is sampled from the mental lexicon on the basis of its role-qualifications for the internal role. The degree of confirmation to such role expectations is set by model parameter. In the present settings roughly one third of the objects will never be combined with a given role and most are only incidentally so. Only the top 5% of most qualified objects is used with a reasonable frequency.

Note that the events that are thus generated consist of instantiated meaning representations (i.e., the mind determines what happens). This is of course not very realistic, as in the real world, the concepts we have are generalizations of the things we perceive. But we can exploit this connection the other way around with the same net result (adding some random noise to the meanings that constitute the events): Both in reality and in the model, there is considerable overlap in what

sponding verb. For example, the same event in which a book first belongs to John and later to Mary can be conceptualized as *John sold a book to Mary* taking the perspective of John or as *Mary bought a book from John*, with Mary as the external argument.

people/agents think and what they experience.

2.4. Word selection, role marking and person indexing

In order to communicate an event, the speaker matches what it perceives to the meaning representations in its mental lexicon. The match between the referents in the virtual-world and the lexical semantics of the words available in the lexicon of the agent is evaluated by calculating the average (absolute) difference per meaning dimension, and subtracting this from 1, in which dimensions that are not specified are ignored. A score of 1 shows a perfect match, a 0 shows maximal deviation. The referential expressions that the speaker selects need not have a perfect semantic match, they only have to be sufficiently distinctive given the distractor objects in the situational context (cf. Grice, 1975). In principle, however, words do have to agree in person with their referents (e.g., a first-person referent should be referred to with a first-person word).

The order in which words are considered for expression does not only depend on their semantic match. Also semantic specificity and usage frequency play a role (Balota & Chumbley, 1985). More frequent and general words with a comparable match are ranked before less frequent and general ones. The first word that is found sufficiently distinctive given the context is selected for production.

In the absence of grammatical argument-marking strategies, agents initially have to use lexical ad hoc solutions to make clear who does what to whom. If the role distribution follows from the semantics of the ingredients for free, nothing extra needs to be done (e.g. the abstract equivalent of a reading event in which a book and a man are involved). If not, an agent has to add a word specifying the predicate role of one of the participants to make it clear. For example, to say that a man saw a woman it could say *woman man looker see*.

For sake of ease, a correspondence is assumed between the dimensions specifying the nouns and those of the verb roles. Thus, the same vector-comparison method determining the match between a word and its referent can be used to establish the role qualification of an argument (its *typing* score, after Aristar, 1997). If the typing score of a participant for its intended role is not significantly higher than the one of the other participant for this role, the role does not follow automatically. The resulting ambiguity only needs to be resolved at one of the arguments, as the role of the other one follows automatically. As the example with *looker* shows, this initially involves lexical expressions that specify idiosyncratic predicate roles. Once more grammatical expressions develop, these can be used instead (cf. Section 3).

Whereas role marking thus serves a clear communicative goal, the reasons for indexing person features are much less obvious. Although some have claimed an identification function (cf. Lehmann, 1995), possibly with a corresponding processing gain (Hawkins, 2002), it might well be nothing but historical junk (although it may be too stable for that; cf. Dahl, 1995, 269; Collin, 2015). Whatever

its synchronic merits, its presence too can be explained as a result of grammaticalization.

2.5. Grouping, event identification, and representation updating

Adding role markers requires the hearer to correctly group together arguments and their markers (for the speaker could also have used *woman man looker see* to say that someone saw a manly woman). In the model, all possible groupings are fully explored, after which the interpretation with matches best with any of the events in the situation is considered the intended one, and the corresponding event is assumed to be the target event. If the target event is correctly identified, the frequency scores and usage histories of the words in the utterance are updated.

3. Development of argument marking

In the course of time, words may grammaticalize. In natural language, grammaticalization is a gradual, diachronic process in which lexical items become less autonomous in various ways (Hopper & Traugott, 2003; Heine & Kuteva, 2007; Haspelmath, 1998). Two important mechanisms in this process are *erosion* (some forms being pronounced sloppily and eventually becoming represented accordingly) and *desemanticization* (some meanings becoming more general as a function of the different contexts in which they are used). If a meaning becomes more general, it can be used in even more contexts, and if a form becomes too short to stand on its own, it is suffixed to its host (Bybee, 1985). From this, it straightforwardly follows that popular lexical role markers may develop into case markers.

The development of indexing requires an extra ingredient. According to Ariel, referential pronouns for highly accessible referents can get reduced to the extent that they are no longer perceived as referring expressions, because of which more pronounced argument copies have to be added. As local persons are consistently highly accessible (Ariel, 1999, 221), this process mostly involves the speaker and addressee. A third-person example is given in (2). In standard French, the pronouns *il* 'he' and *elle* 'she' are referential expressions that make a gender distinction (e.g., *La jeune fille est venue hier soir. Elle est danseuse.* 'The girl came yesterday evening. She is a dancer. '); in non-standard French, *il* has become an agreement marker that no longer refers nor distinguishes gender.

- (2) *Ma femme il est venu.*
my:F wife AGR is come
'My wife has come.' French (Heine & Kuteva, 2002, 234)

In the model, all lexical items (including those for local persons) initially are fully specified semantically and have equally long forms of expression. Words that are frequent or predictable (because of the context or prior usage) are pronounced sloppily, which is instantiated as going back in the alphabet for the last

letter of a form and deleting it altogether if this is no longer possible. Sloppy pronunciation does not lead to a change of lexical representation for the speaker. But if the hearer is still unsure about the form of a word because it has not used it sufficiently frequently yet, it will adapt its representation on the basis of what it hears (Nettle, 1999). Desemanticization is modeled by progressively removing meaning dimensions along which most variation is attested in the usage history. Deletion takes place only after certain frequency thresholds have been reached. For a first dimension to be removed, a word has to be used in 1% of the utterances, for the final dimension it has to be used in 50%.

4. Results

To test the hypothesis that argument-marking systems develop from grammaticalization, two lineages are tested. In the first, desemanticization and erosion do not apply, in the second they do. Examples of typical utterances after 44 generations are given in (3). As can be seen in (3-a), in the absence of grammaticalization, speakers still use lexical ad hoc solutions (as the glossing of *lusolal* means to reflect) and indexing does not develop. In the lineage in which grammaticalization does apply, agents begin in the same way, as shown in (3-b) (word-order variation is random and can be ignored). Over time, however, utterances very similar to the Turkish example in (1) emerge. The marker *-tu* in (3-c) originated as the noun *tusedul*. Because of erosion, it now has to be suffixed to its host, and because of desemanticization, it lost three of its meaning representations, hence its gloss as a general undergoer marker. Note that the etymological source of the second-person index *-da* in (3-c) is also used in (3-b), where it still is fully specified and of maximal length. After 44 generations, it grammaticalized to the degree that it can no longer refer as a result of which a new word with more expressive power has to be recruited.

- (3) a. *nuriret unudede amoduse lusolal*
 1 unudede.V 2 unudeder
 ‘You unudede me.’ (Lineage 1, 44th generation)
- b. *daniset namimin sulalet inenono*
 2 naminin.V 1 namininee
 ‘You naminin me.’ (Lineage 2, first generation)
- c. *otosa namimin-da su-tu*
 2 naminin.V-2 1-U
 ‘You naminin me.’ (Lineage 2, 44th generation)

5. Discussion: results, implications and limitations

Using a cognitively motivated computer simulation of language change, it was shown how the emergence of grammatical argument marking can be modeled as a result of cultural evolution.

Case markers straightforwardly develop as a result of grammaticalization. As relative frequency plays a role in word activation, lexical items that have previously been used for role disambiguation are more likely to be considered again. As there are maximally only two roles to be kept apart, namely the external and internal one, the role marker often need not be very specific. Thus, a previously used marker is often found good enough, as a result of which its frequency of usage increases further, as well as the variation of its usage contexts. Because of the former, its form is likely to erode; because of the latter, its meaning is likely to bleach. Eventually, as illustrated in (3-c), this may lead to a the model equivalent of a case marker (Lestrade, 2010): a maximally short form with a minimal number of meaning dimensions specified that marks its host for its function or type of dependence.

Although the synchronic functionality of indexing is debated, its development too can easily be modeled, especially for local persons. Recall that speech participants are part of the common ground by default and therefore figure in many events. Because of the resulting frequent and varied usage, words referring to local persons are prone to erosion and desemanticization. But differently from role markers, which do not have a referential function, once the form of a referential expression becomes too short to refer properly, a more expressive copy has to be added (following the proposal of Ariel, 1999). As a result, erstwhile local pronouns end up indexing the person of their helpers.

Of course, whether these findings extend to argument marking in natural languages too depends on (the ecological validity of) the model implementation. Care was taken to include independently motivated subroutines only. Moreover, virtually all model assumptions are parameterized and can thus be tested independently. The most problematic assumption at present seems to be the way in which words are selected from the mental lexicon, as we simply don't know yet how this works exactly. Also, language change involves more than just grammaticalization and grammaticalization involves more than erosion and desemanticization. But although the ecological validity of the model can thus be questioned, it still seems that only minimal assumptions about the initial linguistic system have to be made for an argument-marking system to develop.

Acknowledgments

I would like to thank three anonymous reviewers for their valuable comments and Sebastian Collin for fruitful discussion of the codes (which for example led to the improvement of the procedure to generate events).

This research was made possible by the Netherlands Organization for Scientific Research (NWO; Veni grant 27578001, *The exaptation of argument marking*), whose financial support is gratefully acknowledged.

References

- Arbib, M. A. (2015). Language evolution. An emergentist perspective. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (p. 600-623). West Sussex, UK: Wiley/Blackwell.
- Ariel, M. (1999). The development of person agreement markers: From pronouns to higher accessibility markers. In M. Barlow & S. Kemmer (Eds.), *Usage based models of language* (p. 197-260). Stanford: CSLI.
- Aristar, A. R. (1997). Marking and hierarchy. Types and the grammaticalization of case markers. *Studies in Language*, 21(2), 313-368.
- Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency in the pronunciation task: Lexical access and/or production? *Journal of memory and languages*, 24, 89-106.
- Beuls, K., & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLOS ONE*, 8(3). (e58960. doi:10.1371/journal.pone.0058960)
- Bybee, J. L. (1985). *Morphology. a study of the relation between meaning and form*. Amsterdam/Philadelphia: John Benjamins.
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(5), 489-509.
- Collin, S. (2015). *The rise and preservation of argument indexing systems*. (Bachelor thesis, Radboud University Nijmegen)
- Dahl, O. (1995). *The growth and maintenance of linguistic complexity*. Amsterdam/Philadelphia: John Benjamins.
- Deacon, T. (1997). *The symbolic species*. London: Penquin.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547-619.
- Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, p. 41-58). New York: Academic Press.
- Guiraud, P. (1968). The semic matrices of meaning. *Social Science Information*, 7(2), 131-139.
- Haspelmath, M. (1998). Does grammaticalization need reanalysis? *Studies in language*, 22(2), 315-351.
- Hawkins, S. (2002). Symmetries and asymmetries: Their grammar, typology, and parsing. *Theoretical Linguistics*, 28, 95-149.
- Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge, UK: Cambridge University Press.
- Heine, B., & Kuteva, T. (2007). *The genesis of grammar. a reconstruction*. Oxford: Oxford University Press.

- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge: Cambridge University Press.
- Hurford, J. R. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, 77(2), 187-222.
- 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* (p. 157189). London: UCL Press.
- Kirby, S. (2000). Syntax without natural selection. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language* (p. 303-323). Cambridge, UK.
- Lehmann, C. (1988). On the function of agreement. In M. Barlow & C. A. Ferguson (Eds.), *Agreement in natural languages* (p. 55-66). Stanford, CA: CSLI.
- Lehmann, C. (1995). *Thoughts on grammaticalization*. Lincom.
- Lestrade, S. (2010). *The space of case*. Unpublished doctoral dissertation, Radboud University Nijmegen.
- Müller, G. (2002). Remarks on nominal inflection in German. In I. Kaufmann & B. Stiebels (Eds.), *More than words: A festschrift for Dieter Wunderlich* (p. 113-145). Berlin: Akademie Verlag.
- Nettle, D. (1999). *Linguistic diversity*. New York: OUP.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 707-784.
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria.
- Siewierska, A., & Bakker, D. (2009). Case and alternative strategies: word order and agreement marking. In A. Malchukov & A. Spencer (Eds.), *The Oxford handbook of case* (p. 290-303). Oxford: Oxford University Press.
- Smith, K., & Kirby, S. (2008). Cultural evolution: Implications for understanding the human language faculty and its evolution. *Phil. Trans. R. Soc. B*, 363, 3591-3603.
- Steels, L. (1997). Constructing and sharing perceptual distinctions. *Machine Learning, ECML-97*, 4-13.
- Steels, L. (2011). Modeling the cultural evolution of language. *Physics of life reviews*, 8, 330-356.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- van Trijp, R. (2012). The evolution of case systems for marking event structure. In L. Steels (Ed.), *Experiments in cultural language evolution* (p. 169-205). Amsterdam: John Benjamins.
- Van Valin, R. (1999). Generalized semantic roles and the syntax-semantics interface. In F. Corblin, C. Dobrovie-Sorin, & J.-M. Marandin (Eds.), *Empirical*

issues in formal syntax and semantics 2 (p. 373-389). The Hague: Thesus.
Wierzbicka, A. (1996). *Semantics: Primes and universals: Primes and universals*. Oxford University Press.

DIFFERING SIGNAL-MEANING DIMENSIONALITIES FACILITATES THE EMERGENCE OF STRUCTURE

HANNAH LITTLE, KEREM ERYILMAZ AND BART DE BOER

*Artificial Intelligence Laboratory, Vrije Universiteit Brussel
Brussels, Belgium*

hannah@ai.vub.ac.be, kerem@ai.vub.ac.be, bart@arti.vub.ac.be

Structure of language is not only constrained by cognitive processes, but also by physical aspects of the signalling modality. We test the assumptions surrounding the role which the physical aspects of the signal space will have on the emergence of structure in speech. Here, we use a signal creation task to test whether a signal space and a meaning space having similar dimensionalities will generate an iconic system with signal-meaning mapping and whether, when the topologies differ, the emergence of non-iconic structure is facilitated. In our experiments, signals are created using infrared sensors which use hand position to create audio signals. We find that people take advantage of iconic signal-meaning mappings where possible. Further, we use trajectory probabilities and measures of variance to show that when there is a dimensionality mismatch, more structural strategies are used.

1. Introduction

Artificial language experiments have started to use continuous signal-space proxies to investigate the emergence of conventions, patterns and categories within signals (e.g. Verhoef, Kirby, and De Boer (2014) and Galantucci (2005)). However, studies have primarily focused on structural emergence being the result of cognitive processes within cultural transmission (Verhoef et al., 2014) or communication (Roberts & Galantucci, 2012). Here, we investigate how the mapping between the signal space and the meaning space influences the structure which emerges within signals. How signal structure is affected by different linguistic modalities in real world languages is important when considering how and why linguistic structure emerged. Also, when considering emerging structure in experiments using different artificial signal space proxies, it is very important to understand the effects that those proxies have on structure before attributing emerging signal structure to purely cognitive processes.

The dimensionality of a signal space is the number of ways which meaningful distinctions can be made using that signal space, e.g. in speech, voicing or place of articulation. As such, the dimensionality of a signal space will effect a) how quickly semantic distinctions outnumber the number of signal distinctions, and b) how difficult it is to map signals onto complex semantic spaces. Previously,

De Boer and Verhoef (2012) used a model to demonstrate that signal-meaning mappings are optimal when signal and meaning spaces share the same number of dimensions, and that when there is a mismatch between signal and meaning spaces, then more structural strategies are beneficial.

2. Experiments

We experimentally test how differences in the dimensionality of both the signal space, and the meaning space, will have on signal-structure. We are interested in what happens when there is a mapping possible between the dimensionality of the signal space and the meaning space, and what happens when there is a mismatch.

2.1. Experiment 1

2.1.1. Methods

Participants

25 participants, recruited at the Vrije Universiteit Brussel (VUB) in Brussels, took part in the experiment; 10 male and 15 female, with a mean age of 24 (SD = 4.6).

The signal space

An infrared sensor (*Leap Motion*) was used to create auditory signals generated from the hand positions of participants. Signals differed in pitch, volume, or both. Signals with two dimensions (pitch and volume), were created by moving one hand within a two dimensional space, i.e. moving a hand vertically would affect the volume, while a hand moving horizontally would manipulate the pitch. Both pitch and volume scales were non-linear. Signals could not contain gaps. In different phases, participants could either manipulate signals by moving their hand within a horizontal dimension, vertical dimension or both. Participants had time to get used to the mapping between their hand position and sound.

Procedure

There were three phases in the experiment (Figure 1), each phase had a practice round and an experimental round. Each round had a signal creation task and a signal recognition task. Practice rounds and experimental rounds were the same. Only the data from the experimental round was used in the analysis. Participants saw the entire meaning space before each phase. Signals were recorded for the squares one-by-one. Squares were presented in a random order. Participants could play back signals, and rerecord if they were not happy.

Phase 1:1 (matching) The meanings differed along one dimension (size, divided in 5 levels) and the signal space was also one dimensional (either pitch or volume).

Phase 2:2 (matching) Participants described the two-dimensional meaning space

(differing in size and shade), with a two-dimensional signalling space (pitch and volume).

Phase 1:2 (mismatching) Participants created signals for a two-dimensional meaning space (differing in size and shade). However, the signal space had only one-dimension (the same as they used in phase 1:1).

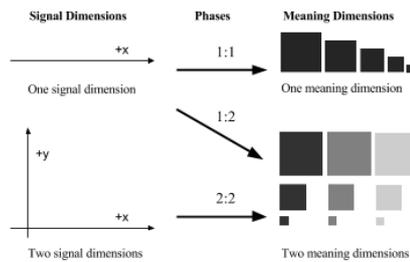


Figure 1. The phases used in the experiment.

Counterbalancing

Participants completed either phase 1:2 or 2:2 first.

Signal Recognition task

Participants heard a signal they had created, and were asked to identify its referent from an array of four randomly generated possibilities. They were given immediate feedback about the correct answer. This task worked as a pressure for expressivity, as participants knew they needed to recognise their own signals.

Post-experimental questionnaire

The questionnaire was free form and asked about the signal-creation strategies that the participant adopted during each phase of the experiment.

2.1.2. Results

Post-experimental questionnaire

Most self-reported strategies used pitch, volume or duration directly to encode size or colour where possible. Participants who saw phase 1:2 before phase 2:2 were more likely to self report using the same signalling strategy throughout, than to change the strategy to take advantage of both dimensions. This association was significant ($\chi^2(1) = 8.7, p < 0.01$).

Signal Recognition Task

Participants recognised a mean of 66% of signals correctly (25% expected by chance). When participants were incorrect, we were able to measure the distance between the answer they gave and the correct answer. Let m_{ij} define a meaning with size i and shade j . The distance between two meanings m_1 and m_2 is then the following:

$$D(m_1, m_2) = \sum_{k=i}^j |m_{1_k} - m_{2_k}|$$

Using this formula, we calculated the distance from the correct answer for both the actual data, and from data generated from choosing an answer at random. Comparing the actual data with the random data using a mixed effect linear model, and controlling for participant number as a random effect, and stimulus number as a fixed effect, we found that with incorrect choices produced in the matching phases, participants were closer to the correct square than if they had chosen at random ($\chi^2(1) = 5.5, p = 0.02$). However, in the mismatching phase there was no difference between actual incorrect choices and random incorrect choices ($\chi^2(1) = 0.01, p = 0.9$). Further, we found that the distance from the correct answer was much higher in the mismatching phases, than in the matching phases, indicating that participants were relying more on iconicity in the matching phases. We again tested this using a mixed effect linear model, and controlling for the same variables ($\chi^2(1) = 5.3, p < 0.05$).

Signal Creation Task

The data collected from the signal creation task consisted of coordinate values designating hand position at every time frame recorded. For this analysis, meaning dimensions were coded to reflect the continuous way in which they differed, e.g. the smallest square was 1 for size, and the biggest square was 5. Across all phases, the mean value of the first dimension that a participant saw in phase 1:1 (either pitch or volume) was predicted most strongly by shade. A mixed linear model, which included participant number as a random effect, and whether their starting dimension was pitch or volume as a fixed effect, showed this interaction to be significant ($\chi^2(1) = 341.4, p < 0.001$). The duration of the signal was predicted most strongly by the size of the square, with each step of size increasing the signal by $75.296 \text{ frames} \pm 7(\text{std errors})$. The mixed linear model for this interaction, controlling for the same fixed and random effects, was also significant ($\chi^2(1) = 103.14, p < 0.001$). These correlations demonstrate a propensity for using iconic strategies. Size and duration are easy to map on to one another, and it makes sense that participants will more likely encode the remaining meaning dimension (shade) with the signal dimension they were first exposed to.

Standard deviations (SD) of the signal trajectories gave us a good idea of the amount of movement in a signal. Signal trajectories produced in the mismatch phase had higher SDs than signals produced in matching phases. Using a linear mixed effects analysis and controlling for participant number as a random effect, and whether they started with pitch or volume as a fixed effect, we found that this finding was significant ($\chi^2(1) = 4.5, p < 0.05$).

Probability of signal trajectories

Another measure we used was the predictability of each signal trajectory, derived from a participant's entire repertoire. We calculated the probability of each individual signal coordinate and used this to find the joint probability of each signal. We did this by taking the negative logarithm of the product of first order conditional probabilities of the coordinates on the signal trajectory. Using a mixed effects linear model and controlling for duration as a random effect, and size of square and participant number as fixed effects, we found that signals generated in phases with matched signal and meaning dimensionality were significantly more predictable than in phases where there was no match ($\chi^2(1) = 3.9, p < 0.05$). Signals produced in the matching phases had higher predictability.

2.2. Experiment 2

Experiment 2 tested the same hypothesis as experiment 1, but the design was altered to counter two problems with experiment 1, a) that duration was used as a dimension by some participants, meaning that if participants used duration, there wasn't a "mismatch" even with the 1:2 phase, and b) that participants produced signals for an entire meaning space in experiment 1 (5 or 9 meanings depending on the phase), meaning that generating holistic signals for each meaning was a possibility.

2.2.1. Methods

Participants

Participants were recruited at the VUB in Brussels. 25 participants took part in the experiment; 8 male and 17 female. Participants had an average age of 21.

Signals

As in the first experiment, there was a continuous signal space which used *Leap Motion*. In this experiment, signals could only be manipulated in pitch. Including duration, the number of signal "dimensions" could not be more than 2.

Meanings

The meaning space consisted of a set of squares which differed along continuous dimensions. In contrast to the first experiment, the number of possible squares from the meaning space outnumbered the number of squares presented to the participants, this created an incentive for participants to create more productive systems which extend to meanings they have not seen yet. Meaning bottlenecks have been shown to encourage the production of structure in experiments such as Kirby, Cornish, and Smith (2008).

Procedure

The procedure in experiment 2 was exactly the same as experiment 1 but with different phases.

Phase 1:1 Signals only differed in pitch. There were 6 squares which differed in 6 degrees of size. Squares were presented in a random order.

Phase 1:2 Participants were presented with 12 squares which differed along two dimensions, 6 degrees of size and 6 shades of grey stripes (See Figure 2.) This made a possible number of 36 squares which were chosen from at random.

Phase 1:3 Participants were presented with 12 squares which differed along three dimensions, 6 degrees of size, 6 shades of grey stripes and 6 shades of orange stripes (See Figure 2.) This made a possible number of 216 squares which were chosen from at random. Stripes were used as pilots showed that this made the squares more easily distinguishable than changing the shade of orange to reflect both darkness and redness.

Signal Recognition task

The signal recognition task was the same as in the first experiment.

Post-experimental questionnaire

Questions were about the strategies that participants adopted during each phase.

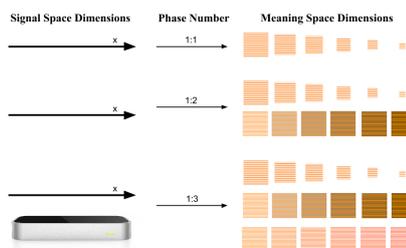


Figure 2. The dimensions used in experiment 2 in each of the 3 phases. All participants had phases presented in the same order: 1:1, 1:2, 1:3.

2.2.2. Results

Post-experimental questionnaire

In phase 1:1, participants self-reported encoding size with pitch or duration (80%). Participants tended to stick with the same strategy for size, but developed new strategies for new meaning dimensions. By phase 1:3, 56% of participants self-reported using a strategy which relied on movements or patterns.

Signal Recognition Task

Participants recognised their signals with a mean of 56% correct (again, chance level was 25%). When participants were incorrect, we were again able to measure the distance between their answer and the correct answer. Using a mixed effect linear model, and controlling for participant number as a random effect and square number as a fixed effect, we found that with incorrect choices produced across phases, participants were closer to the correct square than if they had chosen at random ($\chi^2(1) = 22.4, p < 0.001$) (see figure 3), the difference

between the actual and random result was significant within each phase. In later phases, there was more potential for incorrect distances being higher, because of the larger meaning space, meaning that a comparison between phases is not informative. However, the effect size for the comparison between the actual data and the random data in phase 1:3 was much smaller than in the other two phases, suggesting that in phase 1:3 there was less potential to rely on iconic strategies.

Signal Creation Task

The average duration of signals rose by about 20 frames each phase ($\chi^2(1) = 7.9$, $p < 0.005$)

As in experiment 1, meaning dimensions were coded to reflect the continuous way in which they differed. Across all phases, the size of square was the best predictor for the duration of the signal ($\chi^2(1) = 63.3$, $p < 0.001$). However, in this experiment, size was also the best predictor for the mean pitch of the signals ($\chi^2(1) = 15.7$, $p < 0.001$).

Looking at the SDs of individual signal trajectories, we see that the degree of mismatch affected the amount of movement in the signals. There was no significant difference between phases where there was no mismatch (Phases 1:1 and 1:2), in fact, the mean SD in these phases was nearly identical. However, the SDs from phase 1:3, where there was definitely a mismatch, were significantly higher than in the other two phases ($\chi^2(1) = 6.9$, $p < 0.01$).

Probability of signal trajectories

Calculating probabilities in the same way as in section 2.4m and using a mixed effects linear model, and controlling for duration and participant number as a random effect, and size of square as a fixed effect, we found that whether the signal was produced in a matching phase or not predicted how predictable a trajectory ($\chi^2(1) = 11.2$, $p < 0.001$). The log probability was closer to 0 (more predictable) in phase 1:1 (mean = 95), and got further away from 0 with each phase.

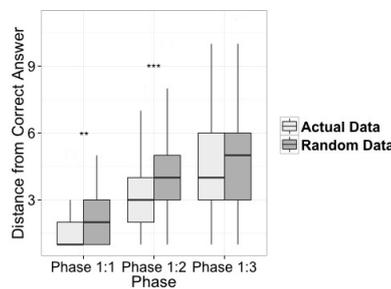


Figure 3. A graph showing the distance from the correct answer participants were in each phase when choosing incorrectly in the signal recognition task.

3. Discussion and Conclusion

In both experiments, we found correlation between structure in signal repertoires and structure in meaning spaces. This was particularly marked when signal and meaning spaces had the same number of dimensions. We also found more movement in signals in phases where there was a mismatch between signal and meaning spaces, suggesting more reliance on structural strategies.

Phases with matching dimensionalities produced signals which were more predictable given a participants entire repertoire, than signals produced within mismatching phases. This, again, may be indicative of the mismatching phases producing signals with more movement, as static, iconic, signals will be easier to predict.

We also found that in matching phases, when participants were incorrect, they were more likely to choose meanings which were closer to the correct meaning than if they'd chosen at random, again suggesting a reliance on iconic strategies.

We have shown that dimensionality, as a physical property of a signalling modality, will affect how, or ever if, structure will emerge. These findings are very important, both when exploring how structure emerges in signals produced using different linguistic modalities, but also when conducting experiments with signal-space proxies which have an effect on structure. Researchers must be careful to consider the physical effects of signal-space dimensionality when considering the emergence of structure. Importantly, understanding these effects will then allow us to isolate effects which are purely cognitive in nature.

Acknowledgments

Project supported by the ERC starting grant project ABACUS (283435).

References

- De Boer, B., & Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Advances in Complex Systems*, 15(3), 1150021-11150021-20.
- Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, 29(5), 737-767.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681-10686.
- Roberts, G., & Galantucci, B. (2012). The emergence of duality of patterning: Insights from the laboratory. *Language and cognition*, 4(4), 297-318.
- Verhoef, T., Kirby, S., & De Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43, 57-68.

CORRELATED EVOLUTION OR NOT? PHYLOGENETIC LINGUISTICS WITH SYNTACTIC, COGNACY AND PHONETIC DATA

GIUSEPPE LONGOBARDI,^{*†} ARMIN BUCH,[‡] ANDREA CEOLIN,[§] AARON ECAY,^{*} CRISTINA GUARDIANO,[¶] MONICA IRIMIA,^{*} DIMITRIS MICHELIODAKIS,^{*} NINA RADKEVICH,^{*} AND GERHARD JÄGER[‡]
**Department of Language and Linguistic Science, University of York, York, United Kingdom*

†Laboratorio di Linguistica e Antropologia cognitiva, Università di Trieste, Trieste, Italy

‡Seminar für Sprachwissenschaft, Universität Tübingen, Tübingen, Germany

§Department of Linguistics, University of Pennsylvania, Philadelphia, USA

¶Dipartimento di Comunicazione e Economia, Università di Modena e Reggio Emilia, Modena, Italy

In this work we compare, on the well explored domain of Indo-European languages, the phylogenetic outputs of three different sets of linguistic characters: traditional etymological judgments, a system for phonetic alignment of lists of cognates, and a set of values for generative syntactic parameters. The correlation and relative informativeness of distances and phylogenies generated by the three types of characters can thus be for the first time accurately evaluated, and the degree of success of the last two, innovative, alternatives to the classical comparative method can be so assessed.

1. Introduction

For many decades now historical linguistics has sought taxonomic procedures alternative and complementary to the classical comparative method, in order to strengthen and expand the results achieved through the latter. Two ideal adequacy standards such new methods should aim at are the possibility of applying also beyond the limits imposed by classical etymological criteria, and that of being easily subjected to quantitative analysis and automated statistical testing. For this purpose, we compare, on a domain where significant historical knowledge is already available (IE languages),

the computational outputs of three sets of linguistic characters: traditional etymological judgments, a system for phonetic alignment of lists of synonyms, and a set of values for generative syntactic parameters. The correlation and relative informativeness of distances and phylogenies generated by the three types of characters can thus be for the first time accurately evaluated.

2. Data being used

For our study we used three different data collections, providing syntactic, cognacy and phonetic information:

Syntactic parameters: The values for syntactic parameters are taken from the database of (Longobardi et al., 2013)^a integrated with data about 4 ancient languages (Latin, Classical Greek, Gothic, and Old English). In such a database, the two opposite values of the 56 binary parameters used were represented by ‘+’ and ‘-’; ‘0’ symbolizes instead a parameter value which is uninformative as fully predictable from the values of other parameters. Hence, such ‘0’s need to be disregarded for the purposes of taxonomic computations, a standard practice in the Parametric Comparison Method (Longobardi & Guardiano, 2009).

Lexical cognacy data: The IELex database (<http://ielex.mpi.nl/>) contains 207-item Swadesh lists for 157 living and extinct languages. Entries are assigned to cognate classes, based on expert judgments.

Phonetic data: The ASJP database (Wichmann et al., 2013) is a collection of 40-item Swadesh lists for more than 6,000 languages and dialects. All entries are given in a uniform phonetic transcription (see <http://asjp.c1ld.org/> for the actual data).

We identified a sample of 22 Indo-European languages (18 contemporary and 4 ancient ones) for which all three databases provide information. Syntactic parameters were organized in a binary matrix with languages as rows and parameters as columns.

From the data supplied by IELex, we constructed a binary matrix with languages as rows and cognacy classes (such as *dog-A* etc.) as columns. A cell has entry “1” if the row-language has an entry for the column-cognacy class, “0” if the row-language does not have an entry for the column-cognacy class but an entry for another cognacy class for that meaning, and “?” (undefined) otherwise. In total, IELex contains 1,566 cognate classes for the 22 languages in our sample. We excluded those cognate classes from

^a Available on line at: <https://benjamins.com/#catalog/journals/jhl.3.1.071on/additional>

consideration that are present in all 22 languages, such as class A for the concept 'I', which comprises English *I*, Hindi *me*, Latin *ego*, Spanish *yo* etc. (all deriving from the paradigm of PIE **h₁egh₂óm*). This leaves us with 1,553 informative binary characters.

Regarding phonetic data, we manually created a lookup table mapping IELex entries to corresponding ASJP entries for all language/concept pair where both databases contain the same word(s). There are several ASJP entries having no counterpart in IELex but clearly belonging to one of the IELex cognate classes. The Russian word **пос** ('dog', i.e. a synonym to **собак3**), for instance, is evidently cognate to the Serbocroatian **пас** and the Polish *pies*. We did not include this kind of information into the cognacy character matrix, but we used it to create sound alignments. (An ASJP entry without IELex counterpart was automatically added to a cognate class if its average string similarity to the members of that class exceeds a certain threshold.)

In a next step, the T-Coffee algorithm (Notredame et al., 2000) was applied to perform *multiple sequence alignment* within each cognate class (see Jäger & List, 2015 for a fuller description of how T-Coffee was adapted to phonetic strings).^b

This is illustrated for the concept *dog* in Tab. 1.

Ideally, sounds within the same column are cognate, i.e. they derive historically from the same ancestor. A “-” represents a gap, i.e. a position where a sound has been deleted or added in the lineage leading to that language. A “.” indicates that the language in question does not contain a word from that cognate class.

As the multiple sequence alignment has been performed automatically, an experienced historical linguist will not agree with every detail. Most alignments arguably capture genuine sound correspondences though. As there are no gold standard data of that type, it is at present not possible to quantify the quality of our alignments.

The multiple sequence alignments can be transformed into a binary character matrix in the familiar way. For a given column, all sound types occurring there define one binary character. Gaps are not treated as characters. If a language has a “.” in a column, all characters from that column are undefined for that language. Again excluding non-informative characters, we end up with a binary character matrix with 1,521 columns.

^bIn (Jäger & List, 2015) a systematic comparison is conducted between Sound Class Based Phonetic Alignment (SCA, List, 2014) and T-Coffee alignment, indicating the latter to be superior.

Table 1. ASJP entries and multiple alignment for cognate classes A, H, and E/concept ‘dog’

<i>language</i>	<i>ASJP entries</i>	<i>multiple sequence alignment</i>
Italian	kane	k---a-ne---.....
Spanish	-
French	Sia	S----ia---.....
Portuguese	kau	k---a-u---.....
Romanian	k3ne	k---3-ne---.....
Greek	-
English	-
German	hunt	h---u-n-t---.....
Danish	hun7	h---u-n-7---.....
Icelandic	hintir	h---i-n-tir.....
Norwegian	hund	h---u-n-d---.....
Bulgarian	-
Serbocroatian	pasp-as
Polish	piEspiEs
Russian	sobak3,pos	sobak3-----.....p-os
Irish	ku	k---u-----.....
Marathi	k3tra7k3tra7....
Hindi	kuttakutta-....
Latin	kanis	k---a-nis---.....
Classical Greek	kion	k--i-o-n---.....
Gothic	hunds	h---u-n-d-s.....
Old English	hund	h---u-n-d---.....

3. Distances and correlations

From each of those three binary matrices we computed pairwise distances between languages. For syntactic parameters, the possible values “+” and “-” are symmetric. Therefore the syntactic distance was defined as the Hamming distance between their parameter vectors. As for the cognacy and phonetic matrices, a “1” represents the presence and a “0” the absence of a certain trait in that language. Therefore the *Dice distance* was used (as proposed originally in Longobardi et al., 2015 in such a context), which is defined as

$$d(A, B) \doteq 1 - \frac{|A \cap B|}{1/2(|A| + |B|)},$$

where A is the set of characters where the first language has entry “1”, and likewise for B .

The three distance measures defined above quantify for each pair of languages on how many syntactic parameters they disagree (syntactic distances), how many basic concepts are expressed by unrelated words (cognacy distance) and how many sound pairs participating in regular correspondences are non-identical (phonetic).

We observe the correlations given in Tab. 2 and depicted in Fig. 1. All

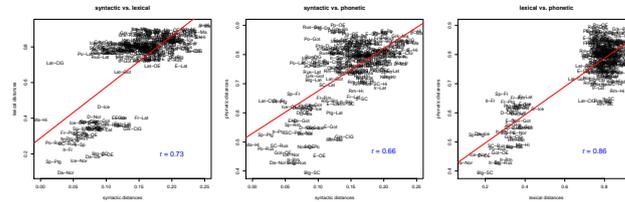


Figure 1. Direct correlations

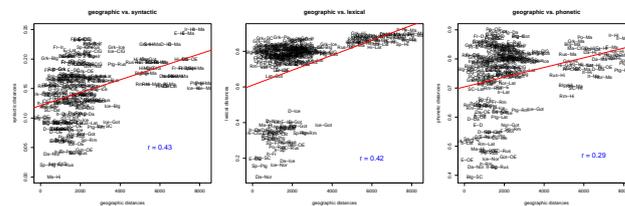


Figure 2. Correlations with geographic distance

correlations are significant ($p < 10^{-4}$ according to the Mantel test).

Table 2. Direct (left) and partial (right) correlations

	syntactic	cognacy		syntactic	cognacy
cognacy	0.73	—		0.67	—
phonetic	0.66	0.87		0.62	0.85

Syntactic ($r = 0.43, p < 10^{-4}$), cognacy ($r = 0.42, p < 10^{-4}$) and phonetic ($r = 0.291, p < \times 10^{-4}$) distances are significantly correlated with geographic distance; cf. Fig. 2.

The correlations with geography may reflect both common descent (as closely related languages tend to be geographically proximate) and effects of language contact. To control for the latter, we computed the partial pairwise correlations between the three linguistic distances conditioned on geographic distance. All three correlations remain strongly significant ($p < 10^{-4}$ according to partial Mantel test) when we control for geography.

4. Phylogenetic inference

Using phylogenetic inference, we can construct the evolutionary scenario best explaining observed data. We performed such inferences for each of

the three data sources separately, using two different approaches.

Distance-based inference The simplest family of methods only rely on the assumption that on average, the distance between species/languages increases after they diverge. The relation between distances and divergence times can be noisy. Distance-based inference take a pairwise distance matrix as input and find a tree (with branch lengths specification) such that the path length between two leaves is as close as possible to the input distance between these two taxa.

Here we will consider the “Kitsch” algorithm from the *Phylyp* software package (Felsenstein, 1989). It uses a weighted least squares method to assess the fit between a distance matrix and a tree, and it finds a tree that minimizes this distance under the

constraint that all leaves have the same distance from the root. This amounts to the assumption that evolution proceeds at the same pace across lineages (the so-called “molecular clock” assumption). It is of course only applicable if all leaves are contemporary. Therefore we will only consider the 18 recent languages in our sample in the sequel.

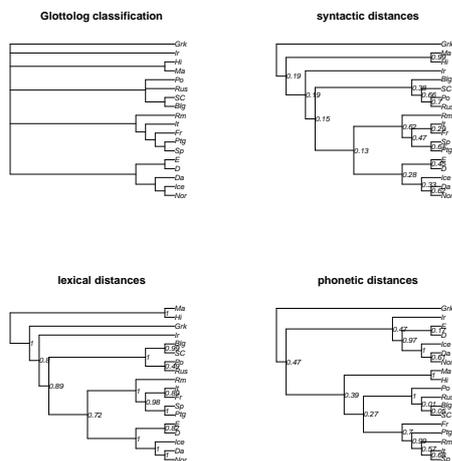


Figure 3. Glottolog expert tree and Kitsch-generated trees; interior node labels are bootstrap support values

giving *bootstrap support values* for the corresponding clade (See Appendix for the technical definition). They indicate how strongly a certain clade is supported by the data.

We mostly find high support values for the established sub-groups of Indo-European and lower values for the other nodes. The support values in the syntactic tree are generally rather low. This does not indicate though that syntactic information is less reliable, but it reflects the relative sparseness

of the syntactic data.

The topologies of the four trees display remarkable similarities; in particular, three trees (i.e. all but the phonetic one) are consistent in singling out (in slightly different clusterings) Greek, Irish and Indo-Aryan languages from the core formed by Romance, Germanic and Slavic. The phonetic tree does not follow the general pattern on this point, but is suggestive that it groups together Slavic and Indo-Aryan, the two satəm families of the sample, in agreement precisely with one of the longest known phonological differentiations within IE (the treatment of velar consonants). Both the cognacy and the syntactic tree favor a closer relation of Germanic and Romance as opposed to Slavic. At this stage we can only notice that the combination of the latter two observations (i.e. the complementarity of the various methods) may usefully capture some plausible areal effect (reflecting the intermediate geographical position of Slavic, of course).

The quality of an automatically generated tree can be measured by quantifying the degree of its disagreement with the expert classification. A suitable distance measure is the *Generalized Quartet Distance* (GQD; see (Pompei et al., 2011), where it is argued convincingly that GQD is more informative than other tree distance measures such as the Robinson-Foulds distance). Briefly put, this is a measure for the recall of an automatically generated tree being evaluated against an expert tree, i.e. its proportion of topological information in the expert tree that is not correctly recovered in the induced tree. (It is not possible to determine a corresponding precision score since the true topology is usually not known; expert trees are generally underspecified in various respects.)

The GQDs of the three Kitsch-generated trees to the Glottolog tree are 0.031 for syntactic, 0.019 for cognacy and 0.054 for phonetic distances.

Character-based inference Distance-based methods infer a tree explaining diversification patterns in a summary fashion. Character-based phylogenetic inference is a family of more advanced methods that model not just an evolutionary tree but the history of change of each feature (a.k.a. character) along the branches of this tree. We chose Maximum-Likelihood estimation as a representative of those methods to infer evolutionary histories for syntactic, cognacy and phonetic characters.^c The resulting trees are shown in Fig. 4.

The GQDs of the three Maximum-Likelihood trees to the Glottolog tree are 0.019 for syntactic and cognacy, and 0.059 for phonetic distances.

^cThe calculations were carried out using the software package *Paup** (Swofford, 2002). For all three data sets, we chose the model with molecular clock and gamma-distributed rates. Rates, base frequencies and proportion of invariant sites were estimated.

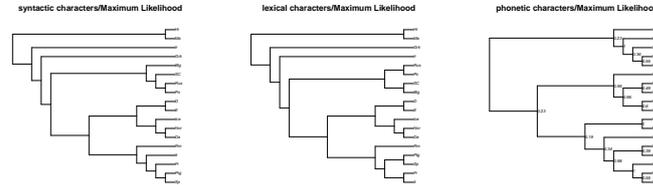


Figure 4. Maximum-likelihood trees

Generally, the syntactic and the cognacy tree are in good agreement both with each other and with the expert tree. The phonetic tree displays some unexpected groupings. To understand this effect better, we performed a Bayesian analysis with the phonetic data (using the Beast software package, <http://beast.bio.ed.ac.uk/>) and used the posterior sample to obtain confidence values for the nodes in the ML tree (displayd in Fig. 4). It turned out that almost all clades with high confidence values (≥ 0.9) correspond to established groupings (Romance, Slavic, Germanic, North-Germanic; the only exception being a group comprising all Romance languages except French). All nodes beyond the sub-families have extremely low support (< 0.5), i.e. they are artefacts of the inference algorithm without real support in the data.

5. Conclusion

By using a variety of methods (correlation studies, distance based phylogenetic inference and character based phylogenetic inference), we provided evidence (a) that different aspects of the language system (syntax, lexicon, sound system) preserve essentially the same historical signal and (b) the insights established by the comparative method are essentially supported by all three signals considered here. These results have a potential impact on future research in phylogenetic linguistics because they indicate that phylogenetic techniques might be suitable to reconstruct earlier language stages by statistical means, and that different linguistic domains inform each other in this endeavor.

Acknowledgements

This work was supported by the ERC Advanced Grant 295733 *LanGeLin* (GL, AC, AE, CG, MI, DM, NR), the ERC Advanced Grant 324246 *EVOLAEMP* (AB, GJ), as well as the DFG-Center for Advanced Studies in the Humanities 2237 *Words, Bones, Genes, Tools* (GJ), which is gratefully acknowledged.

References

- Felsenstein, J. (1989). Phylip — Phylogeny Inference Package (Version 3.2). *Cladistics*, 5, 164-166.
- Jäger, G., & List, J.-M. (2015). Factoring lexical and phonetic phylogenetic characters from word lists. In H. Baayen, G. Jäger, M. Köllner, J. Wahle, & A. Baayen-Oudshoorn (Eds.), *Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics (QITL 6)*. (to appear)
- List, J.-M. (2014). *Sequence comparison in historical linguistics*. Düsseldorf: Düsseldorf University Press.
- Longobardi, G., Ghirotto, S., Guardiano, C., Tassi, F., Benazzo, A., Ceolin, A., & Barbujani, G. (2015). Across language families: Genome diversity mirrors linguistic variation within Europe. *American journal of physical anthropology*. (doi: 10.1002/ajpa.22758)
- Longobardi, G., & Guardiano, C. (2009). Evidence for syntax as a signal of historical relatedness. *Lingua*, 119(11), 1679-1706.
- Longobardi, G., Guardiano, C., Silvestri, G., Boattini, A., & Ceolin, A. (2013). Toward a syntactic phylogeny of modern indo-european languages. *Journal of Historical Linguistics*, 3(1), 122-152.
- Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology*, 302(1), 205-217.
- Pompei, S., Loreto, V., & Tria, F. (2011). On the accuracy of language trees. *PLoS One*, 6(6), e20109.
- Swofford, D. (2002). *Phylogenetic analysis using parsimony (* and other methods)*. Sunderland, MA: Sinauer Associates.
- Wichmann, S., Müller, A., Wett, A., Velupillai, V., Bischoffberger, J., Brown, C. H., Holman, E. W., Sauppe, S., Molochieva, Z., Brown, P., Hammarström, H., Belyaev, O., List, J.-M., Bakker, D., Egorov, D., Urban, M., Mailhammer, R., Carrizo, A., Dryer, M. S., Korovina, E., Beck, D., Geyer, H., Epps, P., Grant, A., & Valenzuela, P. (2013). *The ASJP Database (version 16)*. <http://asjp.cld.org/>.

LINKING THE PROCESSES OF LANGUAGE EVOLUTION AND LANGUAGE CHANGE: A FIVE-LEVEL HIERARCHY

JEROME MICHAUD

*ICMCS, School of Physics and Astronomy, University of Edinburgh,
Peter Guthrie Tait Road, Edinburgh EH9 3FD UK
v1jmicha@staffmail.ed.ac.uk*

The class of problems that can be categorized under *language evolution and change* is very heterogeneous and involves many different timescales and spatial/social scales. In order to better understand the underlying evolutionary processes and the link between language evolution and language change, we introduce a five-system hierarchy based on social structures of the population and apply it to the study of the transition from protolanguage to language. This problem is composed of a biological subproblem and a cultural subproblem. Using our hierarchy, we argue that Bickerton's "living fossils" of language can give some insight only on the cultural subproblem, but not on the biological one or on the coevolution of the two.

1. Introduction

The language faculty is at the heart of what makes us human. It has permitted us to develop increasingly complex societies and cultures, but the way this faculty emerged and subsequently evolved is still to a large extent a mystery, see Hauser et al. (2014). This is due to the complexity of the problem: language evolution and change is not a single and well-defined process. It covers a number of phenomena ranging from language's first appearance in mankind to language acquisition by children, through language use, cultural evolution of languages, language death, new dialects, creole and pidgin formation, to name but a few. In the literature, one usually refers to the language emergence problem by language evolution and to its subsequent evolution by language change. The heterogeneity of language evolution and change problems is such that a single mechanism of evolution is unlikely to be sufficient to explain them all; one therefore has to rely on many different evolution processes. In Kirby (2007), Christiansen and Kirby (2003) it is shown that the biological, cultural and individual evolution processes are the three main processes that drive the evolution of languages. In order to understand the emergence and evolution of a language one has to understand the interplay of these three kinds of processes. As a result, language is a *complex adaptive system*, (Cornish, Tamariz, & Kirby, 2009; Beckner et al., 2009; Steels, 2000).

Given the complexity of language evolution and change problems, it is sometimes difficult to identify which processes are the most relevant to study a specific

question. The main aim of this paper is to provide a systematic approach to analyse a language evolution problem and to identify the most relevant evolutionary mechanisms. Some attempts to achieve this goal have been for example discussed in Enfield (2014) in the broader context of causal relations in language. Our approach is based on an evolutionary point of view, which is best described when adopting a usage-based definition of language. Starting from the three types of evolutionary processes, we then identify five different levels of description characterized by their timescale and their population structure. Joining these two components provides a systematic approach to analyse a given question in language evolution.

In order to illustrate how this new approach can be applied, we consider the problem of the transition from protolanguage to language and discuss in the light of our framework the relevance of different *living fossils* of language first introduced by Bickerton (1995, appendix A). In this paper, the existence of some kind of protolanguage is taken for granted and we will not discuss it in detail. The purpose of this example is to shed light on the relations between language evolution and language change problems.

2. Evolutionary processes and the five-level hierarchy

2.1. Usage-based definition of language and social structures

In this paper, we assume that language is a conventionalised semiotic system composed of LINGUEMES (Croft, 2000). The representation that every individual have of the language is conditioned on her history and is usually called an IDIOLECT. The conventionalisation process mediated by language usage occurs in a complex population. One can identify the homogeneous subgroups of this population as SPEECH COMMUNITIES and their averaged language is called a CODE (Croft, 2000). These subgroups have the property to be both well-connected and socially homogeneous. The next level of social structure is when the population is still well-mixed, but socially structured, that is, when there are many speech communities. In this case, the corresponding subgroups form LINGUISTIC COMMUNITIES. The linguistic system used at this level corresponds to what we call a LANGUAGE. If different subgroups are only loosely connected, many linguistic communities emerge and languages are in contact. This is for example the case in creole and pidgins. We assume that the different subgroups have similar language capacities. This is no longer the case if we consider different interacting species. The level of development of their LANGUAGE FACULTY is different and it is not even clear if other species have such a faculty.

2.2. Evolutionary account

With respect to language evolution, the biological evolutionary processes are concerned with brain evolution and the biological component of our language faculty;

cultural processes are responsible for the propagation and the conventionalization of the language through a population and depend on the social and/or cultural structures of the population; individual processes are responsible for language learning and language use. They are influenced by the cognitive architecture and by the psychology of the individuals. Following Kirby (2007), Christiansen and

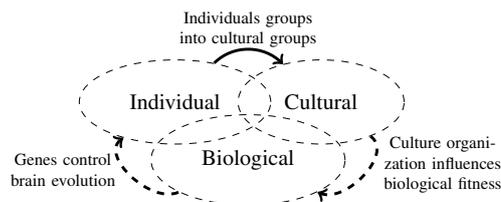


Figure 1. Languages evolve and change by three main processes: individual evolution, cultural evolution and biological evolution. The dashed arrows mean that links are weaker than the others. This figure has been adapted from Kirby (2007), Christiansen and Kirby (2003)

Kirby (2003), Figure 1 sketches the relation between the different types of evolutionary processes. The link between cultural and biological evolution is not really transparent and deserves some comments. In fact, it is not clear how cultural evolution affects the biological fitness and therefore the better survival of the population. Furthermore, the link between biological and individual evolution is weak because of the huge timescale difference between the two types of processes.

2.3. The five systems of language evolution and change

Comparing the language definition and the corresponding social structures with the three evolutionary processes mentioned above, one can assign to every social structure the main corresponding evolutionary mechanism together with the language type and corresponding system as shown in Table 1. An alternative

Table 1. Correspondence between social structures, language type, systems and evolutionary processes.

Level	Social struct.	Lang. type	System	Evo. process
1	Ind.	Idiolect	Ling. System	Individual
2	Homo.	Code	Multi-agent	Cultural
3	Struct.	Language	Syst. of Speech Com.	Cultural
4	Hetero.	Set of lang.	Syst. of Ling. Com.	Cultural
5	Homo.	Lang. faculty	Ecosystem	Biological

representation of this table is given in Figure 2 that represents the five levels of language evolution and change. The solid (red) arrows means “groups into” and each level can be considered as a system where the parts are elements of the previous level, see Table 1. With this association, one can study the dynamics of a

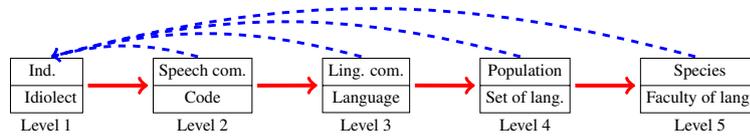


Figure 2. Illustration of the structures underlying the existence of a language in complex populations. The solid (red) arrows means “groups into” and the dashed (blue) arrows symbolize the influence of higher level structures on the smallest one as an individual is aware of the population structures and this influences her behavior.

system by studying the interaction between its parts. Consider for example level 4, a population is a system of linguistic communities (level 3) that interact through the creation and evolution of inter-linguistic speech communities (level 2). The dashed (blue) arrows in Figure 2 underline the fact that an individual is aware of the population structures and acts accordingly. There is therefore not only a micro-to-macro connection between the systems, but also a macro-to-micro relationship. Note that Table 1 corresponds to the state of a fully developed language. At earlier stages, the distinction between levels 2-4 only relies on the social structures of the population.

3. Case study: Protolanguage to language transition

In order to illustrate how this hierarchy of systems can be applied in practice, we consider the case of the transition from protolanguage to language. Since this is a difficult problem, some assumptions are discussed below.

3.1. Assumptions

We assume the existence of a non-holistic protolanguage (Bickerton, 1995; Jackendoff, 1999), that is, a low-structured linguistic system composed of words that can be juxtaposed, but lacking any kind of higher-level grammatical structures. Another kind of protolanguage has been proposed by Wray (1998). She proposed a holistic view of protolanguage, where each utterance is unstructured, that is, it can not be decomposed into words. In the literature, the question of the nature of protolanguage, holistic or not, has been discussed extensively, see for example Hurford (2000), Arbib (2005), Tallerman (2007).

We also assume that the transition from protolanguage to language is gradual, see for example (Kirby, 2007; Sandler et al., 2014) and contrary to Davidson and Noble (1994). One comment has to be made here: since the transition is gradual the categorical distinction between a protolanguage and a language is somehow misleading. To avoid misinterpretation, when we speak about protolanguage, we always mean a low-structured language, whereas a language is high-structured. The emergence of complex grammatical structures seems to be in contradiction with the gradual evolution of the language faculty. In fact, this can be understood

since we are dealing with a complex system. One characteristic of such systems is to have phase transitions, that is, some minor changes can fundamentally change the behaviour of the system. In other words, the emergence of the faculty to handle complex structures may be the consequence of small changes in the brain organisation or be an artefact of the timescale difference between cultural and biological evolution.

3.2. *Analyse of the protolanguage to language transition*

We can now analyse the protolanguage to language transition. In the light of our hierarchy of systems, this transition assumes a change in the cognitive abilities of the population. The phylogenetic change therefore belongs to level 5 of Figure 2^a. This change occurs in the ecosystem to which humans belong. The emergence of the cognitive ability to deal with complex structures and the conditions in which this faculty fixed in the human genotype remains to a large extent a mystery (Hauser et al., 2014). It is really difficult to decide whether this faculty emerges by adaptation, exaptation or simply by random drift.

It is important to note that having the faculty to process complex utterances does not directly imply the existence of a complex linguistic system. This faculty is necessary, but not sufficient. In order for a fully-fledged language to emerge, complex structures have to be innovated and conventionalised by the population. The problem of protolanguage to language transition can therefore be decomposed into two interconnected subproblems: (i) understanding the genetic changes that enable a faculty to deal with complex language structures and (ii) explaining the emergence of a complex language from a protolanguage by a succession of innovation and conventionalisation processes. The first subproblem is biological and the second is cultural. I will argue that “living fossils” of language can be used to gain insight into the second problem, but the first one seems to be out of reach.

Going back to our hierarchy of systems, the first subproblem belongs to level 5, but the second problem belongs to lower levels 2-4, since it is concerned with a cultural process. In fact, we consider that the transition from a protolanguage to a language occurred through a series of language contact problems where (proto)languages of different complexity were in contact. For a successful transition to occur, more complex languages should have outcompeted simpler ones. This problem of language competition therefore belongs to level 4 and should be studied using inter-linguistic speech communities and their internal dynamics (level 2, multi-agent system).

Since the biological and the cultural evolution processes coevolved, the capacity to deal with more complex languages is conditioned on the genetic characteristics of the different linguistic communities and many generations are needed to complete a transition from a low-structured to a high-structured language.

^aFrom now on, the mention of levels assumes the reference to Figure 2.

3.3. *Living fossils*

We now come to the discussion of “living fossils” of language. We discuss the three original examples given by Bickerton (1995, App. A). These three protolanguages are: (i) child language, (ii) pidgin and creole languages and (iii) trained apes language. These living fossils have been claimed by different authors to shed light on the protolanguage to language transition. Such a parallel is for example suggested by Kirby (2007, p.9). In principle one could choose other examples that are probably better examples of protolanguages such as home sign or developing sign languages, see (Sandler et al., 2014; Kegl, Senghas, & Coppola, 1999). However, this choice has the advantage to span all the different levels proposed in our hierarchy and therefore best demonstrates the power of our approach.

3.3.1. *Child language*

In the case of child speech evolution, the transition from a low-structured to a high-structured language is mediated by the learning process. This process is internal to the family speech community (level 2) that can be further decomposed into the children speech community and the adults’ one. That is, the systems which are involved in this process are the linguistic system (level 1) and the multiagent system (level 2). The process analogous to the conventionalisation problem is learning. This analogy is fairly weak, since learning and conventionalisation occur at two different levels of the hierarchy and through different evolutionary processes. Learning is driven by individual evolution with contact with a speech community of speakers of a fully-fledged language, whereas the conventionalisation problem occurs in a relatively well-mixed population of people with similar language abilities. This parallel is therefore pretty weak. However, learning also played a role in the transition from protolanguage to language since each new generation had to learn to communicate with the previous generations. The process analogous to the biological process of language emergence (level 5) is brain maturation (level 1). In our hierarchy, level 5 and level 1 are very different. For this analogy to be explanatory, one has to accept Haekel’s *recapitulation theory* (Haekel, 1874), which states that: “in developing from embryo to adult, animals go through stages resembling or representing successive stages in the evolution of their remote ancestors”, that is, level 1 represents level 5. This is sometimes referred to as the *ontogeny recapitulates phylogeny* assumption (Mufwene, 2008). This theory could justify the biological analogy and it has been shown to be plausible to a certain extent by Kalinka and Tomancak (2012). Note that the recapitulation theory has been disputed in the literature, see for example Payne and Wenger (1998), Løvtrup (1978) or Mufwene (2008). As a result, one can say that the conventionalisation problem between children and adults is relevant for the protolanguage to language transition since during the transition, many generations of learners have been involved. Explanations of biological evolution heavily relies on the recapitulation theory,

but this theory says nothing about the selection mechanisms that have constrained brain evolution. Therefore, child language can not give insight into the biological evolution problem or its coevolution with cultural processes.

3.3.2. *Pidgin/Creoles*

In the pidgin case, the transition between a low-structured and a high-structured language (level 5, since it is assumed to be biologically encoded) is arguably in analogy with the pidgin to creole transition assumed by Mühlhäusler (1986). In the five-level hierarchy, the *creolisation* process involves a system of linguistic communities (level 4) that interact through the creation of speech communities (level 2). The structure of this speech community and its evolution is not much discussed in the literature. But in Mühlhäusler (1986), a pidgin is characterized by the fact that it does not have native speakers, whereas creole languages do. This means that a transition between a pidgin and a creole needs to be driven by a change in the inter-linguistic speech community structure. There is no reason for this structural change to be necessary (Mufwene, 2008). Moreover, it is possible that the inter-linguistic speech community always includes children and in this case, no pidgin will ever be formed and a creole or another type of mixed language can directly emerge. As a result, the transition from pidgin to creole is not necessary and might even be wrong (Mufwene, 2008).

On the one hand, pidgins and creoles are formed inside inter-linguistic speech communities through conventionalisation processes that might be close to those involved in the cultural evolutionary processes active in the protolanguage to language transition, at least inside the homogeneous inter-linguistic community. However, the presence of fully-fledged languages in surrounding communities is weakening this analogy. On the other hand, there is clearly no biological process in this case and the biological subproblem of the protolanguage to language transition has no analogue. As a result, this language fossil might be useful to explain the cultural aspect of the protolanguage to language transition, but the biological aspect and the coevolution of the two seems to be out of reach.

3.3.3. *Trained Apes*

The third language living fossil that Bickerton has proposed is the trained animal language, in particular, the trained apes language illustrated by the case of Koko, a female lowland gorilla, see Patterson (1978). This case is fairly close to the pidgin/creole case, since it concerns a contact between two linguistic communities (level 4), the animal one and the human one, and the subsequent evolution of the corresponding code (level 2). The structure of the resulting code is quite close to the structure of a pidgin (Bickerton, 1995). It is interesting to note that when the animal starts to learn a human language when it is still young, which is the case of Koko, then the structures of the emerging code are more complicated, reflecting in

a different situation the same kind of distinction as between a pidgin and a creole.

The trained apes case can be argued to belong to level 5 of the hierarchy since it involves two different species. However, the type of evolutionary processes involved is cultural, not biological, between two linguistic communities with a different language faculty. This situation is quite close to the competition between communities with different language faculties as is the case in the protolanguage to language transition. The study of this fossil is therefore informative on the cultural evolutionary aspect. As for the pidgin/creole case, the existence of a fully-fledged language is weakening the analogy and one has to be careful when interpreting the results. Furthermore, the main problem is the biological evolutionary aspect of the protolanguage to language transition. If one wants to make a useful analogy, the Haeckel's recapitulation theory (Haeckel, 1874) has to be assumed and the same problems as in language acquisition appear. One can therefore conclude that the cultural aspect of the protolanguage to language transition can be partially studied through experiments with animals, even though a fully-fledged language exists in this case. However, the biological evolutionary aspect of the protolanguage to language transition and the coevolution between biological and cultural processes seems once again to be out of reach.

4. Discussion and concluding remarks

In this paper, we have introduced a five-level hierarchy of systems that serves as a tool to analyse problems in language evolution and change. This tool allows us to quickly and systematically identify the relevant social structures underlying the evolutionary problem at hand. As an illustration, we have considered the protolanguage to language transition and have discussed the relevance of the three living fossils proposed by Bickerton (1995). We have shown that the "living fossils" proposed can give some insight on the cultural evolution subproblem of the protolanguage to language transition, but not on the biological aspect or the coevolution of the cultural and biological problems, since this assumes the acceptance of Haeckel's recapitulation theory. This comparison provides an important link between language evolution and language change processes.

Our example corresponds to a critical application of the five-level hierarchy of systems. One can also apply it in a productive way as a guiding principle to develop evolutionary scenarios or mathematical models of language change. For example, the process of creation and propagation of a convention in a speech community should be treated in a cultural evolution framework, not in a biological one as it is for example the case in the work of Nowak and Krakauer (1999).

Acknowledgements

I would like to thank Richard A. Blythe and Edwige Dugas for interesting discussions and comments. This work was supported by the Swiss National Science Foundation under grant number P2GEP2_159156.

References

- Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and brain sciences*, 28(02), 105–124.
- Beckner, C., Blythe, R. A., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59, Suppl. 1, 1–26.
- Bickerton, D. (1995). *Language and human behavior*. University of Washington Press.
- Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in cognitive sciences*, 7(7), 300–307.
- Cornish, H., Tamariz, M., & Kirby, S. (2009). Complex adaptive systems and the origins of adaptive structure: What experiments can tell us. *Language Learning*, 59, Suppl. 1, 187–205.
- Croft, W. (2000). *Explaining language change: An evolutionary approach*. Pearson Education.
- Davidson, I., & Noble, W. (1994). In K. R. Gibson, K. R. Gibson, & T. Ingold (Eds.), *Tools, language and cognition in human evolution* (pp. 125–155). Cambridge University Press.
- Enfield, N. J. (2014). *Natural causes of language: Frames, biases, and cultural transmission*. Language Science Press.
- Haekel, E. (1874). *Anthropogenie; oder, entwicklungsgeschichte des menschen, keimes- und stammesgeschichte*. Leipzig, W. Engelmann.
- Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., Chomsky, N., & Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in psychology*, 5.
- Hurford, J. R. (2000). Social transmission favours linguistic generalization. *The evolutionary emergence of language: Social function and the origins of linguistic form*, 324–352.
- Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in cognitive sciences*, 3(7), 272–279.
- Kalinka, A. T., & Tomancak, P. (2012). The evolution of early animal embryos: conservation or divergence? *Trends in Ecology & Evolution*, 27(7), 385–393.
- Kegl, J., Senghas, A., & Coppola, M. (1999). *Creation through contact; sign language emergence and sign language change in nicaragua*. Cambridge, MA: MIT Press.
- Kirby, S. (2007). The evolution of language. *Oxford handbook of evolutionary psychology*, 669–681.
- Løvtrup, S. (1978). On von baerian and haeckelian recapitulation. *Systematic*

- Biology*, 27(3), 348–352.
- Mufwene, S. S. (2008). *Language evolution: Contact, competition and change*. Bloomsbury Publishing.
- Mühlhäusler, P. (1986). *Pidgin and creole linguistics*. Blackwell Oxford.
- Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14), 8028–8033.
- Patterson, F. G. (1978). The gestures of a gorilla: Language acquisition in another pongid. *Brain and Language*, 5(1), 72–97.
- Payne, D. G., & Wenger, M. J. (1998). *Cognitive psychology*. Houghton Mifflin College Division.
- Sandler, W., Aronoff, M., Padden, C., Meir, I., Enfield, N., Kockelman, P., & Sidnell, J. (2014). Language emergence: Al-sayyid bedouin sign language. *Cambridge handbook of linguistic anthropology*.
- Steels, L. (2000). Language as a complex adaptive system. In M. Schoenauer et al. (Ed.), *Parallel problem solving for nature—ppsn vi* (Vol. 1917, pp. 17–26). Berlin: Springer.
- Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, 117(3), 579–604.
- Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 10(1), 47–67.

INTERACTION FOR FACILITATING CONVENTIONALIZATION: NEGOTIATING THE SILENT GESTURE COMMUNICATION OF NOUN-VERB PAIRS

ASHLEY MICKLOS

*Applied Linguistics and Anthropology, University of California, Los Angeles,
Los Angeles, CA
amicklos@g.ucla.edu*

This study demonstrates how interaction – specifically negotiation and repair – facilitates the emergence, evolution, and conventionalization of a silent gesture communication system. In a modified iterated learning paradigm, partners communicated noun-verb meanings using only silent gesture. The need to disambiguate similar noun-verb pairs drove these "new" language users to develop a morphology that allowed for quicker processing, easier transmission, and improved accuracy. The specific morphological system that emerged came about through a process of negotiation within the dyad, namely by means of repair. By applying a discourse analytic approach to the use of repair in an experimental methodology for language evolution, we are able to determine not only if interaction facilitates the emergence and learnability of a new communication system, but also how interaction affects such a system.

1. Introduction

The purpose of this experiment is to investigate the evolution of a communication system, along with other discourse features that accompany such a system, in an experimental setting that models natural interaction. The previous studies (see below) examined these questions, albeit in a limited interactive environment. By expanding the interactive abilities and opportunities of the participants, we hope to gain more insights into specific aspects of language evolution, but in a setting that mirrors more natural communicative practices. That is, we aim to observe the interactive strategies that may be co-opted to facilitate alignment to form-meaning matches in a novel communicative system. With a goal-based paradigm, we hope to approximate a context in which communication is required to benefit individuals, and in which negotiation of new form-meaning matches is required.

1.1. Background

Human language learning and use are inherently interactive, however, the role of interaction in the emergence and conventionalization of new communication systems remains under-investigated in language evolution research. Experimentalists (Kirby, Cornish, & Smith, 2008) have demonstrated in the laboratory how languages evolve compositional structures as they are transmitted culturally through iterated learning from one individual to another,

but only recently have others explored the effect of interaction. Fay *et al.* (2010) extend the iterated learning paradigm to social collaboration in using a graphical sign system to convey abstract concepts. The graphical representations become less iconic through iterated participant use, while task accuracy shows the systems were successful in both the dyadic and community groupings. They conclude interaction was a crucial element to the creation of a new, shared communication system between individuals and communities of individuals.

Similarly, Healey *et al.* (2007) incorporate concurrent communication in a graphical drawing task, allowing participants' "mutual modifiability" of one another's outputs. They find an accumulative history of interaction contributes to lineage specific conventions, supporting the hypothesis of a conventionalized communication system emerging from interaction opportunities. Interaction provided for the use of devices that supported coordination (agreement/disagreement) and served editing functions that identified specific aspects of the pairs' drawings, which facilitated alignment to meaningful elements of the representations. Healey and colleagues suggest a repair-driven co-ordination process in communication could likely be observed in emerging gestural systems, a type of interaction that may involve more use of eye gaze, the face, and the body than graphical systems tasks.

By incorporating the study of repair (as in conversation analysis: Jefferson, 1974; Clark & Krych, 2004) into an interaction-based iterated learning paradigm, we hope to demonstrate how certain aspects of interaction facilitate the conventionalization of a gestural communication system and the effect of repair on such a system.

2. Methodology

In an experimental setting, participants (English-speaking, right-handed, non-signers) engaged in an interactive communication game aimed at producing an evolving communication system through negotiation over simulated generations via a gradual turn-over of participants. Participant dyads were required to interact with one another during a card selection task, alternating between Director and Matcher roles. However, their communication was limited to silent gestures (Goldin-Meadow *et al.*, 2008; Schouwstra, 2012), a communication similar to "Charades" in which the participants can use their hands/arm, as well as the face, to communicate without vocalizations. An Observer watched the dyad interact through one "generation," and then participated in the subsequent.

2.1. Stimuli

The stimuli for this experiment were two sets of similar-meaning cards. One set was comprised of verb meanings (with the progressive “-ing”), while the other was comprised of noun meanings (with the indefinite article “a” or null article). However, the verb-noun pairs had minimally contrastive features when gestured with speech (e.g. spontaneous co-speech gesture of “hammering” and “a hammer” are typically produced with a closed fist moving up and down on a real or imagined surface). A total of 32 noun-verb pairs comprise the set (for a total of 64 target tokens). The verb cards were represented in the progressive form (“-ing” suffix), while the noun cards adopted the indefinite article “a” or a null article (as with “snow”). Similar-semantic distractor tokens that contributed to task difficulty and masked the task objective (i.e. some did not conform to the grammatical forms of the target noun-verb pairs) were included as well.

2.2 Procedure

Dyads, using only silent gestures, played an interactive card selection task for four rounds per simulated generation. Each participant in the dyad had a set of cards “in hand” (n=10, 8 target matches, and two similar-semantic distractors) which could only be viewed by the holder. For each round, the target set of noun-verb pairs were grouped for similar semantics (e.g. “shovel” and “rake” cards were in the same round), but were distributed between the participants randomly (for noun-verb pairs, as well as for the number of nouns or verbs in each participants’ set). Participants took alternating turns requesting one card at a time from their partner’s cards-in-hand with silent gesture.

Participants were placed in one of two conditions. In the “Standard” condition, each participant had one opportunity per turn to make a request for the target card. Once the gesture was performed (note that the gesture could be performed or changed any number of times before the Matcher had to provide a guess), the Matcher provided a card. If the card was a correct match, the participants placed it to the side (a “matched matrix”); however, if the card was incorrect, the Matcher returned the card to their in-hand board, and the Director placed the target card to the bottom of their set. The Director would have another opportunity to gesture the target card later in the round. The “Do-Over” condition proceeded in the same manner, except in the instance of an incorrect guess from the Matcher: the Director had an immediate turn for a “do-over” in which they could once again gesture the target meaning. This condition was meant to drive immediate repair and reflect how repair is performed in natural conversation. The hypothesis is that more immediate opportunities to repair would result in more rapid alignment to corrected gesture strategies.

3. Results

Here we present the results collected from three chains (of 4 generations) in each condition.

3.1 Noun Marking Strategies and Handshape

Over generations, participants developed gestures to disambiguate nouns and verbs, focusing especially on nouns. Marking systems were innovated and passed down generations, though not all chains used the same system. The three most common, and consistent, noun markers were “object point” (O-Point, one hand points to the other performing the noun-related gesture), “object emphasis” (O-Emh, in which the emphasis could denote shape, handheld-ness, etc), and use of the “index finger” held up like the numeral “one.” Some generations fluctuated between the use of markers, but eventually one strategy dominated. In addition, each marker conformed to its own word order which was also innovated and recognized by the participants. Nearly all chains made use of a noun marking system to disambiguate noun-verb pairs, and by the final generation, many chains used the noun marker reliably.

Do-Over chains exhibited less variability in marker use, namely in Chains 1 and 2, while Chain 3 showed more alignment to a dominant marker in Generation 3. All Standard chains had non-dominant use of markers initially, wherein Chains 2 and 3 displayed a switch in the marking strategy that eventually overtook the other form in use, occurring in generations 2 and 3, respectively. Overall, the Standard condition marked targets more than the Do-Over condition. The Standard condition participants marked more nouns than were left unmarked, and while the primary overall strategy was pointing to the imagined object, each chain exhibited different dominant strategies.

The semantic category to which a noun belonged influenced how its gestural structure manifested. Instrument nouns (23 of 32 noun targets, such as “A Hammer” or “A Shovel”) were most often gestured as “Handled,” that is, the participant held the imagined object in their hand as they performed the action related to the instrument. This is an action-variant (Ortega et al, 2012), by which an object is represented by the action associated with it. A notable exception is “A Phone,” which is most frequently embodied as a hand in the shape of a phone. Overall, there was a trend to use action-variants (“Handled”) over object/perceptual variants (“Embodied”) for noun targets.

3.2 Accuracy and Timing

Accuracy in the Do-Over condition remained similar over generations (an average of 85%), while accuracy in the Standard condition increased from 75%

to 87%. Accuracy spiked most drastically from Generation 1 to Generation 2, in both conditions. We might attribute this trend to the role of the Observer, who comes into the interacting dyad as a knowing participant having seen the previous dyad perform the task. Using the average accuracy of guesses, each condition demonstrated higher average accuracy when guessing noun-targets with the Index Finger marker. The average accuracy using the Index Finger marker was also higher than not using a marker at all; though, O-Emph and O-Point markers resulted in similar accuracy as not marking at all (with O-Point being slightly more advantageous).

The length of gestures did not vary greatly, with the exception of repair sequence gestures. Overall, gestures were between 2 and 4 seconds, though, on average, Do-Over condition gestures were shorter than Standard condition ones for all generations. Though the general trend of decreased gesture length, by seconds, appeared trivial, within the context of this study, minute changes have large effects. In general, as gestures were transmitted through generations, their length shortened, indicating the potential constricting of the gesture space and the diminished need to elaborate upon initial gestures.

3.3 Repair

All three repair strategies (Repetition, Clarification, and Reformulation) were present in the two conditions, though to differing degrees. Clarifications, or moderate modifications on the initial (trouble-source) gesture, were the most frequent strategy in either condition (Standard=52, Do-Over=63). A clarification repair required the Director to highlight or emphasize an element of the initial gesture, typically done through the shaking of a hand (in “handled” gestures) or emphasizing a marker. Reformulation repairs are complete modifications of the initial gesture, often involving a change in handshape. Reformulations were more frequent in the Standard condition (n=23) than the Do-Over condition (n=9), which may be related to the inability to perform a repair after an incorrect guess, thereby requiring Directors to re-do gestures without knowing what the trouble source or misinterpretation is. Repetition, either full or partial (but crucially not involving emphasis), was similarly used in both conditions (Standard=21, Do-Over=20), and required no modification to the initial gesture.

The evolution of repair strategies over simulated time can be an index for the degree of conventionalization the communication system has achieved. While the amount of repair does not decrease with each generation (as in Fay *et al.* (2010)), it may still be indicative of the conventionalization process in which repairs conform to prior preferences that are then re-constituted through the repair itself. This study evidences the process of conventionalization in the

repair sequences performed by the Directors themselves (self-repair) or those provided by the Matcher (other-repair). Other-initiated repair is an attempt to get the Director to align with the Matcher's conceptions of the silent gesture system. When a Matcher initiates repair with facial gestures (e.g. furrowed brow) or returned eye gaze to the Director, they communicate a misunderstanding of the prior, and attempt to make the Director repair to a more communicable gesture. If, on the other hand, the Matcher performs a repair (i.e. gives the repair options), they are demonstrating a more effective strategy which may then be taken up in subsequent gestures. Both repair sequence types bring about a saliency in the gesture form (and even order) that allows for participants to fixate on more readily transmittable and communicative form-meaning matches.

The participants' repair strategies also played a role in the conventionalization of noun marking systems. Clarification repairs were the most frequently used repair strategies, particularly with noun targets. In the Do-Over condition, the immediate "repair" turn allowed for more repair use, in which we witness the Director's clarification strategy largely involving the use of a marker. Since clarifications require emphasizing or highlighting a previously gestured feature, greater saliency of those features emerges and becomes a substrate from which participants can build meanings, including compound meanings. Over generations, participants develop conventionalized marking systems from the emphasized forms. Repair, a feature of interaction, does facilitate the systematization of the gestures.

4. Discussion

As generations modify the previous generation's output, descriptive gestures become lexicalized such that an identifiable and systematic marking system develops in all chains. Typically, a marking gesture emerges from the need to disambiguate noun-verb pairs following evidence of inaccurate guesses or confusion. The ubiquity of the marking systems in this experiment demonstrates its efficiency in communicating meanings, and, specifically, in disambiguating nouns from verbs. The immediate opportunity to modify the previous gesture (or gesture space) leads to repair-driven markers that conventionalize within a chain. Clarification repairs could be driving the gestural forms that are conventionalized; as the need to clarify a target meaning arises, and repair is performed, the strategy for repair becomes part of the gestural system. Here, the repair strategy "clarification" often results in an emphasis on the object-ness of the target word, including its handheld-ness, shape, or singularity. Marker clarifications are common, and their saliency could drive the marking system into conventionalization.

4.1 The Case for Interaction

The crux of this study is to demonstrate that interaction, namely face-to-face contingent negotiation, has a role in the emergence and evolution of communication systems. Using an iterated learning paradigm, we can simulate language evolution over generations, but incorporating dyadic interaction with an observer highlights the interactive component of language use and transmission. Both conditions followed this model, except that one sought to mimic natural conversation in terms of turn taking in repair sequences. Though some differences existed between the conditions, here we are concerned with the broader role of interaction in facilitating conventionalization.

The modified iterated learning model in this study derived from earlier studies of iterate learning of alien languages, though many did not provide contingent interaction scenarios. Kirby, Cornish, and Smith (2008) found that structure emerged between generations 4 and 9, with variability by chain. These interactions took place in computer-modulated communication, not face-to-face interaction. Nonetheless, a compositional structure emerged that allowed for motion, color, and shape to be encoded by different morphemes that could be strung together to encode meanings such as “black square moving in a circle.” As noted, some chains attained their maximum level of structure by generation 4, while others did not until generation 9; often early-structure chains collapsed again into less-structured use of the language.

Face-to-face interaction, on the other hand, may promote quicker and more efficient transmission of the newly emerged language structure. Fay *et al.* (2010)’s graphical communication task saw isolated pairs reduce drawing complexity through round 3 (an approximate to “generation 3”); however, community pairs took longer to achieve less complexity in drawings (note: simple drawings require less graphical refinement but still convey meaning). The second condition here, though, is more akin to speakers of different languages – or dialects – coming together to communicate. The current study, in contrast, mimics the language learning environment of novices, including children and foreign language learners.

In a more naturalistic setting, Master, Schumann, and Sokolik's (1989) experimental creation of Persian and German pidgins required participants to extend their pidgin lexicon to new meanings. These systems displayed compositionality, regularity, and stabilized compound noun forms by their fourth use. The rapidity of this systematization may have resulted from the extensive negotiation that took place when innovating new lexical items from a given set. In fact, just as with Fay *et al.*, here noun compounds simplified by 66 % (for example, an initial noun compound may have been comprised of 6 individual

tokens, strung together, but after negotiation, decreased to about 3 tokens). It should be noted that this happened at the community level, as pairs exchanged partners in a gradual-turnover fashion, similar to the generational changes incorporated in our study.

In sum, face-to-face interaction might allow for more efficient and reliable transmission over generations. Interacting dyads, and groups to some extent, reach stabilization earlier than some non-interactive conditions, and may lead to the emergence and fixation of many features of systematicity simultaneously (being promoted by negotiation of varied aspects of meaning through moment-to-moment shifts and alignments). Furthermore, allowing for interaction that more closely approximates language use practices (e.g. turn-taking patterns, use of facial gestures, etc), which may be present even pre-linguistically, is essential to understanding how potential communicators rely on interaction structures to facilitate the conventionalization of linguistic systems. Teaching and learning are naturally interactive activities, therefore studies of the cultural transmission of language should consider how features of interaction impact the trajectory, in terms of evolving systematicity, of those languages.

References

- Clark, H.H., & Krych, M.A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1), 62–81.
- Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive science*, 34(3), 351–86.
- Goldin-Meadow, S. et al (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (27), 9163-9168.
- Healey, P.G.T., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: interactional constraints on representational form. *Cognitive science*, 31(2), 285–309.
- Jefferson, G. (1974). Error Correction as an Interactional Resource. *Language in Society*, 2, 181-199.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(31), 10681–6.
- Master, P., Schumann, J., & Sokolik, M.E. (1989). The experimental creation of a pidgin language. *Journal of Pidgin and Creole Languages*, 4(1), 37-63.
- Ortega, G., Sümer, B., & Özyürek, A. (2012). Type of iconicity matters: Bias for action-based signs in sign language acquisition, 1114–1119.
- Schouwstra, M. (2012), *Semantic Structures, Communicative Strategies, and the Emergence of Language*. PhD Thesis, University of Amsterdam.

HOW SELECTION FOR LANGUAGE COULD DISTORT THE DYNAMICS OF HUMAN EVOLUTION

W. GARRETT MITCHENER

*Mathematics Department, College of Charleston
Charleston, SC, USA
mitchenerg@cofc.edu*

The human language faculty is supported by a gene regulatory network that influences the development and operation of the motor, sensory, and cognitive systems. Consequently, this network must be fairly large and complex, and probably includes genes scattered throughout the genome. It is under selective pressure to continue functioning. Many gene loci are therefore in linkage disequilibrium with some element of this network, potentially disrupting their evolutionary dynamics. In the interest of exploring how significant this effect could be, we consider artificial life simulations, in which agents are required to perform an information coding task analogous to the replay of a memorized gesture. The task requires a network of interacting genes. The population is then branched, and phylogenetic trees are constructed based on genetic distances between leaf populations. Distances are determined by comparing genes for a simple task unrelated to the coding task. The process is then repeated without the coding task. Sample runs from these two simulations are compared to data from a neutral drift model. Both a-life simulations result in lower edge weights than the neutral model, giving the appearance of a slower molecular clock. Furthermore, the simulation with the coding task shows even lower edge weights, even though the mutation rate is the same. Therefore, the presence of a large gene network such as the one for language could distort the evolutionary trajectories of unrelated genes.

1. Introduction

The neutral theory of genetic drift is based on the assumption that most variations of a gene within a species are neutral mutants (Kimura, 1984; Ohta, 1992; Nowak, 2006). Beneficial mutations should spread and reach fixation in a process called a selective sweep. Deleterious mutations should die out quickly. Mathematically, each neutrally drifting site in a gene mutates according to a Poisson process. This model is the basis of the *molecular clock*, which asserts that the genetic distance between species should be related in a straightforward way to the time since they diverged. However, biological data is often inconsistent with this model. Similar genes in different species seem to have very different mutation rates, as do different genes within the same species, and sometimes even different sites within the same gene. Often there is no clear explanation for these discrepancies (Ayala, 1999; Cutler, 2000). Suggested explanations include the possibilities that the natural mutation rate varies in unexpected ways, that finite population size effects are

surprisingly significant, and that weak selection may be complicating the dynamics. This third option is the subject of the computational experiments described here.

Selection acting on one set of genes can disturb the dynamics of others, even those that perform unrelated functions. Selective sweeps and hitchhiking (a form of linkage disequilibrium) are known to create such complications, but so could the tangled nature of gene regulatory networks. For example, selective breeding for friendliness during domestication of animals affects many other characteristics, apparently because it places pressure on the timing of the entire development process (Trut, 1999).

Once a selectively favored genetic network is established, purifying selection acts on it, meaning that any mutation causing it to malfunction will die out quickly, which systematically destroys a small amount of genetic diversity. If the network is sufficiently large and performs a highly favored function, there are many possible deleterious mutations. Consequently, purifying selection could reduce genetic diversity enough to give the illusion that the mutation rate is lower than it actually is, even in an unrelated gene, thereby distorting the molecular clock.

The human language faculty combines motor, sensory, and cognitive machinery, which suggests that it is generated by a large gene regulatory network. Supporting molecular evidence comes from the well known speech-related gene FOXP2, which regulates many other genes (Enard et al., 2002). Pressures on other language-related genes may therefore have affected evolutionary dynamics throughout the human genome.

There are statistical tests for selective pressure at the molecular level, usually expressed as whether the data is sufficient to reject a null hypothesis built from the neutral theory, but they do not indicate which aspect of the neutral theory fails to apply and are generally controversial (Nielsen, 2005). They are also not the appropriate tool here because the phenomenon of interest is not whether a specific gene has recently experienced selection, but generally how selection on a large network might affect the evolution of an unrelated gene as a side effect.

In the interest of developing new tools for investigating how selection acting on a large network affects the rest of the genome, consider the following experiment. Begin with a root population of a single species, and place it under selective pressure to evolve and maintain a large gene regulatory network that performs some information processing task analogous to language, as well as genes that perform simpler mundane tasks. Allow the population to evolve a solution to the mundane task, and at least a partial solution to the information processing task. Then make multiple independent copies of this population, and allow them to continue evolving under the same conditions. After some time, branch the populations again. Then construct phylogenetic trees from the resulting leaf populations (subspecies) and compare their statistics with the known process by which they were created. Such trees are commonly reconstructed by examining genes for mundane

functions, such as dehydrogenases, assuming they fall under the neutral model.

Although this experiment is very difficult to conduct physically, it is straightforward to conduct computationally using an artificial life or *a-life* simulation. Essential details of the simulation are presented in Section 2. More details are available in (Mitchener, 2014). The source code and configuration will be available at the author's web site mitchenerg.people.cofc.edu. Section 3 describes the phylogenetic experiment and key results. These are discussed in Section 4.

2. Artificial life simulation

2.1. Digital organisms

Each organism in the a-life simulation, called an agent, has a genome, common to all of its cells, consisting of bit strings that serve as chromosomes. Substrings of 78 bits are interpreted as instructions analogous to biochemical reactions. Instructions are executed in parallel in discrete time steps. Each cell within an agent has an internal state consisting of counts $A[p]$ of how many molecules of each type p are present. Molecules and reactions are abstract, and no attempt is made to simulate molecular structure or chemical bonds. Instead, each type of molecule is an integer called a *pattern*, and each instruction states that if the number of units $A[s]$ of a particular pattern s exceeds some threshold θ , then some units of pattern p are created by adding to $A[p]$, and some units of q are destroyed by subtracting from $A[q]$. A bit of input is provided to a cell by adding to $A[j]$ for a certain designated pattern j on time steps when the input bit is 1, but not when it is 0. Likewise, for each designated output pattern r , a cell generates an output bit of 1 if $A[r]$ exceeds some threshold and 0 otherwise. A cell can have many inputs and outputs. A synapse can be created by connecting an output bit from one cell to an input bit of another cell, thus forming a cellular network. A population of these agents subjected to a selection-mutation process can evolve the ability to solve computational problems.

2.2. Sequential coding and other tasks

The *information processing task* for this experiment is sequential coding and decoding. Each agent consists of two cells. One is the sender, and is given an input word consisting of two bits. The sender can send synaptic spikes to the second cell in the agent, the receiver. Two of the receiver's outputs are interpreted as a two-bit word, and the goal is for these to recreate the original input word. The receiver must also set an output called the *stop signal* to indicate when the calculation is complete. The sending cell is given input all at once, but can only transmit information through a narrow synapse, so it must do so over time. This task is analogous to replaying a gesture, such as speaking a particular sound. During speech, entire words and phrases are present in the mind all at once, but must

be replayed as muscle movements over time.

The *mundane task* is for the receiving cell to generate an additional output called a *beacon*. It indicates that the agent is basically alive, even if it isn't processing any information.

Each agent is given a rating based on how well it performs the above tasks. It earns a very large number of points by setting the beacon output. Thus, a mutation that damages the beacon is almost surely fatal. The agent is given the opportunity to transmit each possible input word, and earns many points for each bit in each input word that it correctly transmits. It earns extra points by stopping after fewer steps, which is the *timing task*. There is also a tiny penalty for using too many reactions, so overly complex mechanisms tend to get simplified.

2.3. Evolutionary dynamics

From each generation of 500 agents, 300 pairs of parents are chosen using tournament selection to produce 300 offspring. They are combined with the top-rated 200 agents of the previous generation to form a new generation of 500. Genomes are diploid, consisting of two pairs of chromosomes. When an agent reproduces, each pair of chromosome is aligned, a random crossover point is selected between genes, and a single recombined chromosome is produced, thus forming a gamete. Two gametes form a new agent. Chromosomes are subject to single-bit substitutions, whole-gene deletions, and whole-gene duplications.

The simulated population typically evolves a single-gene solution to the beacon task right away. Likewise, a single-gene timing mechanism usually develops early on, and enables agents to earn the extra points for stopping early. Solutions to the sequential coding problem evolve in steps. First, a mutation in some gene results in an instruction that is sensitive to one input bit and also activates the synapse in the sender. Another gene must be discovered that reads the synapse and links to the correct output bits. Once that basic connection is in place, gene duplications and other mutations form a link from the other input bit to the synapse in the sender, and from the synapse to the other output bit in the receiver. Typically, the receiver is able to determine which bits it should set from the time at which the sender begins spiking. The sender activates the synapse early if one bit is set, late if the other is set, and very early if both are set. This is usually enough information for the receiver to reconstruct everything except for one bit in one input word. To save time, sample runs used in this article were stopped at this point. (Reaching this stage took 13,000 to 100,000 generations over many hours of computer time. Given much more time, an additional mechanism evolves to handle that one last bit.) The result is a network of 10 to 20 genes that solves all but one part of the coding problem. Once a population reaches this stage, selective pressures on it are dominated by maintaining and possibly simplifying the encoding, decoding, timing, and beacon mechanisms.

There are good reasons to work with such a complicated simulation. The

question at hand is whether a large, selectively favored gene network scattered throughout the genome (such as the one supporting the language faculty) could experience sufficient purifying selection to distort the evolutionary trajectories of unrelated genes. Any model that can address this question must account for how a subset of a genome specifies a network, how that network benefits the organism, and which mutations are beneficial, neutral, and deleterious. An a-life simulation of the complexity described here is a reasonable place to start.

3. Phylogenetic experiment

3.1. Coding problem samples

Begin with a population of 500 agents with random genomes. Let them evolve as described until the first agent capable of transmitting all but one bit dies, which means that a successful mechanism has saturated the population. The population is allowed to continue for 1000 more generations to ensure that the selective sweep has run its course. This is the *root population*. The genomes from this population are copied into four separate populations which are then run independently for 400 more generations each. Those populations are each branched into four and run for 400 more generations. The result is a population tree with four groups of four (4×4) *leaf populations*.

3.2. Mundane problem samples

A second set of leaf populations is created in the same way, except that no points are given for the coding task. Only the timing and beacon tasks affect these populations. Since these mechanisms evolve quickly, these populations are continued for 30,000 more generations before branching so that their history is of the same order of magnitude as that of the coding-problem runs. This extra time is necessary, otherwise some of the phylogenetic tree statistics described below come out artificially different between the coding-problem runs and the mundane-problem runs.

3.3. Genetic distance between a-life populations

The raw distance between two leaf populations from the a-life simulation is the average genetic distance between all maximally rated individuals in each population. Most agents in a leaf population achieve the same high rating, but a few have a lower rating due to deleterious mutations and are ignored. The genetic distance between two individuals is the minimum Hamming distance between beacon genes in each. The genetic distance is divided by 78, the length of a gene, to give a fractional distance p , which is converted to a final distance $-\ln(1 - p)$. This final step is called the Poisson correction (Nei & Kumar, 2000).

3.4. Neutral model samples

As an additional control, a third set of population trees is constructed using a purely neutral process. Each of 1000 individuals consists of a string of 48 bits. To build the next generation, 600 individuals are selected and subject to single bit mutations. The 400 youngest individuals are kept for the next generation. An initial population of strings of all zeros is run for 30,000 generations to match the time required by the coding-problem sample runs, then branched into four copies and run for 400 generations. Each of those is branched into four copies and run for 400 generations, yielding the same population tree structure as the a-life simulation. The distance between pairs of these populations is the average Hamming distance between the bit strings therein divided by 78, then Poisson corrected.

Some explanation is in order for the configuration choices in the neutral simulation. Genes in the a-life simulation consist of 78 bits, 18 of which specify the pattern p such that $A[p]$ is increased, and 12 of which specify the instruction's threshold. That leaves up to 48 bits in the beacon gene that could potentially experience neutral drift. If the neutral simulation is run with strings of 78 unconstrained bits, the distances between populations are proportionally larger, which turns out to be undesirable. Also, the neutral simulation is run on a population of 1000 strings, which is one for each allele in the diploid a-life simulations.

3.5. Phylogenetic reconstruction

Each set of 4×4 leaf populations is analyzed as follows. Each leaf population is paired with each other leaf population that shares its parent. Each such pair is paired with each other pair with the same root. This gives a total of 216 quartet problems, that is, four leaf populations that must be assembled into an un-rooted tree with two internal nodes. The correct branching structure is known. Each branch is then given a non-negative weight so that the total weight along the path from one leaf to another is approximately their genetic distance. Specifically, weight assignments minimize the sum of squares of differences between tree path distance and genetic distance. This method of assigning weights is a standard technique in phylogenetics (Nei & Kumar, 2000).

According to the neutral theory, the distance between two leaf populations should be roughly proportional to twice the time since they branched. The actual time between branches is always 400 generations, and the mutation rate is fixed across all runs. Thus, all phylogenetic trees from all three simulations should ideally have weight w on the edges to the leaves, and $2w$ on the central edge, and all three simulations should result in similar values of w . The central edge should be weighted twice as much because the tree does not include a node for the root population. We therefore consider the following statistics. Given a quartet tree with weights as shown in Figure 1, the *discrepancy* from an ideal tree with

branches weighted w is

$$R(w)^2 = (w_{AX} - w)^2 + (w_{BX} - w)^2 + (w_{CY} - w)^2 + (w_{DY} - w)^2 + (w_{XY} - 2w)^2$$

The optimal w and corresponding R are

$$w^* = \underset{w}{\operatorname{argmin}} R(w) = \frac{1}{8} (w_{AX} + w_{BX} + 2w_{XY} + w_{CY} + w_{DY})$$

$$R^* = R(w^*)$$

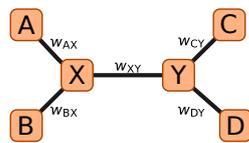


Figure 1. General un-rooted phylogenetic quartet tree for four leaf populations (A, B, C, D) with two ancestral populations (X, Y).

Twenty samples of each of the three models (neutral, mundane, coding) were run, all 216 weighted trees were constructed, and w^* and R^* were calculated (Figure 2). The charts show quartiles for the numbers gathered from each sample run, but it should be noted that numbers from the same run (within each bar) are not entirely independent because they are calculated from phylogenetic trees that can share leaves. Statistics calculated from different runs are independent, however. Weighted trees constructed from biological and simulation data are often not in ideal proportions (Nei & Kumar, 2000, p. 79), so the range of discrepancy values seen in Figure 2 is expected.

It is clear that w^* and R^* are distinctly larger for the neutral sample runs than the a-life sample runs. There is also a distinction in w^* between the a-life models. The Mann-Whitney-Wilcoxon rank sum test applied to sets of median w^* values from each of the sample runs from the mundane problem and coding problem simulations yields a p -value of 0.0154. That is, the difference in the median value of w^* derived from the mundane problem of 0.0245 and the value derived from the coding problem of 0.0201 is statistically significant. Consequently, the weak selectional forces acting directly and indirectly on the mundane task gene in the mundane-only and coding task simulations are not negligible when assigning weights to phylogenetic trees.

4. Discussion and conclusion

When reconstructing weighted phylogenetic trees of the great apes, the human-specific data has several odd features. When mitochondrial DNA is used to estimate genetic distances, a frequently seen anomaly is that the path from root to

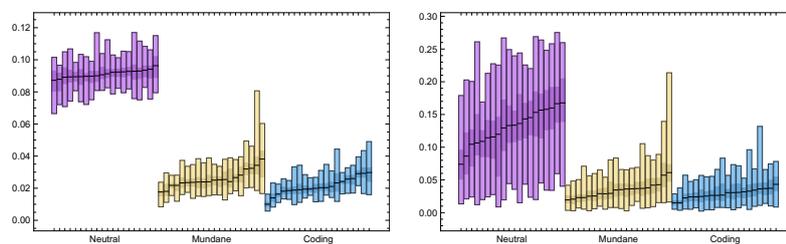


Figure 2. Charts for all sample runs under each model. Each bar ranges over all values from all weighted trees from one sample run. Stripes indicate quartiles. Black lines indicate medians. Runs from each model are ordered by median for easier comparison, and are not in the same order in each chart. Left: w^* ; Right: R^* .

human is slightly shorter (smaller total weight) than paths to the other species, though it is not necessarily statistically significant (Nei & Kumar, 2000; Glazko & Nei, 2003). A related puzzle is that compared to the other great ape species, humans show a distinctly low level of genetic diversity (Deinard & Kidd, 1999). Many forces potentially underlie these anomalies: random noise, differences between mitochondrial and nuclear DNA, natural polymorphism, the fact that speciation is not instantaneous, founder effects due a bottleneck in the history of the human population (Deinard & Kidd, 1999), and a gradual increase in the time between generations (Scally et al., 2012).

The simulation presented here shows that even if the mutation rate, population size, and branch points are fixed and idealized, the phylogenetic trees can be distorted if there is a relatively large and important gene network. Using the neutral model as a baseline, the low edge weights on the mundane-problem and coding-problem trees would underestimate the mutation rate or the time between leaf populations and the first branches. (There is also evidence that phylogenetic methods can significantly overestimate mutation rates (Scally & Durbin, 2012).) Coding-problem trees had even lower edge weights overall than mundane-problem trees. Thus, the selective pressure maintaining a large gene network suffices to distort the molecular clock of unrelated genes.

The relatively low genetic diversity of humans may be partly due to this force. Mitochondrial DNA would be likewise affected. The initial formation of a large, selectively favored gene network, followed by a selective sweep, would contribute to the hypothetical bottleneck and associated founder effects. The language faculty is an obvious candidate for the phenotype of such a network. It is not yet clear how to determine the extent to which this large-network effect might be responsible for anomalies in human genetics, compared to disease, climate, and other non-genetic forces. However, the simulations described here imply that selective pressure on the language faculty, even just to maintain its function, could have significantly altered the evolutionary dynamics of many other human genes.

References

- Ayala, F. J. (1999). Molecular clock mirages. *BioEssays*, 21(1), 71–75.
- Cutler, D. J. (2000). The Index of dispersion of molecular evolution: slow fluctuations. *Theoretical Population Biology*, 57(2), 177–186.
- Deinard, A., & Kidd, K. (1999). Evolution of a HOXB6 intergenic region within the great apes and humans. *Journal of Human Evolution*, 36(6), 687–703.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P., & Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418(6900), 869–872.
- Glazko, G. V., & Nei, M. (2003). Estimation of Divergence Times for Major Lineages of Primate Species. *Molecular Biology and Evolution*, 20(3), 424–434.
- Kimura, M. (1984). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
- Mitchener, W. G. (2014). Evolution of communication protocols using an artificial regulatory network. *Artificial Life*, 20(4), 491–530.
- Nei, M., & Kumar, S. (2000). *Molecular evolution and phylogenetics*. Oxford: Oxford University Press.
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review Of Genetics*, 39, 197–218.
- Nowak, M. A. (2006). *Evolutionary Dynamics: Exploring the equations of life*. Cambridge, MA: Harvard University Press.
- Ohta, T. (1992). The Nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, 23, 263–286.
- Scally, A., & Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. *Nature Reviews Genetics*, 13(10), 745–753.
- Scally, A., Dutheil, J. Y., Hillier, L. W., Jordan, G. E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., McCarthy, S., Montgomery, S. H., Schwalie, P. C., Tang, Y. A., Ward, M. C., Xue, Y., Yngvadottir, B., Alkan, C., Andersen, L. N., Ayub, Q., Ball, E. V., Beal, K., Bradley, B. J., Chen, Y., Clee, C. M., Fitzgerald, S., Graves, T. A., Gu, Y., Heath, P., Heger, A., Karakoc, E., Kolb-Kokocinski, A., Laird, G. K., Lunter, G., Meader, S., Mort, M., Mullikin, J. C., Munch, K., O'Connor, T. D., Phillips, A. D., Prado-Martinez, J., Rogers, A. S., Sajjadian, S., Schmidt, D., Shaw, K., Simpson, J. T., Stenson, P. D., Turner, D. J., Vigilant, L., Vilella, A. J., Whitener, W., Zhu, B., Cooper, D. N., Jong, P. de, Dermitzakis, E. T., Eichler, E. E., Flicek, P., Goldman, N., Mundy, N. I., Ning, Z., Odom, D. T., Ponting, C. P., Quail, M. A., Ryder, O. A., Searle, S. M., Warren, W. C., Wilson, R. K., Schierup, M. H., Rogers, J., Tyler-Smith, C., & Durbin, R. (2012). Insights into hominid evolution from the

gorilla genome sequence. *Nature*, 483(7388), 169–175.

Trut, L. (1999). Early Canid Domestication: The Farm-Fox Experiment. *American Scientist*, 87(2), 160.

FREQUENCY-DEPENDENT REGULARIZATION IN ITERATED LEARNING

EMILY MORGAN AND ROGER LEVY

Linguistics, UC San Diego

San Diego, CA

{eimorgan,rlevy}@ucsd.edu

Binomial expressions are more *regularized*—their ordering preferences (e.g. “bread and butter” vs. “butter and bread”) are more extreme—the higher their frequency. Although standard iterated-learning models of language evolution can encode overall regularization biases, the stationary distributions in these standard models do not exhibit a relationship between expression frequency and regularization. Here we show that introducing a frequency-*independent* regularization bias into the data-generation stage of a 2-Alternative Iterated Learning Model yields frequency-*dependent* regularization in the stationary distribution. We also show that this model accounts for the distribution of binomial ordering preferences seen in corpus data.

1. Introduction

Languages are shaped both by the cognitive architectures of individual speakers and by the process of cultural transmission that acts across generations. In this paper we ask how these two factors jointly contribute to a key dichotomy in language structure: the trade-off between broadly-applicable compositional knowledge and knowledge of item-specific idiosyncrasies. Specifically, we take up the case of frequency dependence in *regularization*—the consistency of a preference for a given form among multiple alternatives.^a Although regularization is a well-attested phenomenon in statistical learning, *frequency-dependent* regularization is not. Here we demonstrate that frequency dependence of regularization can arise as an emergent property of a frequency-*independent* regularization bias in language production, combined with the bottleneck effect of cultural transmission.

Item-specific idiosyncrasies (i.e. exceptions to the rules) are well known to be frequency-dependent. For example, more frequent verbs are more likely to have irregular conjugations (Lieberman, Michel, Jackson, Tang, & Nowak, 2007). More recently, we (Morgan & Levy, 2015) have demonstrated a different type

^aFollowing previous evolutionary linguistics literature, particularly Reali and Griffiths (2009), we define *regularization* as reduction in entropy of a distribution, i.e. reduction in variation. We note that this is different from the notion of “regular” items as those that conform to compositional rules.

of frequency-dependent idiosyncrasy at the level of multi-word phrases, specifically *binomial expressions* of the form “X and Y” (Cooper & Ross, 1975; Benor & Levy, 2006). Word order preferences for these expressions are gradient; for example, “radio and television” is preferred to “television and radio” in a 63 to 37 ratio, while “bread and butter” is preferred to “butter and bread” 99 to 1 (Lin et al., 2012). These ordering preferences are partially compositional, determined by productive, violable constraints, e.g. a constraint to put shorter words before longer words. But these expressions are also subject to learned item-specific idiosyncrasies, e.g. despite a generally strong constraint for words referring to males to precede words referring to females, “ladies and gentlemen” is preferred over “gentlemen and ladies”. In addition to the possibility of the complete reversal of compositional preferences, item-specific idiosyncrasies can also be gradient, e.g. a binomial whose compositional preference predicts a 60/40 distribution might instead be used in a 90/10 ratio. Morgan and Levy (2015) showed that, as is the case with irregular verbs, the distribution of idiosyncrasies in binomial ordering preference is frequency-dependent: more frequent binomial expressions deviate more from compositional preferences. In particular, more frequent binomials are more strongly regularized.

Regularization is a well-established phenomenon in statistical learning. In a variety of tasks, both linguistic and non-linguistic, in which participants learn and reproduce probability distributions over alternates, both children and adults tend to regularize their productions (Hudson Kam & Newport, 2005; Reali & Griffiths, 2009; Ferdinand, Kirby, & Smith, 2014). For example, Reali and Griffiths (2009) found that when exposed to two labels for a novel object, subjects on average reproduced the more frequent label *even more frequently* than that label was seen in training. Although this tendency was weak, they demonstrated that even such a small bias towards regularization can have significant long-term impacts, as the bias acts across successive generations to shape language over time. Bickerton (1981), Hudson Kam and Newport (2005), and others have argued that children’s tendency to regularize is an important mechanism of language change, e.g. for forming more consistent languages out of pidgins.

However, standard iterated-learning theories of language evolution do not, in general, lead to frequency-dependent regularization. Thus our finding in Morgan and Levy (2015) is unexpected, and poses a challenge to models of language evolution. In this paper, we review the key data (Section 2) and show that standard iterated-learning models fail to account for frequency-dependent regularization (Section 3). We then show that frequency-dependent regularization emerges when the data-generation stage of a standard iterated learning model is augmented with a frequency-independent regularization bias, and that this augmented model accounts for the empirical distribution of binomial ordering preferences (Section 4). Section 5 concludes.

2. Dataset

We take advantage of a uniquely appropriate real-world data set: Morgan and Levy (2015)’s corpus of 594 binomial expression types hand-annotated for a range of semantic, phonological, and lexical constraints known to affect binomial ordering preferences, and with frequencies of each ordering extracted from the Google Books corpus (Lin et al., 2012). Morgan and Levy also reported a model estimating the quantitative compositional ordering preference for each binomial expression, as expected on the basis of the above constraints (independent of actual occurrence frequencies). The dataset and model thus give us three key measures for these expressions:

- The *overall (unordered) frequency* of an expression: $\text{freq}(\text{“X and Y”}) + \text{freq}(\text{“Y and X”})$
- The *observed preference* for occurrence in a given order, expressed as a number between 0 and 1: $\text{freq}(\text{“X and Y”}) / (\text{freq}(\text{“X and Y”}) + \text{freq}(\text{“Y and X”}))$
- The *compositional preference* for occurrence in a given order, expressed as a number between 0 and 1, given by Morgan and Levy’s model.

Observed preferences are multimodally distributed, with modes at the extremes as well as around 0.5 (Fig. 1a). Crucially, this pattern is not predicted by compositional preferences, which predict only a single mode (Fig. 1b). This pattern reflects the key generalization to be accounted for in the present paper: that expressions with higher overall frequency diverge most from compositional preferences, and are more regularized (Fig. 1c).

3. Regularization is Frequency-Independent in Standard Iterated Learning

We use 2-alternative iterated learning—specifically following Reali and Griffiths (2009) (see also Smith, 2009)—to simulate the evolution of binomial expressions over generations of speakers. A learner hears N tokens of a binomial expression, with x_1 of them in a given order—we use alphabetical order as a neutral reference order—and then infers a hypothesis $\theta_1 \in [0, 1]$ which is the proportion of time a binomial should be produced in alphabetical order. The learner then generates new data using θ_1 .

The prior probability $P(\theta_1)$ of a binomial being preferred in a given order can be expressed using the beta distribution. We can treat the compositional preference as a form of prior knowledge of ordering preferences for a binomial. To incorporate this prior knowledge, we use a parameterization of the beta distribution with a parameter μ that determines the mean of draws and a concentration parameter ν that determines how tightly clustered around the mean those draws are. (ν can also be thought of as reflecting how confident in the prior we are, e.g. $\nu = 10$ would indicate confidence equivalent to having seen ten instances of a

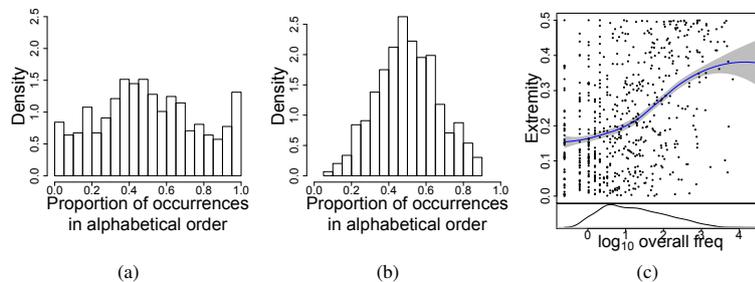


Figure 1. Results from Morgan and Levy (2015). (a) Histogram of binomial types' observed preferences. (b) Histogram of binomial types' compositional preferences. (c) We define an expression's *extremity* as the absolute difference between its observed preference and 0.5. More frequent expressions have more extreme/regularized preferences; see Morgan & Levy (2015) for alternative ways to quantify extremity that yield similar conclusions. Lower panel shows density of overall frequency counts (scaled as described in Section 4.2). The distribution is non-Zipfian because the corpus is restricted to binomial types with at least 1000 occurrences in the Google Books corpus to ensure accurate observed preference estimates.

given binomial expression type before.) Under this parameterization,

$$P(\theta_1) = \frac{\theta_1^{\mu\nu-1}(1-\theta_1)^{(1-\mu)\nu-1}}{B(\mu\nu, (1-\mu)\nu)} \quad (1)$$

where B is the beta function. Because μ represents compositional ordering preferences, it varies for each binomial, and is set according to Morgan and Levy's model. All learners are assumed to have the same μ value for a given binomial. ν is constant for all binomial expressions for all learners, and is a free parameter. Given θ_1 , data is generated binomially:

$$P(x_1|\theta_1) = \binom{N}{x_1} \theta_1^{x_1} (1-\theta_1)^{N-x_1} \quad (2)$$

We define a chain of learners under this model by initializing a single learner with some hypothesis. This first generation produces N utterances according to the distribution defined in Eq. 2. The learner in the next generation applies Bayes rule and chooses a hypothesis from the resulting posterior distribution over hypotheses. This process continues iteratively.

Reali and Griffiths (2009) have demonstrated that regularization occurs in iterated learning models with sparse priors (i.e. those that favor hypothesis close to 0 and 1); given our parameterization of the beta distribution, these are hypothesis with $\nu < 2$. However, this regularization is not dependent on the expression's overall frequency. We demonstrate this by modeling chains of learners with different values of N . We model a single binomial expression with prior probability

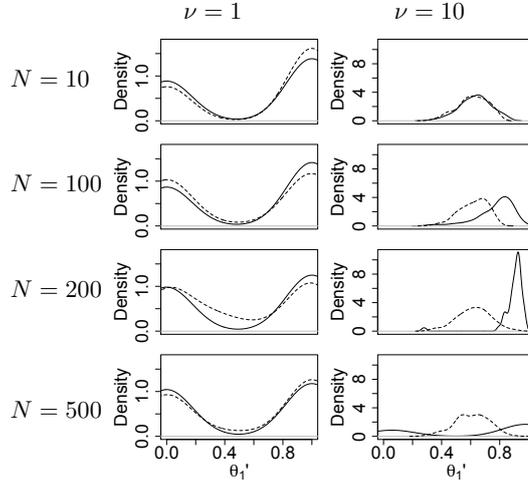


Figure 2. Simulated distribution of binomial ordering preferences for a single expression type with $\mu = 0.6$ in a standard 2-Alternative Iterated Learning Model (dotted lines) and one with an explicit regularization bias in data production of $R = 1.1$ (solid lines). Note that $\theta_1' = \theta_1$ in the standard model. Regularization depends upon N only in the model with an explicit regularization bias.

$\mu = 0.6$. We explore different values of ν , specifically $\nu = 1$ (a sparse prior) and $\nu = 10$ (a dense prior), and values of $N = 10, 100, 200, 500$. For each combination of ν and N , we approximate the distribution over expression preferences by running 100 chains of learners for 500 generations each and taking the hypothesis of the final generation in each chain, except in the $N = 500, \nu = 1$ case where chains are run for 1000 generations each because convergence to the stationary distribution is slower for higher values of N . (For all chains in all simulations in this paper, we initialize $\theta_1 = 0.5$ and use maximum a posteriori (MAP) estimation to choose θ_1 in each new generation. Results are qualitatively similar under posterior sampling.) Regularization in the resulting distributions does not depend on N (Fig. 2, dashed lines; the small apparent sensitivity to N for a given value of ν is due to the finite number of chains used in the simulations.) The number of times an expression is seen in each generation does not affect its ultimate degree of regularization.

4. Emergence of Frequency-Dependent Regularization in Iterated Learning

The standard 2-Alternative Iterated Learning Model does not predict frequency-dependent regularization. We now demonstrate that we can predict frequency-dependent regularization by introducing a frequency-independent regularization

bias into our model. Under this model, frequency-dependent regularization is an emergent property of the interaction of the frequency-independent regularization bias with the bottleneck effect of cultural transmission.

We augment the learning and transmission process as follows. After hearing data, the learner chooses a hypothesis θ_1 as before, then applies a regularization function to produce a new hypothesis θ'_1 , then generates data from θ'_1 .

The regularization function is the regularized incomplete beta function (equivalently, the cumulative distribution function of the beta distribution), restricted to be symmetric such that it has a single free parameter R :

$$f(x; R) = \frac{\int_0^x t^{R-1}(1-t)^{R-1} dt}{B(R, R)} \quad (3)$$

As shown in Fig. 3, the bias parameter R controls strength of regularization. When $R = 1$, this is the identity function, i.e. no explicit regularization is added, hence the standard model is a special case of the augmented model with $R = 1$. As R increases, the regularization bias grows stronger.

4.1. Results: Frequency-dependent regularization

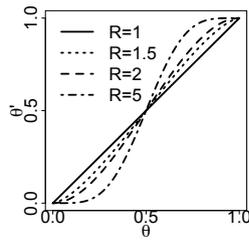


Figure 3. Regularization function with different values of R

When we repeat the simulations from above using a non-trivial regularization bias $R = 1.1$, we see frequency-dependent regularization in the case with a dense prior (Fig. 2). Although the regularization bias itself is frequency-independent, frequency-dependence emerges from the interaction of the regularization bias with the process of cultural transmission: At lower frequencies, there is not sufficient data for the regularization bias to overcome the prior. At higher frequencies, the regularization bias becomes increasingly dominant as

there is increasingly enough data for the effects of this bias to be carried across generations. Even a relatively weak bias ($R = 1.1$) can produce noticeable regularization when compounded across generations. However, the prior always continues to exert some influence; thus, even the highest frequency expressions do not become completely regularized.

Another linguistically accurate property of this model is that for sufficiently high values of N , the distribution over hypotheses includes a mode on the opposite side of 0.5 from the prior. Thus the model correctly predicts that at high enough frequencies, an expression can become idiosyncratically preferred in the opposite of its compositionally predicted direction (as in “ladies and gentlemen”).

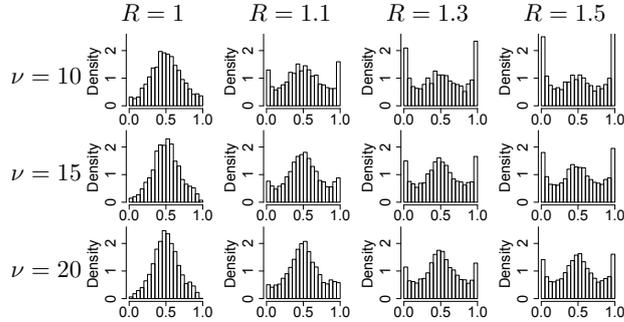


Figure 4. Predicted distribution of θ_1 . We see a trade-off between effects of the prior and the regularization bias. When the prior is stronger (high ν , low R), we see a unimodal distribution of preferences, similar to Fig. 1b. When the regularization bias is stronger (low ν , high R), we see too much regularization. At appropriate values of R and ν , we see the correct multimodal distribution of preferences as seen in corpus data (Fig. 1a).

4.2. Results: Simulating corpus data

Having demonstrated that our augmented model produces frequency-dependent regularization, we now show that it additionally predicts the true language-wide distribution of binomial preference strengths seen in corpus data. The target distribution to be accounted for is shown in Fig. 1a.

We take the true properties of each binomial expression in the corpus: its compositional preference determines μ and its overall frequency determines N . We scale overall frequency counts based on estimated lifetime exposure to 300 million total words (Levy, Fedorenko, Breen, & Gibson, 2012, footnote 10). The resulting distribution of values N is shown in Fig. 1c. For each binomial in the corpus, we approximate the stationary distribution by modeling 10 chains of learners for 200 generations each and taking the hypothesis θ'_1 of the final generation of each chain.

Our model has two free parameters, ν and R . We model the corpus data as described above for a range of values of both of these parameters. As shown in Fig. 4, our model displays a trade-off between the prior and the regularization bias as a function of these parameters. At appropriate values, our model correctly predicts the multimodal distribution of corpus data as seen in Fig. 1a.

5. Conclusion

We have demonstrated that a frequency-independent regularization bias in data generation, combined with cultural transmission, can produce the pattern of frequency-dependent regularization of binomial ordering preferences seen in corpus data. Cultural transmission creates frequency-dependence by introducing a

bottleneck effect (i.e. a limit on the number of tokens of a binomial seen by each generation) that favors prior knowledge at lower frequencies while allowing the regularization bias to be increasingly well transmitted at higher frequencies. This finding sheds light on the origins of linguistic structure in two important ways: one, it confirms earlier demonstrations of a bias to regularize when learning stochastic linguistic items. Second, it shows that this bias can apply equally across all levels of frequency, but that the distribution of idiosyncrasy seen in the language emerges from the interaction of individuals' cognitive biases with the bottleneck effect of cultural transmission. Additionally, we have expanded the empirical coverage of iterated learning models, showing that they can account for not only qualitative generalizations in natural language and data from laboratory experiments, but also detailed patterns of naturalistic corpus data. As we hope to have shown, binomial ordering preferences are a particularly suitable test case for iterated learning models, at once theoretically interesting, data-rich, and computationally tractable.

Acknowledgements

Thank you to Wednesday Bushong and Melodie Yen for assistance with preparation of the corpus data. We gratefully acknowledge support from the Alfred P. Sloan Foundation to Roger Levy.

References

- Benor, S., & Levy, R. (2006). The Chicken or the Egg? A Probabilistic Analysis of English Binomials. *Language*, 82(2), 233–278.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma.
- Cooper, W. E., & Ross, J. R. (1975). World Order. In R. E. Grossman, L. J. San, & T. J. Vance (Eds.), *Papers from the parasession on functionalism* (pp. 63–111). Chicago: Chicago Linguistics Society.
- Ferdinand, V., Kirby, S., & Smith, K. (2014). Regularization in language evolution. *Proceedings of the 10th International Evolution of Language Conference*.
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, 1(2), 151–195.
- Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, 122(1), 12–36.
- Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449(7163), 713–716.
- Lin, Y., Michel, J.-B., Aiden, E. L., Orwant, J., Brockman, W., & Petrov, S. (2012). Syntactic Annotations for the Google Books Ngram Corpus. *Pro-*

ceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 169–174.

- Morgan, E., & Levy, R. (2015). Modeling idiosyncratic preferences: How generative knowledge and expression frequency jointly determine language structure. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *37th Annual Meeting of the Cognitive Science Society* (pp. 1649–1654). Austin, TX: Cognitive Science Society.
- Real, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*(3), 317–328.
- Smith, K. (2009). Iterated learning in populations of Bayesian agents. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 697–702.

SELF-ORGANIZATION IN SOUND SYSTEMS: A MODEL OF SOUND STRINGS PROCESSING AGENTS

ROLAND MÜHLENBERND, JOHANNES WAHLE
Seminar für Sprachwissenschaft, University of Tübingen
Tübingen, Germany
{roland.muehlenbernd, johannes.wahle}@uni-tuebingen.de

Several typological universals of sound systems of human language are assumed to be a result of self-organization in a population's communication performance. This has been shown for human vowel system in diverse studies (c.f. de Boer, 2000b; Jäger, 2008). In these studies computational models were designed of agents communicating with single vowel sounds. In our study we present a computational model where agents communicate with concatenations of single sounds forming complex expressions. The goals of this study are i) to examine decisive factors that contribute to the emergence of realistic sound systems in artificial societies of interacting agents, and ii) to discuss ways to evaluate artificially emerged sound systems.

1. Introduction

Human sound systems reveal various global regularities on different levels. There are implicational universals describing the composition of sound inventories (c.f. Maddieson, 1984; Plank & Filimonova, 2000; Hyman, 2008) and particularly the structure of human vowel inventories (Schwartz, Boë, Vallée, & Abry, 1997). There are also universal tendencies characterizing the structure of syllables (c.f. Bell & Hooper, 1978; Vennemann, 1988; Hammond, 1997). Finally, the existence of so-called *cross-linguistic phoneme correspondences* (Tiberius & Cahill, 2000) is assumed, i.e. regularities of the realization of the phonemic inventory, particularly the relationship of allophones and phonemes. A number of studies addresses a synthetic approach – a simulation model of interacting agents – to show that a number of regularities might not be a consequence of an innate human disposition, but the result of *self-organization processes* of interacting individuals (c.f. Lindblom & Maddieson, 1988; de Boer, 2000b; Jäger, 2008). In line with Blevins (2006), we address the following research question: *Can self-organization explain particular universal tendencies of human sound systems and syllable structures?* Additionally, we are interested in the application of procedures for the evaluation of synthetically emerged sound systems.

2. The Simulation Model

Simulation models proved themselves as a valuable technique to gain insights into the way some properties of human language emerge, such as compositionality (c.f. Nowak & Krakauer, 1999), complex syntactic patterns (c.f. Kirby & Hurford, 2002), or regular patterns in syllable structure (c.f. de Boer, 1997, 2000a), and sound inventory (c.f. Jäger, 2008). One essential goal of such studies is to show that the presence of particular features can be explained by functional factors – such as self-organization effects in communication (Lindblom & Maddieson, 1988; de Boer, 2000b) – without the need for postulating innate dispositions.

Following this line of research, we want to create a model of interacting agents to study the concurrent emergence of i) a sound inventory and ii) a lexicon. We apply a model framework that has proven fruitful for studying the emergence of vowel inventories: the *imitation game* model (de Boer, 2000a, 2000b).^a Furthermore, we want to analyze how realistic particular aspects of the emergent language systems are. We evaluate the quality by two factors: i) the syllable structure of the agents' lexical entries, and ii) the composition of the agents' sound inventories.

2.1. The Model Conception

The conception of our model is an altered version of de Boer's (2000b) imitation game model. It is altered in two main aspects:

1. Agents produce expressions instead of vowels. As a consequence they have two different types of inventories: i) a sound inventory of the sounds used to build expressions; ii) prototypes of expressions the agent produces.
2. Instead of having a very exact model of phonetic information, such as an articulatory and perceptual space, we abstract away from this level of detail by taking phonetic codes as acoustic units.

We adopted most of the other elements of de Boer's model, i.e. a similar agent architecture and an imitation game as interaction protocol (see Figure 1). To adapt this infrastructure to a communication model of complex expressions, we had to answer a number of design questions, of which three were most essential:

1. *What is the phonetic code of the sound inventory?* We used the ASJP code (c.f. Wichmann et al., 2013) (41 symbols), since it allows us to utilize real language data from the ASJP data base for evaluation. Subsequently, we reduced this set to 35 symbols, which enables us to integrate acoustic data from Mielke (2012) for setting up an agent's *production bias*.

^aIn his study, de Boer has shown that the prototypical structures of human vowel systems can be reproduced as a result of self-organization effects of agents that produce and imitate vowel sounds. In subsequent work, de Boer (2000a) also proposed a more realistic model by assuming agents to interact via more complex expressions.

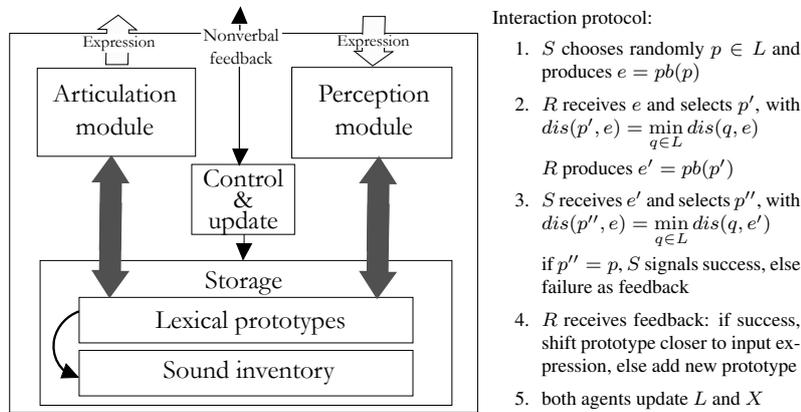


Figure 1. **Left:** the architecture of an agent. The *storage* module entails a) lexical prototypes of expressions L that is a subset of all possible expressions E , and b) the sound inventory X that entails exclusively sounds that occur in at least one lexical prototype. The *articulation module* realizes the production bias function $pb : E \rightarrow E$ which alters an expression $e \in E$ in a speaker biased way. The *perception module* is a search engine that finds the most similar prototype $p \in L$ for a received expression e . The control unit is informed by feedback signals and regulates the update of the prototypes, also by access to the sound inventory. **Right:** the interaction protocol of a communication step. The sender S chooses a prototype p from her storage and produces an expression e chosen from her prototypes and altered by production bias. If the storage of lexical prototypes is empty, a random expression is generated. The receiver R receives e , searches for the most similar prototype in her lexicon, and produces it (e'). S does the same and compares the new with the old prototype. If the prototypes are the same, S gives a feedback for success, otherwise for failure. Prototypes are updated according to the feedback: if success, R shifts her prototype closer to the input expression; if failure, either R adds the input expression to her prototype storage or S deletes prototype p , depending on p 's former communicative success. Finally, both agents update their storages (see de Boer, 2000a).

2. *How is the similarity between expressions computed?* To deal with strings of ASJP code, we used a *weighted alignment algorithm* (Jäger, 2013) to compute the distance between whole expressions.
3. *How is a random expression generated?* A random expression is generated by concatenating n randomly chosen symbols, whereby n itself is a random number between 1 and 10.

Note, that a realistic communication model has to entail *noisy signals* such that the produced signal does not totally resemble the intended prototype. To implement unbiased random noise seems to be an unrealistic assumption, since any alteration of the agent's production process of a complex expressions must be influenced by context effects. In this sense, we implemented a tendency of agents to produce a particular alteration to an expression as *production bias*: a probability function that determines the probability of each sound x of an expression to be replaced by

another sound x' in the production process, defined as follows:

Definition 1 (Production Bias) Let X be an alphabet of single sounds and let $e = x_1x_2 \dots x_{n-1}x_n$ be an expression, whereby $n \in \mathbb{N}_{>0}$ and $x_i \in X$ is the symbol of expression e at position i , $1 \leq i \leq n$. The production bias pb_α for replacing x_i with x' is defined as follows:

$$pb_\alpha(x_i \rightarrow x') = \frac{\alpha \times sim(x_i, x') + (1 - \alpha) \times P(x'|x_{i-1}, x_{i+1})}{\sum_{x'' \in X} (\alpha \times sim(x_i, x'') + (1 - \alpha) \times P(x''|x_{i-1}, x_{i+1}))}$$

In a nutshell: the production bias defines the probability to replace a sound x_i in context $x_{i-1}x_ix_{i+1}$ with a sound x' , which depends on a) how similar both sounds are to each other, thus $sim(x_i, x')$, and b) how conventional it is to position x' in such a context, given by probability $P(x'|x_{i-1}, x_{i+1})$. All in all, the production bias is defined as the weighted and normalized sum of similarity and conventionality. The free parameter α gives weight to similarity versus conventionality. The important issue in this part of the model is how to define i) similarity between two sounds $x, x' \in X$ and ii) conventionality of a sound being in a specific context.

For the similarity value we used a metric of production similarity that we gained from experimental data obtained by Mielke (2012). Based on different aspects such as nasal and oral airflow, larynx height and vocal fold contact area he computed similarities between sounds using a principal component analysis.

We define the conventionality of a sound in a given context as the probability of a sound x_i appearing between x_{i-1} and x_{i+1} , thus $P(x_i|x_{i-1}, x_{i+1})$, estimated on the basis of trigram frequencies in the ASJP data. The resulting counts were smoothed using an adapted version of Kneser-Ney-smoothing (Kneser & Ney, 1995).

Definition 2 (Conventionality) Let X be an alphabet of single sounds and let $P_{ENV}(x_i)$ be the probability of x_i in the context of x_{i-1} and x_{i+1} this is $P_{ENV}(x_i) = \frac{|\{x_{i-1}, x_{i+1} : c(x_{i-1}, x_i, x_{i+1}) > 0\}|}{\sum_j |\{x_{j-1}, x_j, x_{j+1} : c(x_{j-1}, x_j, x_{j+1}) > 0\}|}$, where $c(z)$ is the plain count of z . Let $\lambda(x_{i-1}, x_{i+1})$ be $\lambda(x_{i-1}, x_{i+1}) = \frac{d}{c(x_{i-1}) + c(x_{i+1})} \cdot |\{z : c(x_{i-1}, z, x_{i+1}) > 0\}|$, then

$$P(x_i|x_{i-1}, x_{i+1}) = \frac{\max(c(x_{i-1}, x_i, x_{i+1}) - d, 0)}{c(x_{i-1}) + c(x_{i+1})} + \lambda(x_{i-1}, x_{i+1}) \cdot P_{ENV}(x_i)$$

The smoothing parameter d was calculated in accordance with Sundermeyer, Schlüter, and Ney (2011) $d = \frac{n_1}{n_1 + 2n_2}$, where n_1 is the number of observed trigrams whose count is one and n_2 is the number of trigrams whose count is two.

2.2. Simulation Experiments

To gain a lot of data in a reasonable time frame, we decided to conduct first experiments with a very small population of three agents. The gist of our investigation

was to figure out how the implementation of the production bias influences the resulting lexicon and sound inventory. Thus, we made experiments with 4 different settings: 3 settings apply the production bias function pb as defined in Definition 1, each with different α parameters: 0.2, 0.5 and 0.8. The fourth setting does not involve a production bias, but a random alteration. In this setting randomly chosen sounds are occasionally replaced, deleted or added.

To neglect any assumption for a prior bias or disposition, each agent has an empty prototype storage and sound inventory at the beginning of a simulation run. With a probability of 0.01 an agent adds a new entry to her prototype storage. In each simulation step every agent communicates with every other agent according to the interaction protocol (Figure 1). As a primary result, in all runs the number of lexical prototypes and the size of the sound inventories both increased over time. Simulation runs were stopped after 4,000 simulation steps, where the size of emerged lexicons (number of expression prototypes) and the size of the emerged sound inventories of the agents roughly corresponded with sizes of concept lists and sound inventories of the languages of the ASJP database.

2.3. Evaluation

The results of the simulation were evaluated with respect to two different aspects: i) to measure the characteristics of the emergent lexicons of the agents, we analyzed the syllable structure^b of the resulting expressions (sound strings), and ii) to measure the quality of the emerged sound inventories of the agents, we analyzed the composition of sounds they entail. Therefore a list of rules for each of these aspects was compiled. These rules should represent universals regarding sound inventories or syllable structure. The rules were selected in such a way that they can be evaluated as true or false and can be answered within the provided framework. Some of the universals found in the literature could not be used for evaluation since they either focus on distinctions which are not measurable by using the ASJP code (rounded vs. unrounded vowels) or since the necessary information was not present in the system at all (minimal pairs). This results in 15 rules which were used to examine the sound inventory (Table 1) and five rules for the syllable structure (Table 2).

2.4. Results

We made 100 simulation runs for each setting (random alteration and $\alpha = 0.2, 0.5, 0.8$) and analyzed the resulting sound inventories and syllable structures according to the rule set. To compare the results with empirical data, we applied the rules to the 6895 lexicons of the ASJP data base.

Figure 2 (left) depicts the results for the estimation of the sound inventories: the box plots over percentage values of satisfied rules (15 sound rules) for all

^bWe used a self-implemented syllable parser inspired by Brunson (1989).

Table 1. Examples of implemented rules to check the quality of the resulting sound inventories.

Nr	Rule	Source
01	All languages have place distinctions of high and low and of front and back in their vowel systems; hence vowel systems minimally include /i, a, u/	
02	IF there are palatoalveolar consonants, THEN there are dental consonants.	
03	IF there are uvular stops, THEN there are velar stops.	
04	IF there is a glottal stop, THEN there must be a primary oral stop	
05	IF there is a voiceless palatal approximant, THEN there is also a voiceless labial-velar approximant	Plank and Filimonova (2000)
06	IF there are fricatives, THEN there will be stops	
07	IF there are back consonants, THEN there will be front consonants.	
08	IF there is any other lateral, THEN there will be a voiced lateral approximant.	
09	All languages have a high or a lower high front vowel	
10	IF there is the voiceless labial fricative phoneme /f/, THEN there will be the voiced labial fricative phoneme /w-v/.	
11	Every phonological system contrasts phonemes which are [-cont] (= stops) with phonemes that are specified with a different feature.	
12	Every phonological system has coronal phonemes	Hyman (2008)
13	Every phonological system has at least one front vowel or palatal glide /y/.	
14	Every phonological system has at least one back vowel.	
15	Every phonological system has stops	

Table 2. Implemented rules to check the quality of the resulting syllables.

Nr	Rule	Source
R_1	Syllables must have onsets.	Hammond (1997)
R_2	Syllables can't have codas.	Hammond (1997)
	A syllable is more preferred, ...	
R_3	...the steadier speech sound is.	Vennemann (1988)
R_4	...the smaller the number of speech sounds in the coda.	Vennemann (1988)
R_5	...the closer the number of speech sounds is to one.	Vennemann (1988)

sound inventories per setting. The results show that the production bias produces inventories that are significantly closer to the ASJP data than random alternation, whereas alternation of the α -value does not have a significant impact. This result reveals that the production bias – although operating locally on single strings – affects the way human sound inventories are compounded. Nonetheless, the inventories that emerged with production bias have still significantly lower values than the ASJP data. As a next step, further assumption should be tested that might improve the results.

The evaluation of the syllable features did not reveal significant results. All four lines of experiments produced lexicons with syllable structure with similar quality, all lower than those of the ASJP data. Nevertheless, by analyzing the runs with the lowest values for each setting (worst case values), we found that random alternation generally produces the lowest values, whereas the intermediate production bias parameter $\alpha = 0.5$ guarantees the greatest lower boundary for almost all rules, as shown in Figure 2 (right). This result was not significant,

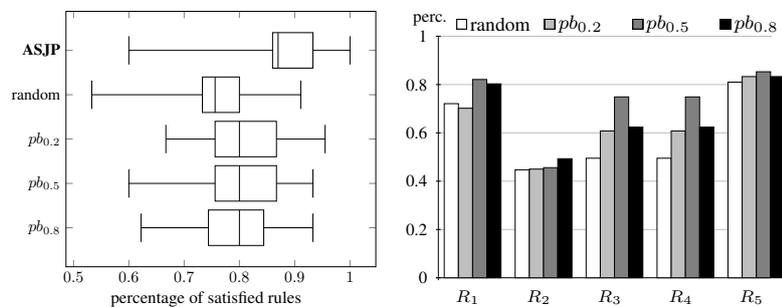


Figure 2. **Left:** result for the evaluation of the emerged sound inventories: box plots over percentage values of satisfied rules (15 sound rules, see Table 1). **Right:** result for the evaluation of the syllable features of the emergent lexicons: worst case values for each rule (see Table 2).

but might be a hint that an intermediate production bias prevents from producing syllables of low quality ($< .7$). Further investigations are necessary to test this suggestion and to detect more concretely the role of production bias in the process of syllable structuring.

All in all, the evaluation of the data reveals that a production bias supports the emergence of more realistic sound systems, though there is still room for improvement. Furthermore, a mid weighting (α -value) between similarity and conventionality for the production bias might increase the lower boundary of the quality of syllable structures.

2.5. Discussion

The results show that the some of the tendencies of syllable structure and sound inventory can be modeled by self-organization. The analysis of the resulting syllable structures is of course affected by the properties of the syllable parser. But since the syllable parser is only used for the analysis and not part of the simulation process it will not introduce any bias into the emergence of the syllable features itself. Nevertheless, a more sophisticated parser, e.g. statistical syllable parsing, may be necessary to better identify a syllable and classify its parts.^c

Another important aspect of the experiments is the presence of multiple agents, even though the population is considerably small. An innovation resulting from the production bias can only survive if it succeeds in communication. Thus there are two forces at work here. On the one hand, agents strive to produce strings in accordance with the *production bias*, i.e. strings which are more common and easier to produce; on the other hand, *communicative success* requires a

^cSuch an approach demands large corpora to successfully train the parser, which poses a problem in itself in the cross-linguistic case shown here.

stable code, i.e. maintenance of a form established in the population.

3. Conclusion and Outlook

The goal of this article is twofold: the first goal is the implementation of a model that can explain universal features of human languages by means of self-organization. The essential extension to former models in this area are i) a virtual society interacts through complex expressions, and ii) the members' lexicon and sound inventory emerges in parallel. In our study we used an extended version of de Boer's *imitation game*. The second goal is to establish evaluation methodology for synthetically emerged language system. In this study we proposed a first approach by implementing a rule-based matching system for universal tendencies of human sound inventories and syllable structure. This is but the first step to the development of a more elaborated evaluation system. Using the described evaluation methodology, it has been shown that self-organization can explain some particular universal tendencies in human sound systems and syllable structure.

Further work involves the development of a more fine grained evaluation mechanism. This involves a more sufficient syllable analysis mechanism and a more sophisticated and larger set of rules. Another aspect for improvement should focus on the production bias and its parts. Supplementary studies should investigate the influence of different similarity measures and agent structures.

Acknowledgements

The authors have been supported by the ERC Advanced Grant 324246 EVOLAEMP - Language Evolution: the Empirical Turn (<http://www.evolaemp.uni-tuebingen.de>), which is gratefully acknowledged. The authors especially thank Gerhard Jäger, Armin Buch, Igor Yanovich, Christian Bentz, Johannes Dellert and Marisa Köllner for insightful discussions.

References

- Bell, A., & Hooper, J. B. (1978). *Syllables and segments*. North-Holland Publ.
- Blevins, J. (2006). *Evolutionary phonology: The emergence of sound patterns* (3 ed.). Cambridge: Cambridge University Press.
- Brunson, B. (1989). Parsyl: A computer model of syllable parsing and acquisition. *Toronto Working Papers in Linguistics*, 10, 3-20.
- de Boer, B. (1997). Emergent cv-syllables. *Vrije Universiteit Brussel AI-lab AI-memo*.
- de Boer, B. (2000a). Imitation games for complex utterances. *Proceedings of BNAIC*.
- de Boer, B. (2000b). Self-organization in vowel systems. *Journal of Phonetics*, 28(4), 441-465.
- Hammond, M. (1997). *Parsing in OT* (Tech. Rep.). University of Arizona.

- Hyman, L. M. (2008). Universals in phonology. *The Linguistic Review*, 25(1-2), 83–137.
- Jäger, G. (2008). Applications of game theory in linguistics. *Language and Linguistics Compass*, 2/3, 408–421.
- Jäger, G. (2013). Phylogenetic inference from word lists: Using weighted alignment with empirically determined weights. *Language Dynamics and Change*, 3, 245–291.
- Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). Springer.
- Kneser, R., & Ney, H. (1995). Improved backing-off for M-gram language modeling. In *International conference on acoustics, speech, and signal processing* (Vol. 1, p. 181-184). Detroit, MI.
- Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In M. Larry, C. Hyman, & N. Li (Eds.), *Language, speech and mind* (pp. 62–78). Routledge.
- Maddieson, I. (1984). *Patters of sounds*. Cambridge University Press.
- Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua*, 145–163.
- Nowak, M. A., & Krakauer, D. (1999). The evolution of language. *PNAS*, 96(14), 8028–8033.
- Plank, F., & Filimonova, E. (2000). The universals archive: brief introduction for prospective users. *STUF - Language Typology and Universals*, 53(1), 109-123.
- Schwartz, J. L., Boë, L.-J., Vallée, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of Phonetics*, 25, 233–253.
- Sundermeyer, M., Schlüter, R., & Ney, H. (2011). On the estimation of discount parameters for language model smoothing. In *Interspeech* (p. 1433-1436). ISCA.
- Tiberius, C., & Cahill, L. (2000). Incorporating metaphonemes in a multilingual lexicon. In *Proceedings of the 18th conference on computational linguistics - volume 2* (pp. 1126–1130). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change : with special reference to german, germanic, italian, and latin*. Berlin: Mouton de Gruyter.
- Wichmann, S., Müller, A., Wett, A., Velupillai, V., Bischoffberger, J., Brown, C. H., Holman, E. W., Sauppe, S., Molochieva, Z., Brown, P., Hammarström, H., Belyaev, O., List, J.-M., Bakker, D., Egorov, D., Urban, M., Mailhammer, R., Carrizo, A., Dryer, M. S., Korovina, E., Beck, D., Geyer, H., Epps, P., Grant, A., & Valenzuela, P. (2013). *The ASJP Database (version 16)*.

EDWARD SAPIR AND THE ORIGIN OF LANGUAGE

ALBERT F. H. NACCACHE

Archaeology Department, Lebanese University, (ret.)

Beirut, Lebanon

anaccash@nidal.com

The field of Language Evolution is at a stage where its speed of growth and diversification is blurring the image of the origin of language, the “prime problem” at its heart. To help focus on this central issue, we take a step back in time and look at the logical analysis of it that Edward Sapir presented nearly a century ago. Starting with Sapir’s early involvement with the problem of language origin, we establish that his analysis of language is still congruent with today’s thinking, and then show that his insights into the origin of language still carry diagnostic and heuristic value today.

1. Introduction

The origin of language is a challenging problem to focus upon. Language, like mind and intelligence, is a phenomenon that “we find intuitive but hard to define” (Floridi, 2013, p. 601), and a century of progress in Linguistics has only heightened our awareness of the protean nature of language while exacerbating the fuzziness of its definition. Meanwhile, research in the field of Language Evolution is blooming and the sheer variety of available approaches, though highly promising, has blurred, momentarily at least, our perception of the issue at the core of the field: the origin of language.

We do not propose a solution to this problem, just an attempt to put it in perspective by looking at it through the writings of a researcher who had nearly no data with which to tackle the issue and could only rely on his analytical abilities. Published in 1921, Edward Sapir’s “Language. An introduction to the study of speech” is still in print and regularly studied and quoted, even in the field of Language Evolution (see Carstairs-McCarthy, 2012; Hurford, 2011; Mufwene, 2003). The claim here is that Sapir’s insights on the nature of language and its origin still offers diagnostic and heuristic clues to today’s researchers, in addition to the pleasure of glimpses at a keen mind at work.

We start by looking at Sapir’s early involvement with the problem of language origin, then establish that his analysis of language is still congruent

with today's thinking on the subject, and then review Sapir's insights into the nature and origin of language, those that have already been recognized by researchers in the field of Language Evolution, as well as those still challenging the field with their heuristic clues for future research.

2. From the origin to the nature of language

Edward Sapir, a student of Franz Boas, wrote his Master's thesis on Herder's 1781 "Essay on the Origin of Language," and published his work in 1907. In this paper, Sapir identified Herder's "epoch making" contribution as of having replaced the age-old perspectives on language, either as "given or revealed to man by God" or as invented, by a new perspective that looked at language as a faculty that had undergone a gradual evolution (1907, p. 4). Sapir also assessed that progress in Linguistics during the 19th century had "speedily relegated Herder's treatise to the limbo of things that were" (1907, p. 32). He concluded his paper by writing that "it seems to me that the path for future work on the prime problems, more especially the origin, of language lies in the direction pointed out by evolution," and, still upbeat and with impressive insight, the 23 years-old scholar proposed as main elements of such a research program "the careful and scientific study of sound-reflexes in higher animals," and "a very extended study of all the various existing stocks of languages, in order to determine the most fundamental properties of language" (1907, p. 34).

By 1921, when he published "Language," Sapir had operated a "Turing-type" shift in level of analysis (Floridi, 2013), that is, he had moved his attention away from the ill-defined problem of language origin and focused it on the more tractable problems of language change and variation. Still, in the "Introductory: Language Defined" chapter, and while analyzing the nature of language, Sapir briefly outlined some cardinal issues about language origin, issues he had identified through a logical analysis.

Sapir started by dismissing, as he had done in 1907, the then-current interjectional and onomatopoeic theories about the origin of speech, showing "how little the essential nature of speech is concerned with the mere imitation of things" (1921, p. 7 -page references are to the 2004 "Dover Books on Language" edition). Having thus cleared the way, Sapir then defined language as "a purely human and noninstinctive method of communicating ideas, emotions, and desires by means of a system of voluntarily produced symbols" (p. 7).

Sapir qualified this definition as "serviceable." How does it stand today, especially as seen from the field of Language Evolution? Though there is no orthodoxy in the field, a case can be made that Tallerman and Gibson's

“Oxford Handbook of Language Evolution” (2012), which represents the *zeitgeist* of a core group of researchers, could be used as a benchmark. In their general introduction, which shows the uniqueness of language among animal communication systems and reviews the various approaches being pursued in studying the subject, Tallerman and Gibson do not present a definition of language *per se*, but nothing they say conflicts with Sapir’s definition.

Not only that, but consider the following two quotes. Sapir writes: “Speech is not a simple activity that is carried on by one or more organs biologically adapted to the purpose. It is an extremely complex and ever-shifting network of adjustments—in the brain, in the nervous system, and in the articulating and auditory organs—tending towards the desired end of communication... Physiologically, speech is an overlaid function, or, to be more precise, a group of overlaid functions. It gets what service it can out of organs and functions, nervous and muscular, that have come into being and are maintained for very different ends than its own” (1921, p. 7). While Tallerman and Gibson write: “‘language’ is not a monolithic entity, but rather a complex bundle of traits that must have evolved over a significant time frame, some features doubtless appearing in species that preceded our own. Moreover, language crucially draws on aspects of cognition that are long established in the primate lineage, such as memory: the language faculty as a whole comprises more than just the uniquely linguistic features.” (2012, p. 2). The similarity in the analysis of the complex nature of language displayed by the two texts, written 91 years apart, is clear.

It would be easy to multiply Sapir’s quotes that are modern in spirit, if not terminology, but space prevents us from doing so. The following quotes, selected to show that Sapir’s insightful characterizations of the nature of language can still provide fruitful guidelines to the research program of Language Evolution, will also have to serve to show that his analysis withstood the test of time.

3. Sapir’s insights on the nature of language

After pointing out that “the essence of language consists in the assigning of conventional, voluntarily articulated, sounds, or of their equivalents, to the diverse elements of experience,” Sapir engaged in an analysis of the symbolic associations of words, and wrote that: “The elements of language, the symbols that ticket off experience, must therefore be associated with whole groups, delimited classes, of experience rather than with the single experiences themselves. Only so is communication possible, for the single experience lodges

in an individual consciousness and is, strictly speaking, incommunicable. To be communicated it needs to be referred to a class which is tacitly accepted by the community as an identity” (p. 9). Sapir and his generation of “anthropological linguists” were pioneers in the scientific analysis of the social aspect of language, which had long been ignored even though Zhuangzi had already written in the 3rd Cent. BC that, “like a path is formed by constant treading on the ground, things are called by their names through the constant application of the names to them” (Book II, 4, my phrasing based on Legge’s and Wiegner’s translations). Still, even today, the social dimension of language is not wholeheartedly endorsed by linguists, nor fully exploited in *Language Evolution* (Gibson & Tallerman, 2012, p. 248-9). What gives Sapir’s analysis its enduring heuristic value is that he focused it on the elements of experience shared by the community. In doing so, he directs us to search for the social and technological settings of the hominin groups in which language emerged, a task in which the field has, as yet, barely engaged.

Language not only depends on a social consensus to function as a communication device, it also shapes what it makes common. Sapir wrote: “It is, indeed, in the highest degree likely that language is an instrument originally put to uses lower than the conceptual plane and that thought arises as a refined interpretation of its content. The product grows, in other words, with the instrument, and thought may be no more conceivable, in its genesis and daily practice, without speech than is mathematical reasoning practicable without the lever of an appropriate mathematical symbolism” (p. 10). Hurford refers to Sapir’s insight that “the product grows with the instrument” as a positive feedback loop leading to “the significant effect on individual thought that possession of publicly shared symbols can have, even as yet without any syntax to combine them” (2012, p. 172), adding “Humans have evolved a unique type of complex system for expressing their thoughts. (And their thoughts have become more complex as a result.)” (2012, p. 175). Of course, “the co-evolution of language and the human brain is not a new idea. Darwin (1871) himself proposed a co-evolutionary spiral of language and brain” (Számadó & Száthmary, 2012, p.165). Still, Sapir’s analysis should help us remember that the linguistic communication that first emerged was not language as we know it now, and also that we have to account not only for language-brain but also for language-culture co-evolution.

In pursuing such attempts, we can again take advantage of Sapir’s insights: “We may assume that language arose pre-rationally—just how and on what precise level of mental activity we do not know—but we must not imagine that a highly developed system of speech symbols worked itself out before the

genesis of distinct concepts and of thinking, the handling of concepts. We must rather imagine that thought processes set in, as a kind of psychic overflow, almost at the beginning of linguistic expression; further, that the concept, once defined, necessarily reacted on the life of its linguistic symbol, encouraging further linguistic growth” (p. 10). Schoenemann suggested one such “Sapirian” scenario in which “syntax is more properly understood as an emergent characteristic of the explosion of semantic complexity that occurred during hominid evolution” (1999). Hurford reached, on a theoretical basis, a conclusion similar to that of Sapir, writing that “the human capacity for syntax evolved ... Plausible evolutionary accounts should conform to general evolutionary theory, and this consideration tends strongly to recommend a gradual trajectory” (p. 180). All attempts to reconstruct the trajectory from words to sentences, or from proto-language to language, would benefit from incorporating Sapir’s analysis of the “beginning of linguistic expression” as a positive feedback system.

What Sapir wrote next has important implications: “communication, which is the very object of speech, is successfully effected only when the hearer’s auditory perceptions are translated into the appropriate and intended flow of imagery or thought or both combined. ... The concordance between the initial auditory imagery and the final auditory perceptions is the social seal or warrant of the successful issue of the process” (p. 11). A review of verbal interaction, “the prime ecological niche for language, the context in which language is learned, in which the cultural forms of language have evolved” (Levinson & Torreira, 2015, p. 1), confirms this insight. The “social seal” provided by shared experience is required to make prediction possible and insure successful conversation. This social aspect of linguistic communication should be reflected in all scenarios of language emergence and accretion.

4. Sapir’s challenges on the origin of language

Sapir concluding thought in the introductory chapter is very challenging to the field: “The universality and the diversity of speech lead to a significant inference. We are forced to believe that language is an immensely ancient heritage of the human race, whether or not all forms of speech are the historical outgrowth of a single pristine form. It is doubtful if any other cultural asset of man, be it the art of drilling for fire or of chipping stone, may lay claim to a greater age. I am inclined to believe that it antedated even the lowliest developments of material culture, that these developments, in fact, were not strictly possible until language, the tool of significant expression, had itself taken

shape” (p. 13). Hurford writes that “it is quite possible that *Homo erectus*, perhaps for over a million years, had symbolic pre-syntactic communicative behaviour. They may have had inventories of learned conventional ‘words’ referring to things and events” (2012, p. 483), and Tallerman reports that “many proposals link the *Homo* genus with the first protolanguage, perhaps 2 million years ago (mya)” (2012, p. 479). Still, this insight is probably the hardest to fit within the present models of language evolution (Gibson & Tallerman, 2012), even with those proposing the earliest emergence (such as Naccache, 2012). There was already a problem with the 2.6 million-year-old lithics found in Gona, Ethiopia, which predated by more than half a million years the earliest *Homo*. And now the problem is compounded by the recently found 3.3 million-year-old stone tools uncovered at Lomekwi 3, in Kenya (Harmand, 2015). Addressing this issue is one of the biggest challenges facing Language Evolution. We can already state confidently that it will require abandoning saltational scenarios of recent (50 to 100,000 years ago) language emergence with *H. sapiens*. It will also require facing up to the fact that, though cultural evolution today is explosive, and has been so since the end of the Paleolithic, at its origin it was excruciatingly slow, and this because it lacked the appropriate ratchetting mechanisms that had yet to be socially and culturally implemented.

All Sapir’s previously mentioned analytical insights into the origin of language appeared in the introductory chapter. In the penultimate chapter, “Language, Race and Culture,” Sapir provides us with hints at two potential research avenues that might gain us access to material correlates of the origin and evolution of language.

The first approach is inspired by Sapir’s characterization of language as “a collective *art* of thought,” and/or “a particular *how* of thought” (p. 104). Looking on language as an art of thought is heuristically fruitful because, while it is true that language and/or thought do not fossilize, thoughts have material realizations that do. Once recovered by archaeology these thought realizations can be used to infer back to the thought processes that made them possible. For example, it is clear today that the syntactic operation of predication is one of “the most fundamental property of language” (Sapir, 1907), distinguishing language from all other animal communication systems, and this independently of the evolutionary role that Bolhuis et al. postulate for syntactic predication, which they refer to as “*merge*” (2014). Since we can identify the cognitive operations that need to be deployed for predication, or *merge*, to be possible (Bogdan, 2009), and since we can find reflections of the same cognitive operations in the lithic record (Stout et al., 2015), we have a way of estimating the date at which

our ancestors would have had the cognitive potential to include *merge* in their communicative behavior (Naccache, 2014).

Sapir's second clue to potential material correlates of language evolution is presented quite clearly as a challenge by Sapir himself: "The drift of culture, another way of saying history, is a complex series of changes in society's selected inventory—additions, losses, changes of emphasis and relation. The drift of language is not properly concerned with changes of content at all, merely with changes in formal expression ... If it can be shown that culture has an innate form, a series of contours, quite apart from subject matter of any description whatsoever, we have a something in culture that may serve as a term of comparison with and possibly a means of relating it to language. But until such purely formal patterns of culture are discovered and laid bare, we shall do well to hold the drifts of language and of culture to be non-comparable and unrelated processes. From this it follows that all attempts to connect particular types of linguistic morphology with certain correlated stages of cultural development are vain" (p. 104).

Nearly a century after the publication of "Language" we are still far from having discovered the "formal patterns of culture" that would allow us to correlate them to "particular types of linguistic morphology." However, such goal now seems to be nearly within sights of researchers. For an example from the collateral field of the emergence of rituals, Marcus & Flannery have documented in Oaxaca, Mexico, "changes in religious ritual that accompanied the evolution of society from hunting and gathering to the archaic state" (2004, p. 18257), thus relating the evolution of patterns of culture with those of ritual performances. In a previous Evolang contribution, I have correlated between, on the one hand, social relationships and collective memory and on the other, not linguistic morphology, but the progressive attainment of Hockett's "design features" of the human linguistic communication system, and this over the last 2M years (Naccache, 2012). And Lupyán and Dale, based on a large scale statistical analysis of structural language properties, have documented that, today, "language structures appear to adapt to the environment (niche) in which they are being learned and used" (2010), a finding which opens up the potential of extending such analysis to the whole of the historical period, and eventually to the Neolithic and Paleolithic periods.

In view of these developments, the academic goal of differentially characterizing the linguistic morphology, or the *how* of thought, available to Plato and the headhunter of Assam on the one hand from that which had been available to *H. heidelbergensis*—or even to *H. ergaster*, should not seem

anymore impossibly out of reach for the coming generations of researchers in the field of Language Evolution.

5. Conclusion

Hopefully, this brief review of the insights of Edward Sapir on what he termed the “prime problem” of language, that is, the problem of the origin of language, has shown that Sapir was able to achieve a solid characterization of the nature of language while identifying some important conditions that must have held to make the emergence of language possible. What is worth pondering for the researchers in the field of Language Evolution today is that Sapir was able to achieve these characterizations and identifications nearly a hundred years ago, when the empirical database at his disposal was extremely limited, if not to say inexistent. Sapir succeeded in achieving enduring characterizations of the conditions necessary for the human linguistic communication system to emerge and accrete, purely by relying on a logical analysis of the problem of language seen as a whole in its communicative context.

Today, it is very encouraging to see the field of Language Evolution luxuriating in mining its highly multidisciplinary databases and studying all aspects of the linguistic phenomenon, formal, embodied, socialized and encultured. But, in addition, and no matter how hard it is to do, it cannot do any harm to try to look at the problem of language origin and evolution through an overarching logical analysis. No doubt, Edward Sapir is hard to emulate, but he has proved that such an approach is fruitful, and therefore worth attempting.

References

- Bogdan, R. J. (2009). *Predicative Minds. The Social Ontogeny of Propositional Thinking*. MIT Press.
- Bolhuis, J., I. Tattersall, N. Chomsky & R. C. Berwick (2014). How could language have evolved? *PLoS Biology*, **12**, 8, e1001934
- Carstairs-McCarthy, A. (2012). The evolutionary relevance of more and less complex forms of language. p. 435-441, in Tallerman, M. & K. R. Gibson, eds, *The Oxford Handbook of Language Evolution*. Oxford University Press.
- Floridi, L. (2013). The Turing Test and the Method of Levels of Abstraction. in Cooper, S. B. & J. van Leeuwen eds., *Alan Turing: His Work and Impact*. New York: Elsevier

- Gibson, K. R. & M. Tallerman (2012). The Prehistory of Language: When and Why did language evolve? p. 239-249 in Tallerman, M. & K. Gibson, eds. *Oxford Handbook of Language Evolution*. Oxford University Press.
- Harmand, S. et al. (2015). 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature*, **521**, 310-315.
- Herder, J. G. (1781). *Essay on the Origin of Language*. Translated by J. H. Moran and A. Gode. University of Chicago Press. 1966.
- Hurford, J. (2002). *Origins of Grammar: Language in the Light of Evolution II*. Oxford University Press.
- Levinson, S. C. & F. Torreira (2015). Timing in turn taking and its implications for processing models of language. *Frontiers in Psychology*, **6**: 731.
- Lupyan, G. & R. Dale (2010). Language structure is partly determined by social structure. *PLoS ONE*, **5**, 1, p. 1-10.
- Marcus, J. & K. V. Flannery (2004). The coevolution of ritual and society: New 14C dates from ancient Mexico. *PNAS*, **101**, 52, p. 18257–18261.
- Mufwene, S. S. (2003). Analogs anywhere: the flow of highway traffic and language evolution. *Contemporary Linguistics*, **3**.
- Naccache, A. (2012). Hominin Cooperation and Language Evolution. p. 250-257, in: Th. C. Scott-Philips, M. Tamariz, E. A. Cartmill & J. R. Hurford (eds), *The Evolution of Language*. World Scientific.
- Naccache, A. (2014). *Homo praedicans*. p. 197-204, in: E. A. Cartmill, S. Roberts, H. Lyn & H. Cornish (eds), *The Evolution of Language*. World Scientific.
- Sapir, E. (1907). Herder's "Ursprung der Sprache." *Modern Philology*, **5**, 1, p. 109-142.
- Sapir, E. (1921). *Language. An introduction to the study of speech*. New York. Harcourt, Brace. Reprint: Dover Books on Language, 2004.
- Schoenemann, P. Th. (1999). Syntax as an Emergent Characteristic of the Evolution of Semantic Complexity. *Minds and Machines*, **9**, 3, p. 309–346.
- Stout, D., E. Hecht, N. Khreisheh, B. Bradley & T. Chaminade (2015). Cognitive Demands of Lower Paleolithic Toolmaking. *PLoS ONE*, **10**, 4, e0121804.
- Számadó S. & E. Száthmary (2012). Evolutionary biological foundations of the origin of language: the co-evolution of language and brain. p. 157-167 in Tallerman & Gibson, eds. *Oxford Handbook of Language Evolution*. OUP.
- Tallerman, M. (2012). Protolanguage. p. 479-491 in Tallerman, M. & K. Gibson, eds. *Oxford Handbook of Language Evolution*. OUP.
- Tallerman, M. & K. R. Gibson (2012). Introduction: The Evolution of Language. p.1-35 in Tallerman, M. & K. Gibson, eds. *Oxford Handbook of Language Evolution*. Oxford University Press.
- Zhuangzi (3rd cent. BC). Legge's translation: <http://oaks.nvg.org/chuang.html>.
Wieger's translation: http://classiques.uqac.ca/classiques/wieger_leon/B15_pere_systeme_taoiste/les_peres.pdf

QUANTIFYING THE SEMANTIC VALUE OF WORDS

DILLON NIEDERHUT

*Department of Anthropology, University of California
Berkeley, USA
dillon.niederhut@berkeley.edu*

Hypotheses about the evolution of human language often posit a role for the informativeness of speech acts. However, there has yet to be an accessible method for measuring the semantic value of a word or group of words. This paper outlines a novel test statistic for determining the relative semantic value of words in a language, given a corpus that approximates ecologically valid use of that language. Future work could use this test to disambiguate between social and cognitive factors in linguistic change.

1. Introduction

The extent to which human language systems evolved because it was beneficial for early hominins to communicate well has been a topic of heavy debate for some time. While this debate has advanced to the point of some researchers arguing that communicating poorly might be even more beneficial than doing it well, there has yet to be a method for measuring the communicative worth of a speech act (Pinker, Nowak, & Lee, 2008).^a Were such a method to exist, hypotheses that posit a utility motive for the origin of language could be tested in natural experiments involving the acquisition or change of that language.

2. Natural word use follows a Zipfian distribution

In conversational English, a few words are used many times, and the rest are used with shocking rarity (Fig. 1). Even words that seem relatively common, like CARPET, SCALE, and WEIRD, appear fewer than once in every twenty thousand words. To put this in context, it has been estimated that the average person speaks 16,000 words per day (Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007).

Intuitively, words that are used very frequently don't seem to carry much semantic value. The word MY, for example, really only tells you that something is

^aOne can measure the Shannon Information of human language, but high entropy words are not necessarily the same as semantically valuable words.

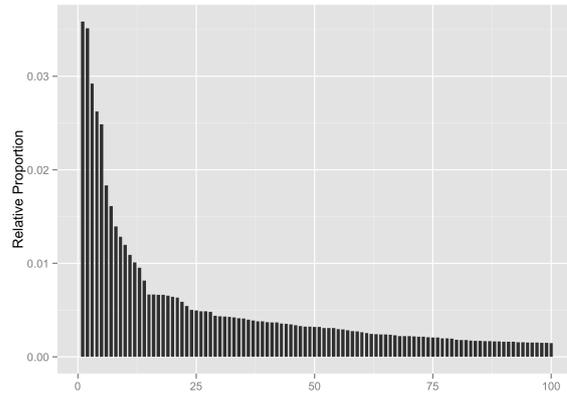


Figure 1. Relative proportion of the 100 most common words in English, from the sample described below.

being possessed by the speaker, but does not tell you anything about that something or about the speaker. To take another example, the word YEAR probably indicates a period of 365 days, but might not have the expected start date (i.e. a fiscal year) or might only refer to nine months out of those 365 days (i.e. an academic year).

On the other hand, uncommon words seem to carry a lot of semantic value. For example, were the speaker to use the word MULTICOLLINEAR, a native speaker of English familiar with statistics could assume the following statements (ordered by decreasing likelihood):

1. the speaker is a fairly educated person who does statistical analyses, speaking to one or more persons who are fairly educated and understand statistics
2. the speaker, in particular, is well informed about the general linear model
3. the context under discussion is a model with two or more predictors that are highly correlated, and thus has unstable linear coefficients
4. the speaker will go on to mention ways to measure this, like the variance inflation factor, and ways to correct it, like principal component analysis
5. the speaker does not work with very large datasets or machine learning methods

In this particular case, a single word is giving us a rich set of inferences about who is involved in the conversation, their knowledge state, the topic under dis-

discussion, and what will happen next. It should be obvious that this is much more information than was provided by the words MY and YEAR.

However, *context* is difficult to define and even harder to measure. To continue the example above, measuring the education level of every person involved in a conversation, along with the major topic under discussion and the fields in which the speaker does not work in a dataset large enough to draw meaningful inferences about language is not currently feasible. However, the fourth point above – measuring the frequency of associated words like VARIANCE INFLATION FACTOR – is easily and frequently implemented.

3. Semantic value is a change in that distribution

We can imagine, then, that a rare word is one which refers to a rare context; and conversely, that rare contexts tend to be described with rare words. In that case, we may posit that the semantic value of a word is related to the distance between the words associated with it relative to their frequency of use in the language as a whole (Wittgenstein, 1953; Salton, Wong, & Yang, 1975; Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990). To state this another way, an informative word is one which changes the relative proportions of the words surrounding it each time it appears.

This argument differs from prior work which assumes that the semantic content of a word is defined by the words that appear nearby (Gentner, 1983; Furnas, Landauer, Gomez, & Dumais, 1983). The model in this paper is the nearby words point to the same latent variable, which is the difficult-to-measure real-world context of the speech act. The implementation, however, is very similar to modern methods in computational semantics like pairwise mutual information or latent semantic analysis (LSA) (Lund & Burgess, 1996; Turney & Pantel, 2010). The difference here is that we are not trying to assign a similarity between the distributions of one single word versus another single word; we are attempting to assign a single value that demonstrates the relationship between a single word and the distribution of all words in the English language.

We can calculate the distance between the word distribution of all events that contain a single word of interest with the overall English distribution using the chi squared test:

$$\sum_i^k \frac{(p_{sample} * n_{sample} - p_{population} * n_{sample})^2}{p_{population} * n_{sample}}$$

where k is the list of unique words in the population; i is one word in that list; n_{sample} is the total number of words used in communicative events that also include the word being measured; and, p_{sample} is the probability that any one in n words is word i .

To make this a little more concrete, imagine looking at the informativeness of

the word MULTICOLLINEAR. Communicative events that contain the word MULTICOLLINEAR also tend to contain words like VARIABLE, CLUSTER, and MODEL. These words have very high relative proportions in our sample, but low relative proportions in our population. So, for each of these, we might add something like:

$$\frac{(1E+01 - 1E-06)^2}{1E-06} = 1E+08$$

to the total sum. However, these events also contain many common words like TO, OF, and A, at close to their relative proportion in the total population of words. Each of these words adds a smaller value to the total sum.

This method of measuring distance is favorable in that the appearance of rare words is heavily weighted. It is unfavorable in that it is also sensitive to the total number of words, n . A very common word like MY has a very large n , so even small differences between the proportions of a word in the sample and the population produce values that are in the zeroth or first of magnitude. A very uncommon word with a small n will frequently produce values that are equal to the expected frequency of each word in the population, which is typically below $1E-05$ – five orders of magnitude less. When summed across the number of unique words in the population (the length of k), the effect of n dominates the calculated value.

4. Correcting the magnitude of that change produces a test statistic

To produce a test statistic from the chi squared values, one first needs to correct the bias produced by the size of the sample. Then, the distribution of the test statistic must be characterized to produce population parameters for the expected mean value and variance. Both of these steps require real word linguistic data that has been decomposed into a distribution of frequency counts.

An English word distribution was created using a Python library written by the author, available at <https://github.com/deniederhut/medicorpus>, by randomly sampling comments from the discussion board at [reddit.com/r/AskReddit](https://www.reddit.com/r/AskReddit) over a period of 22 weeks. This particular discussion board was chosen both for its high traffic rate and the broad topics and conversational nature of the discussions there. This resulted in a total sample size of $4.29E+06$ communicative events, with $1.08E+08$ total unigrams and $2.93E+05$ unique unigrams appearing more than once.

To describe the sampling distribution of chi square values, comments were randomly sampled from the total corpus twenty times each at the probabilities $1E-02$, $1E-03$, $1E-04$, and $1E-05$. In each of the 100 samples, term frequency was set to equal the number of comments, and the chi squared value of the term frequencies in the sample was calculated. As predicted above, the magnitude of the chi squared statistic is dependent on the size of the sample used to calculate the statistic (Fig. 2).

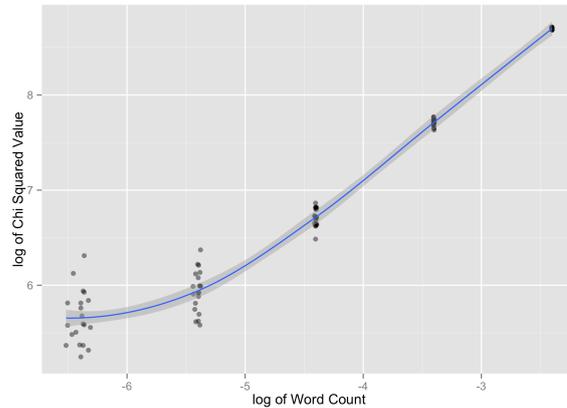


Figure 2. Chi squared values produced by comparing word distributions are convolved with the size of the sample.

The relationship between chi squared values and sample size linearizes when each variable is square root transformed. Happily, this transformation also causes the variability in chi squared values to become constant. After linearization, the chi squared values from each random sample are almost perfectly predicted by the total number of words in all the comments used to compute the value (Fig. 3).

The stable variability means that a simple test value can be created using the chi square value, corrected for word count, and divided by the constant standard deviation. For the sake of brevity, we'll call this the Zipf test.

$$z_{correct} = \frac{(-274.85 + \frac{\sqrt{n}}{6.37})}{288.33}$$

5. The test statistic conforms to expected behavior

As a proof of concept, the Zipf statistic was calculated for several words from the corpus (Table 1). Generally speaking, it produces values concurrent with our intuition. Very common words, like MY, DAY, and FEEL, have negative Zipf statistics that increase in linearly in magnitude with the square root of their relative proportion in the population. Words with a relative proportion around 1E-05 have Zipf statistics that are close to zero, or close to the corrected values derived from randomly sampling the entire population. Words that appear less frequently than this have an increasing change of producing positive Zipf statistics, indicating words with high semantic content.

The Zipf test provides a method for quantifying the relative informativeness

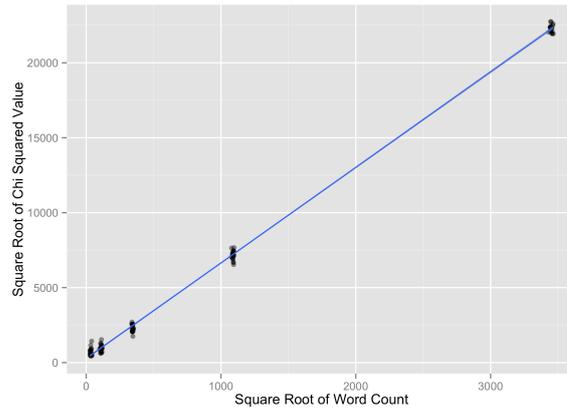


Figure 3. Square-root transforming the chi squared values and sample sizes linearizes both the relationship and the variability in y.

Term	Relative Proportion	Test Value
my	1.30E-02	-9.41E+01
day	2.12E-03	-5.82E+01
feel	1.71E-03	-5.04E+01
record	1.24E-04	-7.13E+00
unclear	1.11E-05	-8.12E-01
hither	6.09E-07	6.80E-01
dill	2.92E-06	8.25E-01
omlette	3.42E-07	1.78E+00
multicollinear	3.69E-08	3.28E+00

of any word, given a language corpus that is divided by communicative events. Additionally, this is an objective method that can be implemented largely without human intervention, and in any language that is easily tokenized or lemmatized. It measures the semantic value of a given word by comparing the frequencies of other words used in the same context with the distribution of all the words that appear in the corpus.

6. Hypotheses made tractable by this test

If the evolution of human language was driven by a need for effective communication, we would expect evolutionary pressures to produce cognitive systems that prioritize the acquisition of information-heavy terms. Specifically, we would

hypothesize that words with high semantic value would:

- be learned sooner in infancy; and,
- spread more quickly through a population; and,
- be preferentially adopted across languages.

More generally, the Zipf test should be useful in testing predictions of language change and use that include social and cognitive factors. For example, one could ask whether the diffusion of linguistic variants is better predicted by the utility of the word, or its use as a marker of social identity (Eisenstein, O'Connor, Smith, & Xing, 2014). It should also be possible to create historical data on the rate of semantic bleaching of words, and to investigate if a relationship exists between that rate and populations employing the word.

7. Supplementary information

Data were collected in Python 2.7.8 on Ubuntu Server 14.0.4, and were analyzed with Revolution R Open^b based on CRAN release v. 3.2.1, “World-Famous Astronaut”, (R Core Team, 2015). Tables were produced with *xtable*, and figures were produced with *ggplot2* (Dahl, 2014; Wickham, 2009). The code and data necessary to reproduce this paper are available at <https://github.com/deniederhut/quantifying-semantic-value>.

Acknowledgements

The author would like to acknowledge Gabe Doyle and Dan Yurovksy for their methodological advice; and Terrence Deacon, Madza Virgens, and Drew Halley for their comments on this manuscript.

References

- Dahl, D. (2014). *xtable: Export tables to latex or html*. (R package version 1.7-4)
- Deerwester, S., Dumais, S., Landauer, T., Furnas, G., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*, 391–407.
- Eisenstein, J., O'Connor, B., Smith, N., & Xing, E. (2014). Diffusion of lexical change in social media. *PLoS One*, *9*, e113114.
- Furnas, G., Landauer, T., Gomez, L., & Dumais, S. (1983). Statistical semantics: Analysis of the potential performance of keyword information systems. *Bell System Technical Journal*, *62*, 1753–1806.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.

^b<https://mran.revolutionanalytics.com/open/>

- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, 28, 203–208.
- Mehl, M., Vazire, S., Ramirez-Esparza, N., Slatcher, R., & Pennebaker, J. (2007). Are women really more talkative than men? *Science*, 317, 82.
- Pinker, S., Nowak, M., & Lee, J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, 105, 833-838.
- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria.
- Salton, G., Wong, A., & Yang, C. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18, 613–620.
- Turney, P., & Pantel, P. (2010). From frequency to meaning; vector space models of semantics. *Journal of Artificial Intelligence Research*, 37, 141–188.
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.
- Wittgenstein, L. (1953). *Philosophical investigations*. Hoboken: Blackwell.

SEMANTIC APPROXIMATION AND ITS EFFECT ON THE DEVELOPMENT OF LEXICAL CONVENTIONS

BILL NOBLE, RAQUEL FERNÁNDEZ

*Institute for Logic, Language and Computation
University of Amsterdam*

We define a signaling games setting for investigating how short- and long-term conventions are established in a community of interacting speakers. Using simulations, we model a particular type of non-literal use of linguistic expressions, *semantic approximation*, and investigate its effects on lexical alignment, ambiguity, polysemy, and communicative success. Critically, in our approach agents do not only keep track of a lexicon reflecting conventions at the level of the community, but also of a *discourse lexicon* that stores information agreed upon by the participants in a specific dialogue. We find that semantic approximation creates opportunities for discourse-level lexicalization, which boosts the expected utility of the discourse lexicon, and that it can have a profound effect on the evolution of community-level lexical resources.

1. Introduction

One of the most striking features of human language is its flexibility—despite the presence of ambiguity and possibly mis-aligned linguistic resources, speakers manage (most of the time) to communicate effectively with each other. We argue that the use of *semantic approximation* is a key mechanism for achieving communicative success. Notable examples of semantic approximation are found in the context of language acquisition, as illustrated in (1) from Chaix, Barry, and Duvignau (2012), where young children extend the meaning of a known-to-them term (in this case the verbs ‘*undress*’ and ‘*turn on*’) to approximate the concept they want to express, for which presumably they lack a more suitable word:

- (1) a. Undress the potato? [*age 2; context: mother peels a potato*]
b. Go on mum, turn on your eyes [*age 3; context: mother has her eyes closed*]

Semantic approximation, however, is also common in regular language use amongst adults, often signalled by hedging (e.g., ‘*kind of*’, ‘*-ish*’). For instance, consider the following dialogue excerpt from Brennan and Clark (1996):

- (2) A: A docksider.
B: A what? [...] Is that a kind of dog?
A: No, it’s a kind of um leather shoe, kinda pennyloafer.
B: Okay, okay, got it.

In (2), after having used an unsuccessful term (*'docksider'*) to jointly refer to a particular object, speaker A proposes an alternative expression via semantic approximation (*'kinda pennyloafer'*). From then onwards in the conversation, these speakers successfully use the (unhedged) term *'pennyloafer'*.

This illustrates two key features of language use. First, by resorting to semantic approximation (rather than to coining a completely new expression), speakers exploit the potential for polysemy present in natural language, which arguably can have a beneficial effect on efficient communication (Piantadosi, Tily, & Gibson, 2012; Juba, Kalai, Khanna, & Sudan, 2011). Second, speakers can make use of semantic approximation to establish *ad-hoc* conventions over the course of a dialogue. These factors raise several intriguing questions, which we set out to investigate in this paper. How do novel conventions emerge in one-on-one conversation? What effect do semantic approximation and ambiguity have on the lexicon of interacting agents? To what extent does local interaction shape the lexicon of a linguistic community?

2. Formal Model

We use signaling games (Lewis, 1969) to model semantic approximation and investigate its effect on the communicative success of a linguistic community. Our model consists of a community of agents, A , who share a set of atomic expressions (or messages) M , and a conceptual domain C . A similarity measure is defined over pairs of concepts in C . A simulation consists of a series of dialogues between pairs of agents in A . In a dialogue, the interlocutors take turns producing and interpreting single-expression utterances from M . For each utterance, the speaker chooses an expression based on a *speaker intention* modeled as a concept chosen at random from C . The addressee then guesses the speaker's intention, selecting an interpretation for the uttered expression from C as well. Communicative success is determined by the degree of similarity between the speaker's intention and the addressee's interpretation.

Lexicons. A lexicon is a linguistic resource that agents use to keep track of the agreed-upon meanings of words. In natural language there is no centralized objective lexicon. Instead, a lexicon is a subjective representation of joint information—it is what each agent takes to be the lexical common ground (Stalnaker, 1978). We follow Clark (1996) in the distinction between different types of common ground: Personal common ground is built up between individuals based on shared experience, whereas communal common ground holds based simply on mutual community membership. In this model, the two kinds of common ground correspond to two different lexicons that agents $a \in A$ keep: the community lexicon (CL_a) and the discourse lexicon (DL_a). The discourse lexicon is based on the community lexicon, but also incorporates lexical information agreed upon by the participants of a particular dialogue (i.e., the interlocutors' personal common

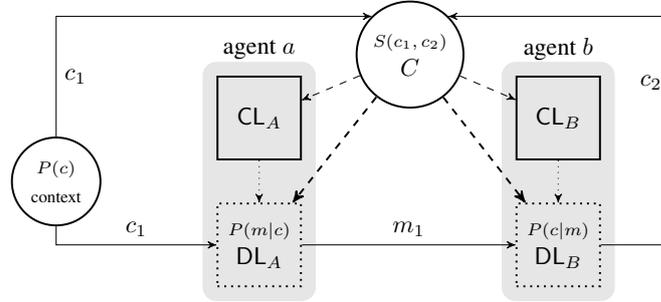


Figure 1. At the beginning of a dialogue, agents construct a discourse lexicon based on their representations of the community lexicon (dotted arrows). In this diagram, agent a expresses c_1 to b using expression m_1 , and agent b interprets m_1 as c_2 . The interlocutors update their lexicons (dashed arrows) based on their communicative success as determined by the similarity of c_1 and c_2 . The discourse lexicons are more radically influenced by this update.

ground based on the discourse itself as a shared experience).

For each lexicon (C|D)L (community or discourse), agents $a \in A$ keep a $|C| \times |M|$ matrix L_a , where $L_a[c, m] \in \mathbb{R}^+$ represents the sum of a 's evidence that it is common ground that m means c . From this matrix, agents may derive conditional probability distributions L_a^p over messages given concepts (for production) and L_a^i over concepts given messages (for interpretation).^a An agent's discourse lexicon is initially just a normalized copy of her community lexicon—the initial discourse lexicon replicates the community lexicon in its semantic content, but disregards the magnitude of evidence from previous discourses.

After each turn, DL_a and CL_a are updated in parallel by Roth-Erev reinforcement learning (Erev & Roth, 1998) according to the production/interpretation choices made by a and how successful the communication was. Although the same learning scheme is applied to both lexicons, the discourse lexicon is more sensitive to these updates since the magnitude of evidence for each concept-expression pair tends to be smaller. The evolution of the discourse lexicon is thus sensitive to the communicative successes of a *particular* dialogue, while the accumulation of utility across dialogues causes the community lexicon to become more stable with each conversation. In our simulations we compare linguistic communities that make use of a discourse lexicon to those that only keep track of a community lexicon.

^aConditional probabilities are represented by row-normalized matrices. L_a^p is a $|C| \times |M|$ matrix where row $L_a^p[c]$ is the distribution of messages given concept c . Likewise, L_a^i , a $|M| \times |C|$ matrix, is a distribution of interpretations (concepts) given message m .

Alignment. Alignment measures how much agents' representations of a lexicon agree (i.e., the degree of actual, rather than subjective, lexical common ground). For a set of agents $A' \subseteq A$ and lexicon $(C|D)L$, alignment is defined by the Jensen-Shannon divergence (JSD) of the agents' respective representations:^b

$$\text{Alignment}(A') = 1 - \text{JSD}(L_a) = 1 - H\left[\sum_{a \in A'} \frac{L_a}{|A'|}\right] + \frac{1}{|A'|} \sum_{a \in A'} H[L_a] \quad [1]$$

Under this definition, a community in perfect semantic alignment (i.e., every agent's representation of the given lexicon is identical) has an alignment of 1. Alignment is 0 when no two agents ever interpret any expression in the same way.

Ambiguity and Polysemy. In our model, the meaning of an expression m in a lexicon $(C|D)L_a$ is a discrete probability distribution over concepts; that is $L_a^i[m]$. We define ambiguity of m in L_a as the probability that two independent draws from $L_a^i[m]$ will result in distinct concepts:

$$\text{Ambiguity}(m, L_a) = 1 - \sum_{c \in C} L_a^i[m, c]^2 \quad [2]$$

The ambiguity of an agent's lexicon is defined as the average ambiguity of its expressions. The ambiguity of a lexicon as an abstract joint entity is simply the average ambiguity of all the relevant agents' representations of the lexicon.

An expression (or lexicon) that is perfectly deterministic has ambiguity 0. Maximum ambiguity is bounded by the size of the conceptual domain (but is always less than 1). An expression whose meaning is a perfectly flat distribution over all concepts has ambiguity $1 - \frac{1}{|C|}$.

A word is *polysemous* if its meaning is ambiguous among conceptually similar interpretations. We define polysemy as the probability that two different interpretations of m are similar (weighted by how similar they are):

$$\text{Polisemy}(m, L_a) = \sum_{c_1 \neq c_2 \in C} L_a^i[m, c_1] \cdot L_a^i[m, c_2] \cdot \text{sim}(c_1, c_2) \quad [3]$$

By definition, $0 \leq \text{Polisemy}(m, L_a) \leq \text{Ambiguity}(m, L_a) \leq 1$. One may think of polysemy as measuring how much ambiguity is due to ambiguity among similar concepts. Polysemy of an agent's lexicon is likewise defined as the average polysemy of its expressions and polysemy of the abstract joint lexicon is the average polysemy of all agents' representations.

^bHere H is Shannon entropy defined with the base $|A'|$ logarithm. This ensures that the Jensen-Shannon divergence (Lin, 1991) is bounded by 1 for communities of any size.

Communication Strategies. Agents may employ a *literal* strategy (LL), simply producing expressions and selecting interpretations according to the conditional probabilities encoded in their respective lexicons: $P(m | c, LL) = L^p[c, m]$ for production and $P(c | m, LL) = L^i[m, c]$ for interpretation. However, we also endow our agents with the possibility to employ a *semantic approximation* (SA) strategy. With this, we attempt to model situations such as those exemplified in the Introduction, where, for example, an agent wants to communicate a concept c , but there are no lexical items that express that concept especially well, so she instead utters some expression m because there is a concept c' similar to c such that $L^p[c', m]$ is high.

When choosing a message to express a concept c with semantic approximation, the speaker considers all the concepts each message m expresses and how similar they are to the intended concept.

$$P(m | c, SA) = \frac{\sum_{c'} \text{sim}(c, c') \cdot L^p[c', m]}{\sum_{m', c'} \text{sim}(c, c') \cdot L^p[c', m']} \quad [4]$$

Interpretation according to SA is defined similarly:

$$P(c | m, SA) = \frac{\sum_{c'} \text{sim}(c, c') \cdot L^i[m, c']}{\sum_{m', c'} \text{sim}(c, c') \cdot L^i[m', c']} \quad [5]$$

Which communicative strategy is being used may change from utterance to utterance within a single dialogue. We assume, however, that the current speaker always chooses the agents' communicative strategy.^c In our model, how likely a speaker is to use semantic approximation is proportional to how much worse than average her lexicon is at expressing her intention:

$$P(SA) = \min\left(0, 1 - |C| \sum_{m \in M} L[c, m]\right) \quad [6]$$

3. Experimental Setup & Results

We are interested in investigating the interplay between semantic approximation and the local common ground built up by a pair of agents over the course of a dialogue (encoded in the transient discourse lexicon), as well as how this interplay may shape the lexical resources of an overall linguistic community in the long term. To this end, simulations were run varying two parameters: communication strategy (whether agents were restricted to the literal strategy or could also use semantic approximation) and discourse lexicon (whether agents only kept track of the community lexicon or also made use of a local discourse lexicon).

^cHow coordination on a given strategy is achieved is beyond the scope of this paper, but recall (§1) that, for example, speakers may explicitly indicate their use of semantic approximation by hedging.

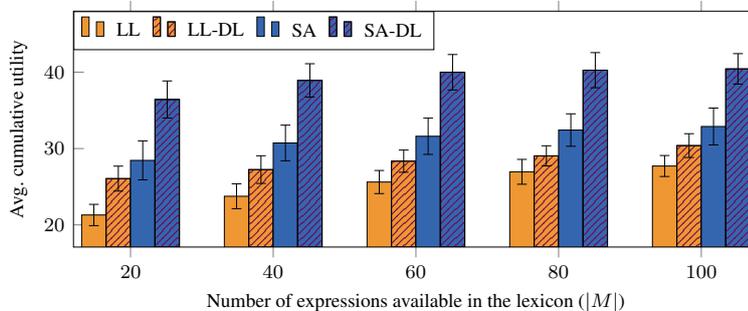


Figure 2. Average cumulative utility (as a percentage of total available utility) across 50 random concept domains after 1000 dialogues.

Settings. All four combinations (SA, LL, SA-DL and LL-DL) were simulated with $|M|$ ranging from 20 to 100 expressions (intervals of 20) on 50 random concept domains. Communities of 5 agents and concept domains of 100 concepts were used. We generate concept domains using the Holme-Kim growing random graph algorithm, which encourages clustering (Holme & Kim, 2002). That is, if c_1 and c_2 are similar and c_2 and c_3 are similar, it is more likely than average that c_1 and c_3 are similar as well. All simulations lasted for 1000 dialogues of 50 utterances each. Agents employed Roth-Erev reinforcement learning with a discounting rate of $\lambda = 0.99$.

The concepts speakers intend to express are typically non-uniformly distributed in a given conversation—concepts that have appeared once are likely to come up again in the same discourse than are concepts that have not been expressed. To model this intuition, each discourse is accompanied by a “context”^d (a probability distribution over concepts) which is itself drawn from a symmetric Dirichlet prior ($\alpha = 0.1$). Because the prior distribution is symmetric, no concept is more likely than any other to come up across discourses, but since the Dirichlet prior tends to produce (more so with lower α), in any given discourse, some concepts will tend to be expressed more than others.

Results. To measure communicative success, we consider a simulation’s cumulative utility; that is, the sum of the similarity between speaker intention and addressee interpretation over all turns in all dialogues (Figure 2). We find that the option to use semantic approximation consistently improves communicative success, but that this improvement is more pronounced when agents keep a discourse

^dWe assume that interlocutors do not exploit any knowledge of the prior probability of an intention in their interpretation and production strategies. In this way, the distribution over concepts that accompanies a discourse is different than in other models.

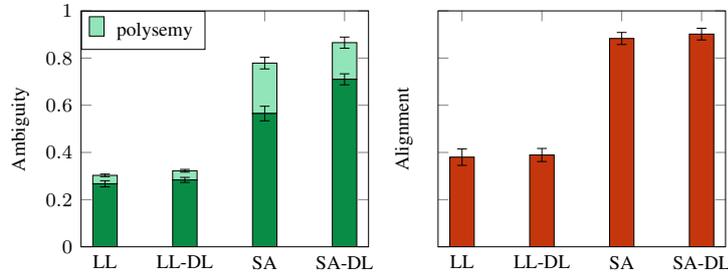


Figure 3. Characteristics of agents' joint community lexicon after 1000 dialogues (with $|M| = 40$). Polysemy is shown as the top portion of ambiguity (represented by the whole bar).

lexicon ($d > 5.01$ for all lexicon sizes with DL vs. $d < 3.47$ without DL).^e Furthermore, the discourse lexicon improves communicative success whether or not semantic approximation is used ($d > 1.41$ for all lexicon sizes and both strategies). We also note that using semantic approximation mitigates the disadvantage of having a smaller lexicon. For example, the effect on cumulative utility of going from 100 to 20 expressions is smaller if SA is used ($d = -1.79$ and -1.80 , for SA and SA-DL versus $d = -4.63$ and -2.70 , for LL and LL-DL respectively).

Figure 3 shows the effect of semantic approximation on community-level lexical alignment, ambiguity and polysemy. The use of SA results in higher alignment amongst agents, and it also leads to higher levels of ambiguity and polysemy in the joint (abstract) community lexicon. In both cases this effect is exaggerated for larger lexicons. However, when SA is used, the presence of a discourse lexicon reduces the proportion of polysemy ($p < 0.01$).

Finally, Figure 4 shows how agents' discourse lexicon compares to their community lexicon after the former has evolved over the course of a dialogue in terms of expected utility—the utility yielded by an additional hypothetical utterance by one of the interlocutors. At the end of each dialogue (after 50 utterances), we computed the expected utility of the interlocutors' discourse and community lexicons. We observe that overall the discourse lexicon has higher expected utility than the community lexicon ($p < 0.01$ for all dialogues after the 250th), but this advantage is greater when semantic approximation is allowed.

4. Discussion

Discourse-specific conventions allow agents to express concepts that have poor coverage in the community lexicon. We see, however, that the discourse lexicon is most helpful when semantic approximation is an option (Figure 2). In order for the discourse lexicon to tune its expressivity to a particular discourse, there

^eAll reported effect sizes (Cohen's d) accompany a dependent student's t-test with $p < 0.01$.

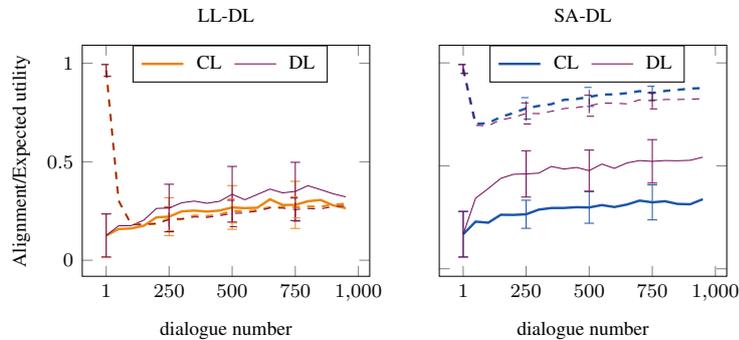


Figure 4. Expected utility (solid lines) and alignment (dashed lines) of the interlocutors' discourse lexicons (after a just-completed dialogue) versus their community lexicons.

must be some extra-lexical way of establishing those conventions not present in the community lexicon. Non-literal communication strategies such as semantic approximation fill this role. As in example (2) in the Introduction, a message-concept pairing is first suggested through semantic approximation, and then, when it is found to be successful, conventionalized for use in the remainder of the discourse, and possibly beyond.

In addition to creating opportunities for discourse-level lexicalization, semantic approximation has a profound effect on the long-term character of community-level lexical resources. For example, when semantic approximation is used, community lexicons are more ambiguous (Figure 3). Ambiguity is a paradox that has long puzzled linguists—it is commonplace in natural language despite the fact that it is intuitively detrimental to communicative success. One explanation for lexical ambiguity (Piantadosi et al., 2012) is that the re-use of expressions improves communicative efficiency, assuming that context is informative about meaning. But this answer doesn't explain some of the specific features of natural language ambiguity—particularly the prevalence of polysemy. The results of these simulations suggest yet another explanation for ambiguity in natural language: the use of semantic approximation (which is communicatively beneficial on a discourse-level) leads to polysemous *ad hoc* conventions, which, with enough use, are lexicalized at the community level.

Semantic approximation is not the only method of non-literal semantic coordination in natural language, however. In addition to further investigation of semantic approximation, future work should investigate how other local communication strategies such as metaphor contribute to discourse-level communicative success and influence the development of community-level semantic consensus.

References

- Brennan, S., & Clark, H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology*, 22(6), 1482–1493.
- Chaix, P. Y., Barry, I., & Duvignau, K. (2012). Semantic approximation in SLI and normal development. *Revue Française de Linguistique Appliquée*, XVII(2),
- Clark, H. H. (1996). *Using language*. Cambridge University Press.
- Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 848–881.
- Holme, P., & Kim, B. J. (2002). Growing scale-free networks with tunable clustering. *Physical Review E*, 65.
- Juba, B., Kalai, A. T., Khanna, S., & Sudan, M. (2011). Compression without a common prior: An information-theoretic justification for ambiguity in language. In *Innovations in Computer Science* (pp. 79–86).
- Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press.
- Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1), 145–151.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280–291.
- Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Pragmatics* (Vol. 9).

THE EVOLUTION OF IM/POLITENESS

MONIKA PLEYER

*Department of English, Universität Heidelberg, Kettengasse 12, Heidelberg, D-69117,
Germany, monika.pleyer@ix.urz.uni-heidelberg.de*

MICHAEL PLEYER

*Department of English, Universität Heidelberg, Kettengasse 12, Heidelberg, D-69117,
Germany, pleyer@stud.uni-heidelberg.de*

Im/politeness is a fundamental feature of human language and communication. However, there is hardly any research on the evolution of im/politeness and the cognitive and social factors underlying its emergence. In this paper we argue that the evolution of politic, polite, as well as impolite behaviour is an important research question for language evolution research. We investigate the evolutionary foundations of im/politeness, present an evolutionary model of the emergence of im/politeness and discuss the evolutionary functions of im/politeness. In this way, we illustrate that investigating im/politeness from an evolutionary perspective can make significant contributions to our understanding of the evolution of pragmatic competencies, language, and also im/politeness research in general.

1. Introduction

There has been a growing consensus that the dimension of pragmatics is of fundamental importance to accounts of language evolution and has not been addressed in enough detail so far (e.g. Scott-Phillips 2014: 141; Tomasello 2008; Wacewicz, Żywiczyński, & McCrohon 2014: 82). Among this growing number of researchers investigating the evolution of pragmatics, Żywiczyński (2012) and Wacewicz et al. (2014) have stressed the central role of politeness in pragmatics and language use and have drawn attention to the importance of studying politeness from an evolutionary-ethological perspective. In addition, Wacewicz et al. (2014: 92) also mention impoliteness as an important topic for language evolution research in its own right. Among im/politeness researchers, Leech (2014: 26) has posed the question “How has the evolution of politeness – so widely observed in human societies and languages – come about?” and identified it as a key component of an explanation of politeness. Im/politeness researchers

so far have paid little to no attention to the question of the evolution of im/politeness, however.

In this paper, we will make a contribution towards this important research question. We take as our point of departure the consensus in pragmatics research that im/politeness is a fundamental and universal feature of language use that is present in all interactions (e.g. Sell 2005: 114-115). Moreover, one of the problems of previous discussions of the evolution of politeness is that they treated politeness and impoliteness as distinct phenomena. However, here we adopt an integrated model that sees politic, polite and impolite behaviour as scalar (see Watts 2003) and based on the same cognitive and evolutionary foundations. In this way, we will sketch an integrated evolutionary account of the foundations, emergence, and functions of im/politeness in language.

2. Cognitive and Evolutionary Foundations of Im/Politeness

From a cognitive perspective an understanding of im/politeness is based on a number of sociocognitive skills that allow humans to negotiate and coordinate social relations in interactive contexts in a given speech community. Like other coordinative and collaborative interactive skills it rests on the *shared intentionality infrastructure* (Tomasello 2008), that is the ability to attend jointly to something with others and have shared goals, intentions, and a common ground that underlies the negotiation of a shared perspective on a situation. These skills arise quite early and a nascent understanding of joint attentional states and others' perspectives is evident even in prelinguistic infants (Tomasello 2008). These capacities lay the groundwork for an emerging understanding of politic and im/polite behaviour as based on conventions and social perspectives. In addition, children have a high motivation to align themselves with others, be like, and cooperate with them (Tomasello 2009, 2014). This represents another important foundation for acquiring and internalizing politic and im/politeness triggers. In the process of socialisation, children acquire a concept of other interactants' social face and face needs (Leech 2014: 26). This process also depends on the capacities for cooperation, alignment, perspective-taking and the emerging capacity for Theory of Mind – that is the capacity to attribute complex mental states to others. However, the acquisition of im/politeness is not only tied to these kinds of interactional perspective-taking and cooperative proclivities. Crucially, they rest on an understanding of conventions and normativity that is operative in cooperative contexts (cf. Waciewicz et al. 2014).

Tomasello (2014: 144) proposes that before age three, young children's social cognition operates on a second-personal, interactive mode, and that "they

do not fully understand how such things as language, artifacts, and social norms work as conventional creations” based on group-agreement. This mode is termed *joint intentionality* by Tomasello. For an understanding of conventions, norms, and in turn also for an understanding of im/politeness, children need to develop what Tomasello calls *collective intentionality*, an understanding that conventions are based on collective agreements that guide and coordinate normative behaviour. As Tomasello argues, this understanding starts developing around children’s third birthday and is evident in a number of behaviours, for example in their concern for how they are socially evaluated by others, their active management of impressions they make on others, and especially their enforcement of “social norms on others even when they are not personally involved or affected in any way, often using normative language about what one should or should not do in general” (Tomasello 2014: 87; cf. Haun & Tomasello 2011; Rakoczy, Warneken & Tomasello 2008; Schmidt & Tomasello 2012).

From an evolutionary perspective, it is interesting that chimpanzees do not enforce and monitor norms in this way and do not punish others for acts that are not done to them directly (Tomasello 2014: 75; Riedl, Jensen, Call & Tomasello 2012). In contrast to preschool children, chimpanzees are also not very interested in what impression they make on others (Engelmann, Herrmann & Tomasello 2012) and do not show a bias for conformity as children do (Tomasello 2014: 75; Haun, Rekers & Tomasello 2014). Polite behaviour and im/politeness as a social activity also depend on such judgements and understandings of normative standards. Thus, the capacities mentioned above represent the cognitive and evolutionary foundations not only of normative cognition, but also for the ontogenetic and phylogenetic emergence of understanding polite behavior and im/politeness.

3. An Evolutionary Model of Im/Politeness

As was shown in the previous section, certain cognitive and evolutionary foundations of language must therefore be present as a basis for pragmatic language use.

But what exactly are the minimal requirements for the evolution of an im/politeness-ready brain and for individuals to communicate in an im/polite manner? We argue that this requires three concepts: 1) face, “the positive social value a person effectively claims for himself [sic!] by the line others assume he has taken during a particular interaction” (Goffman 1967: 5); 2) norms of behaviour, based on sociocognitive skills such as perspective-taking, an understanding of normativity, Theory of Mind and collective intentionality; and

3) protolinguistic capacities for meaning construal enabling flexible and intentional use of meaningful items in interactive contexts.

Face in Goffman's view is an individual as well as a group phenomenon, as it is on loan, accepted and perpetuated during a given interaction. There are two prerequisites for individuals to express and maintain face: First, a mutual understanding that the other has a face and is an intentional being, as expressed in the sociocognitive infrastructure underlying shared intentionality and perspective-taking. Second, the acceptance of face as being in operation in an interaction is a prerequisite to successful communication (Goffman 1967). While face, as well as holding and expressing im/polite beliefs, seems to be universal, its constituents and characteristics can and do vary in particular cultures and contexts (see e.g. Matsumoto 1988; Culpeper et al. 2010, 2014).

Humans are group-minded beings, and as such "come together around mutual engagement in an endeavor. Ways of doing things, ways of talking, beliefs, values, power relations-in short, practices emerge in the course of this mutual endeavor" (Eckert & McConnell-Ginet 1992: 464). Within these communities of practice and in the group's common ground, emerging linguistic conventions of performing and evaluating social roles are shared and become normatively grounded in a community standard (Tomasello 2014: 92; 114). Current research widely accepts that interactants' behaviour is evaluated according to these interactional norms (Eelen 2001: 42-43); this ties into the social norm view of politeness (Fraser 1990: 220).

For the development of pragmatic competence, this means that within the community, a set of linguistic means is developed which speakers use to express the appropriate behaviour as per the community standard. This default behaviour which abides by the Conversational Contract (Fraser & Nolen 1981; Fraser 1990) is unmarked and likely goes unnoticed. As such, it is likely to have developed first and to have been conventionalised quicker.

Any behaviour that goes beyond this appropriate or 'politic' (Watts 2003) behaviour is salient. Depending on usage, it can be positively or negatively marked and thus be open for an interpretation as im/polite (see Eelen 2001, Watts 2003, Culpeper 2011, Locher & Bousfield 2008). Some of these salient behaviours are selected more often in certain contexts. For impoliteness, for instance, these are contexts in which negative emotions or negative consequences occur, and contextual expectations are infringed upon (see Culpeper 2011: 23). As these behaviours are commonly selected in negatively marked contexts, they become conventionalised. A knowledge of these (non-) conventionalised im/politeness triggers (Culpeper 2015) forms an integral part of the pragmatic competence of a particular language user.

Children acquire these preexisting representational systems and abstract them from instances of actual im/polite and politic language use in context (Tomasello 2014: 139); these systems become part of the individual speaker's habitus (Bourdieu 1991) and are passed on to and stabilised in further generations. These systems are thus subject to cultural transmission and historical development within the communities of practice they are embedded in (Mesoudi 2011). For this reason, evaluations of behaviours as politic or im/polite can change markedly over time and can vary strongly in different cultures (Nevala 2010).

As both impoliteness and politeness rely on the same cognitive mechanisms, they should not be treated as two distinct phenomena, but as the two end points of a scale. An integrated model of im/politeness that includes a notion of a default or politic/appropriate behaviour therefore better captures this understanding of im/politeness as scalar (see Watts 2003). On the positively marked end of the spectrum, politeness is behaviour that caters to the other's face. If the line the speaker takes during an interaction is accepted or boosted, she will feel good and wish to perpetuate this line in further interactions. Similarly, she will choose triggers to support the hearer's face, as both are mutually aware that face loss will lead to negative emotions (e.g. shame, confusion, anger). On the other, negatively marked end, impoliteness constitutes "behaviour that is face-aggravating in a particular context" (Bousfield & Locher 2008: 3), or is understood by the target to be in disregard of face concerns. As such, it comprises any behaviour that violates contextual norms, is evaluated negatively, and has negative consequences for at least one participant (cf. Culpeper 2011: 23).

Im/politeness is thus a universal feature of human communication. The expression of im/politeness, however, and the number and characteristics of im/politeness triggers are specific to culture and context, as well as to specific points in time (see e.g. Ide 1989; Sell 2005).

4. The Evolutionary Function of Im/Politeness

Once a community standard of doing im/politeness and catering to face has been established, it can be used for various interpersonal goals, e.g. to gain praise, or attain a better social position (Goffman 1967: 24).

Speakers use their capacities for joint attention and joint goal formation to create and share perspectives which have to be negotiated in the group (Tomasello 2014: 3). Im/politeness is tied to this negotiation process of social roles in that it offers ways of establishing and negotiating perspectives on the

social standing of participants in social situations. Thus it is a way of coordinating perspectives and social relations in a collective intentionality context. As such it forms a basis for “more abstract cultural practices” that are “structured – indeed, created – by agreed-upon social conventions and norms” (Tomasello 2014: 3; see Tomasello 2009).

Leech (2014: 27) suggests that from an evolutionary perspective “politeness, building on empathy [...] and theory of mind [...], is one of the forms of reciprocity” that enable humans to live in stable and cohesive societies. In showing concern for the face, wants and rights of each individual, members of a group can use politeness to attain and maintain group coherence. Meeting the face needs of an individual thus has immediate benefits in terms of group cohesion, group survival, and individual fitness. This proposal ties in with theories that discuss the evolution of language in the context of maintaining and managing group cohesion and social relations (e.g. Dunbar 2003). Politeness also offers ways of reputation management. As long as the line the speaker has taken is accepted and boosted over multiple interactions, politeness helps secure one’s social standing within the group, contributing to the establishment and protection of a stable group hierarchy.

Impoliteness, on the other hand, should not be conceptualised as a handicap or failed politeness, since it can be used intentionally as a means to achieve individual goals (Sell 2005: 115-116). For a protolinguistic community, two ways of using impoliteness are of interest.

First, a producer uses a coercive action that clashes with the target’s interests or restricts the target to gain benefits or reinforce existing ones (Culpeper 2011: 226-227); this so-called coercive impoliteness can be used to reduce the target’s symbolic power (Bourdieu 1991) by infringing upon their Equity Rights (Spencer-Oatey 2002) or negative face. So for instance “[I]n aggressive utterances the winner [...] demonstrates that as interactant he can handle himself better than his adversaries” (Goffman 1967: 25). The producer thus uses impoliteness in situations of power imbalance to gain (social) benefits, e.g. to appear powerful by insulting the target, to get power over actions, or to get power in conversation, for instance by managing the floor. In early human evolution, such capacities were certainly integral for an individual’s fitness within social structures characterized by a high degree of collaboration, cooperation, but also competition and social negotiation.

A second, different form of impoliteness might even predate coercive use. This “affective impoliteness is the targeted display of heightened emotion, typically anger, with the implication that the target is to blame for producing that negative emotional state” (Culpeper 2011: 223). Impoliteness is expressed

purposefully to harm a target who hurt the speaker (or her self-esteem). It is thus used to manage speaker aggression and express one's own perspective on the behaviour of a target (an individual or group), situation or event. Hence it can be used to manage in-groups and out-groups. For instance, impolite beliefs can be expressed with the aim of excluding the target from the in-group; further, a group of individuals can be construed as other, i.e. an out-group who lacks participation rights in the in-group community. For the in-group, this might entail evolutionary benefits, e.g. when resources are sparse. Attacks on and loss of face can have severe consequences such as loss of social status and ostracism. These in turn can have significant detrimental effects on evolutionary fitness. The evolutionary importance and fragility of face in the social domain is also underscored by the close tie between central behaviour-motivating emotive categories such as joy, surprise, anger and sadness on the one hand, and im/politeness and rapport management on the other (Spencer-Oatey 2005: 116).

Overall, then, politeness and impoliteness are ways of taking, setting, and sharing perspectives. In particular, they function as attempts to coordinate and create views of an interlocutor's face, as well as attempts to manage group coherence and group membership. As such, im/politeness is an integral mechanism involved in the development of culture, which is essentially a process of social coordination (Tomasello 2014).

5. Conclusion

On the basis of research into the evolution of language, group mindedness, and shared intentionality we proposed prerequisites for an im/politeness-ready brain: concern for interactants' face, a shared set of norms, and capacities underlying the goal-directed use of proto-linguistic items in the context of social coordination. We used this as a basis for an evolutionary model of im/politeness with an emerging standard default politic mode of interaction and conventionalised ways of using im/politeness triggers. The model is coherent with one proposed in current im/politeness research and the ways in which modern speakers negotiate im/politeness in interaction. While triggers differ in each community, the expression of im/politeness is unique to all human societies. Early forms of politic and im/polite behaviour had evolutionary benefits for protolinguistic hominins in terms of the maintenance and management of growing groups and social hierarchy, as well as group and individual benefits. On this view, impoliteness is not a handicap or a failure to do politeness, but an important feature in the evolution of pragmatics and language in its own right that deserves further scientific exploration.

References

- Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge: Polity Press.
- Bousfield, D., and M. Locher (eds.) (2008). *Impoliteness in Language*. Berlin: Mouton de Gruyter.
- Culpeper, J. (2015). Impoliteness strategies. In A. Capone and J. L. Mey (eds.), *Interdisciplinary Studies in Pragmatics, Culture and Society* (pp. 421-445). Heidelberg [et al.]: Springer.
- Culpeper, J. (2011). *Impoliteness. Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Culpeper, J., Schauer, G.A., Marti, L., Mei, M. & Nevala, M. (2014). Impoliteness and emotions in a cross-cultural perspective. *SPELL: Swiss Papers in English Language and Literature*, 30, 67-88.
- Culpeper, J., Marti, L., Mei, M., Nevala, M., & Schauer, G.A. (2010). Cross-cultural variation in the perception of impoliteness: A study of impoliteness events reported by students in England, China, Finland, Germany and Turkey. *Intercultural Pragmatics*, 7(4), 597-624.
- Dunbar, R. I. M. (2003). The social brain: Mind, language, and society in evolutionary perspective. *Annual Review of Anthropology*, 32, 163-181.
- Eckert, P. and S. McConnell-Ginet. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology*, 21, 461-488.
- Eelen, G. (2001). *A Critique of Politeness Theories*. Manchester: St. Jerome Publishing.
- Engelmann, J., E. Herrmann, and M. Tomasello. (2012). Five- year olds, but not chimpanzees, attempt to manage their reputations. *PLoS ONE*, 7(10), e48433.
- Fraser, B. (1990). Perspectives on politeness. *Journal of Pragmatics*, 14, 219-236.
- Fraser, B., and W. Nolen. (1981). The association of deference with linguistic form. *International Journal of the Sociology of Language*, 27, 93-109.
- Goffman, E. (1967). *Interaction Ritual. Essays on Face-to-face Behavior*. Garden City: Doubleday.
- Haun, D. B. M., and M. Tomasello (2011). Conformity to peer pressure in preschool children. *Child Development*, 82, 1759-1767.
- Haun, D. B. M., Rekers, Y. and Tomasello, M. (2014). Children conform to the behavior of peers; other great apes stick with what they know. *Psychological Science*, 25(12), 2160-2167.
- Ide, S. (1989). Formal forms and discernment. Two neglected aspects of universals of linguistic politeness. *Multilingua*, 8(2-3), 223-248.
- Leech G. (2014). *The Pragmatics of Politeness*. Oxford: Oxford University Press.

- Matsumoto, Y. (1988). Reexamination of the universality of face: politeness phenomena in Japanese. *Journal of Pragmatics*, 12(4), 403-426.
- Mesoudi, A. (2011). *Cultural Evolution: How Darwinian Theory can Explain Human Culture and Synthesize the Social Sciences*. Chicago: University of Chicago Press.
- Nevala, M. (2010). Politeness. In A.H. Jucker & I. Taavitsainen (eds.), *Historical Pragmatics* (pp. 419-450). Berlin: de Gruyter.
- Rakoczy, H., F. Warneken, and M. Tomasello (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, 44, 875-881.
- Riedl, K., K. Jensen, J. Call, and M. Tomasello (2012). No third-party punishment in chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 14824-14829.
- Schmidt, M., and M. Tomasello (2012). Young children enforce social norms. *Current Directions in Psychological Science*, 21, 232-236.
- Scott-Phillips, T.C. (2014). *Speaking our Minds: Why Human Communication is Different, and How Language Evolved to Make it Special*. Houndmills: Palgrave Macmillan.
- Sell, R. D. (2005). Literary texts and diachronic aspects of politeness. In R. J. Watts, S. Ide and K. Ehlich (eds.), *Politeness in Language: Studies in its History, Theory and Practice* (pp. 109-130). Berlin: Mouton.
- Spencer-Oatey, H. (2005). (Im)Politeness, face and perceptions of rapport: Unpackaging their bases and interrelationships. *Journal of Politeness Research*, 1(1), 95-119.
- Spencer-Oatey, H. (2002). Managing interpersonal rapport: Using rapport sensitive incidents to explore the motivational concerns underlying the management of relations. *Journal of Pragmatics*, 34, 529-545.
- Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2009). *Why we Cooperate*. Cambridge, MA: MIT Press.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Wacewicz, S., P. Żywiczyński, and L. McCrohon (2014). Linguistic politeness from an ethological perspective: Theoretical questions and empirical issues. *Theoria et Historia Scientiarum*, 11, 81-98.
- Watts, R. J. (2003). *Politeness*. Cambridge: Cambridge University Press.
- Żywiczyński, P. (2012). An attempt at a proxemic description of politeness from the ethological-evolutionary perspective. *Theoria et Historia Scientiarum*, 9, 71-98.

WHAT KIND OF GRAMMAR DID EARLY HUMANS (AND NEANDERTHALS) COMMAND? A LINGUISTIC RECONSTRUCTION^a

LJILJANA PROGovac

*Wayne State University
Detroit, USA
progovac@wayne.edu*

Here I pursue a linguistic reconstruction of the earliest stages of grammar, following a precise syntactic theory. This reconstruction arrives at the initial stages of grammar which are in consort with crosslinguistic variation attested in the expression of various syntactic phenomena, including transitivity and tense marking. Interestingly, in making an argument for the antiquity of language, Dediu & Levinson (2013, p. 11) express their hope “that some combinations of structural features will prove so conservative that they will allow deep reconstruction.” I propose that the earliest stages of syntax as reconstructed here provide just such a conservative platform from which all the subsequent variation could arise, and which could have been commanded also by our cousins and the common ancestor. The reconstruction is at the right level of granularity to exclude some hypotheses regarding the hominin timeline, and to support others. It leads to specific and testable hypotheses which can be explored in e.g. anthropology, neuroscience, and genetics.

1. Introduction

One of my goals is to find connections and common ground between the recent proposals in Dediu and Levinson (2013) (D&L) regarding the antiquity of language and the syntactic reconstruction of syntax proposed in Progovac (2015, and previous work, e.g. 2009). I believe that this is exactly the kind of interdisciplinary dialogue that is necessary in order to shed light on language evolution: we need to rely on linguistic theories and linguistic variation to reconstruct the stages of language evolution, and we also need to cross-fertilize these findings with various other types of evidence coming from e.g. genetics, skeletal morphology, cultural artifacts, neuroscience, etc. Linguistic reconstructions are needed because the data from genes and bones alone cannot

^a I am grateful to Dan Dediu (personal communication 2015), for very useful feedback on the first draft of this paper.

tell us with any precision what kind of language e.g. early humans or Neanderthals may have commanded, and thus do not lead to further and more precise hypothesizing. D&L (p. 11) are aware of this, and express their hope “that some combinations of structural features will prove so conservative that they will allow deep reconstruction.” I propose that the earliest stages of syntax reconstructed in Progovac (2015) provide just such a conservative grammatical platform that D&L are looking for.

2. What Can the Bones and Genes Tell us?

D&L review a number of recent findings suggesting that at least *H. heidelbergensis* had some form of language, based on the comparative evidence among its descendants: *H. sapiens*, Denisovans, and Neanderthals, as also suggested by Finlayson (2009, 116) (but see Berwick, Hauser, & Tattersall, 2013, for criticism of this view). According to D&L (p. 10), “language as we know it must then have originated within the ~1 million years between *H. erectus* and the common ancestor of Neanderthals and us.” D&L (p. 5) conclude that Neanderthals and Denisovans “had the basic genetic underpinning for recognizably modern language and speech, but it is possible that modern humans may outstrip them in some parameters (perhaps range of speech sounds or rapidity of speech, *complexity of syntax*, size of vocabularies, or the like). If we adopt the widely accepted uniregional hypothesis regarding human origins, then the linguistic reconstruction I follow leads to the conclusion that human grammars are at least more complex (as they are hierarchical) than any grammar commanded by the common ancestor, *H. heidelbergensis*, as explained below.

In addition to genetics, D&L also review evidence from the skeletal morphology, the morphology of the vocal tract, infant maturation, Broca’s area, brain size, and cultural artifacts, and they conclude that all the evidence is consistent with their proposal. If this is on the right track, then the question to be addressed is the following: what kind of grammar characterized *H. heidelbergensis* or Neanderthals? Moreover, one also needs to hypothesize about what kinds of pressures contributed to the innovation and evolution of more complex grammars, and how our brains evolved to be able to support first the simpler grammars, and then the more complex grammars. The only way to arrive at specific and testable hypotheses regarding these questions is to pursue a reconstruction based on a linguistic theory and on linguistic variation. Neuroimaging experiments can then be designed to test such specific hypotheses (as in e.g. Progovac, Ofen, Crabtree, Angell, and Liddane, In Preparation).

3. What Can Linguistic Theories Contribute: Reconstructing Early Stages of Grammar

The reconstruction proposed in Progovac (2015, and work cited there) is based on the influential framework of Minimalism and its predecessors, as outlined in e.g. Chomsky (1995). In this syntactic framework, modern sentences and phrases are treated as hierarchical constructs, consisting of several layers of structure, built in a binary fashion. The following is the partial hierarchy of projections involved in the construction of a typical sentence:

(1) TP > vP > SC/VP

[where TP is a Tense Phrase layer; vP a transitive (higher) verb Phrase; VP the basic Verb Phrase, and SC a Small Clause. For the relevance of using both SC and VP as labels for the inner layer, see Progovac (2015).]

To derive a sentence such as *Deer will eat fish*, we first assemble the most basic, inner layer, the SC/VP *eat fish*. At this point in the derivation, it is still not determined if the fish is going to be the eater or the eaten, and thus there is really no subject/object differentiation. On the other hand, superimposing the transitivity layer (vP) enables the grammatical differentiation between subjects and objects (as in *Deer eat fish*; *Eat fish (by) deer*). The TP layer, in this case headed by *will*, is then superimposed over the vP layer, to create three layers of syntactic structure. Sentences and phrases in this framework can exhibit additional layers of structure.

Importantly, the layer upon which the whole sentence rests is the inner, foundational, SC (*eat fish*) layer, which is reconstructed as the initial stage of grammar. The logic behind the proposed reconstruction is straightforward: while VP/SC can be composed without a vP or a TP layer, a vP or a TP can only be constructed upon the foundation of a VP/SC. Moreover, while imposing an additional layer of structure upon the foundational SC necessarily results in a hierarchical, layered construct, the SC itself can be a flat, headless, paratactic creation. That is exactly the kind of proto-grammar that this reconstruction arrives at: a flat, tenseless, intransitive two-slot mold, consisting of just one verb-like and one noun-like element, and in which the subject/object distinction could not be expressed grammatically.

As we begin to wonder if this kind of grammar is feasible at all, consider that we find “living fossils” (in the sense of Jackendoff, 2002) of such grammars in

various constructions in present-day languages. One example would be verb-noun compounds, such as English: *cry-baby*, *kill-joy*, *tattle-tale*, *turn-coat*, *scatter-brain*, *tumble-dung* (insect); Serbian *cepi-dlaka* (split-hair; hair-splitter), *ispi-čutura* (drink-up flask; drunkard), *vrti-guz* (spin-but; fidget), *jebi-vetar* (screw-wind; charlatan); and Twi (spoken in Ghana) *kukru-bin* (roll-feces; beetle). These are essentially small clauses created by two-slot grammars, with just one verb and one noun, without a possibility for distinguishing subjects from objects. If we compare e.g. *turn-table* and *turn-coat*, we observe that the first describes a table that turns (table is subject-like), and the second describes somebody who turns his/her coat, metaphorically speaking (coat is object-like). But, grammatically speaking, these two compounds are identical.^b

4. What Can Language Variation Tell us?

The unspecified role of the noun in this two-slot grammar can be characterized as the absolutive role, as such roles are not directly sensitive to the subject/object distinction. Absolutive-like roles are found not only in languages that are classified as ergative-absolutive, but probably in all languages, in some guise or another, including in the compounds above. Human languages in fact differ widely with respect to how they express transitivity, and this reconstructed absolutive-like basis provides the common denominator, the foundation from which the variation can arise. The reconstruction offered above is thus synergistic with the findings in linguistic typology, the field concerned with language variation. What makes this synergy possible is the precision of the reconstruction, and the consideration of specific linguistic data.

In ergative-absolutive languages, such as e.g. Tongan (spoken in Tonga; Tchekhoff, 1973), there is special case marking for an additional, second argument, typically agent, and this case marking is called ergative, resulting in structures comparable to: *Eat (by) deer_{ERG} fish*. On the other hand, intransitive structures comparable to *Eat fish* in Tongan are vague/unspecified with respect to whether the fish (absolutive) is eating or being eaten (see also Gil, 2005, for Riau Indonesian). In nominative/accusative languages, such as Serbian, there is

^b In addition to being illustrative of a most rudimentary grammar, it is intriguing that verb-noun compounds in many languages specialize for derogatory reference and insult when referring to humans, and are often crude or obscene. In medieval times alone, thousands of such compounds were used, certainly many more than nature needs. Such abundance, indeed extravagance, is usually associated with display and sexual selection, the force that has also created the peacock's tail.

special case marking for objects (accusative), and here one encounters a transitive structure of the kind: *Deer eat fish*._{ACC} or *Fish*._{ACC} *eat deer*. There are also languages which make use of the so-called serial verb constructions, where two small clauses get strung together to express semantic transitivity. One illustrative example is *Dog catch, fish eat*, from Anyi-Sanvi, Niger-Congo, meaning roughly “dog catches it: fish gets eaten” (Van Leynseele, 1975, 191-2).

It is of great significance that these different strategies for expressing transitive meanings all share the common ground or foundation, and all can be reduced to the initial absolutive-like small clause. The different strategies are just different solutions to the same problem: the problem of having only a two-slot grammar, able to fit a verb and only one noun, but desiring to describe a transitive event, which requires two nouns. Given the reconstruction results and the attested variation in transitivity, it is now possible to formulate specific hypotheses regarding the hominin timeline, as well as regarding the timing of the emergence of different stages of grammar. Again, what makes this possible is the precision of the syntactic reconstruction considered here.

5. The Potential and the Limits of Two-Slot Grammars

Even though disarmingly simple, the beauty of this two-slot grammar is in its ability to combine not only two words, but also two flat small clauses, paratactically, as illustrated in the following AB-AC formulaic “living fossils” from English and Twi (spoken in Ghana) (see also serial verb constructions mentioned in the previous section):

- (2) Monkey see, monkey do. First come, first serve.
Come one, come all. Card laid, card played.
Like father, like son. So far, so good. Easy come, easy go.
- (3) Wo dua, wo twa. (You sow, you reap)
Wo hwehwea, wo hu. (You seek, you find)

Just like the compounds of Section 3, these two-by-two formulae can support an abundance of tokens, demonstrating that even these simple grammars have an amazing creative and expressive potential.^c

^c Especially rich in such paratactic AB AC formulae is Hmong, spoken in China and northern Southeast Asia (Martha Ratliff, p.c., 2015). Of note is, perhaps, that such symmetric, parallel combinations would have been easy to fit onto simple melodies, and to develop musical protolanguage from. Such paratactic structures rely on prosodic glue to hold them together, and if

What all human languages and constructions undoubtedly have in common is the paratactic platform, that is, the ability to combine two words or two small clauses paratactically, essentially the properties of the reconstructed flat, binary, two-slot stage. All the complex hierarchical phenomena, including transitivity and subordination, have alternative routes, as well as precursors, in parataxis (Progovac, 2015). This is therefore a deep, conservative property of human language, the foundation upon which all else rests.

6. Drawing some Conclusions about the Grammatical Abilities of our Ancestors

There are certain scenarios for the evolution of grammar/syntax that are inconsistent with the approach pursued here, which means that this reconstruction is at the right level of granularity to engage the questions regarding the hominin timeline. For example, a great degree of crosslinguistic variation in how human languages build upon the foundational paratactic stage suggests that the hierarchical stage did not emerge in all its complexity and in a uniform fashion only once (in Africa), but instead multiple times, and independently, either within Africa, or after the dispersion from Africa. If it had emerged only once, before *H. sapiens* spread out, it would be difficult to explain why there is so much variation across languages of the world in how they express transitivity (by ergative, accusative, serial verb, or other means), or in whether or not they grammaticalize tense/aspect/mood systems, to name just some parameters of variation.

Under the uniregional hypothesis scenario, this reasoning leads to the conclusion that *H. heidelbergensis*, our common ancestor with Neanderthals and Denisovans, did not command hierarchical syntax, but most probably only the basic, paratactic, two-slot platform. This would be consistent with the slightly smaller size of the *H. heidelbergensis* brains, in comparison to either humans or Neanderthals. Neanderthals would have, in that case, inherited this paratactic grammar, but could not have inherited any hierarchical grammar from *H. heidelbergensis*. Of course, there always remains the possibility that Neanderthals developed their own kind of hierarchical syntax, or some other complexity, independently. On the other hand, Neanderthals could have stayed

there was musical protolanguage at any point in human evolution, then it would have been most useful in these earliest stages of grammar.

with the grammar they inherited from *H. heidelbergensis*, the paratactic two-slot grammar. Even though grammatically simple, this kind of grammar has an amazing creative potential for expressing a variety of meanings (see Section 5). If it was there at that juncture, it would have allowed *H. heidelbergensis* and Neanderthals, among many other communicative opportunities, to hurl insults at each other in the form of flat compounds (e.g. *crybaby*, *hunchback*), as well as to express eternal wisdoms and observations in the *form* of AB AC formulae (e.g. *You seek, you find*; *Monkey see, monkey do*).

Consistent with these considerations, it is likely, even though not certain, that the paratactic proto-syntax stage already characterized the *H. heidelbergensis* species, which would place the emergence of the proto-syntactic stage to at least as far as half million years ago. In fact, my proposal also cannot exclude the possibility that *H. erectus* also had some form of proto-syntax, especially considering that their brain doubled in size relative to that of the Australopithecus, who lived sometime between 4 million years ago and 2 million years ago. There was nothing else at that juncture that would have required as much brain capacity as the early stages of language, accompanied by an increase in expressive abilities and vocabulary size.

However, the linguistic considerations explored here, as they stand now, are not capable of choosing between the uniregional and multiregional hypotheses about human origins. It has been established that *H. erectus* traveled out of Africa around 1.7 million years ago, spreading to Europe and Asia. According to the much less accepted multiregional hypothesis, the local *H. erectus* populations in Africa, Asia, and Europe differentiated into *H. sapiens* independently, by a process of parallel evolution, as well as some admixture among the populations (e.g. Stone and Lurquin, 2007; Finlayson 2009). If this hypothesis turns out to be correct, or a weaker version of it (Harris 2015, 163-164), then, under my approach, one would have to say that *H. erectus*, prior to the migrations out of Africa, already commanded the foundational paratactic grammar, and that the more complex hierarchical grammars emerged separately in different geographical locations, after the dispersion. On this scenario, the hierarchical grammars could have emerged much earlier than with the uniregional hypothesis, given that the dispersion took place much earlier, around 1.7 million years ago.^d

^d There may be another scenario for the timeline, which would also allow for an earlier timing for hierarchical syntax. Namely, it is possible that hierarchical syntax emerged independently among

References

- Berwick, R. C., Hauser, M. D., & Tattersall, I. (2013). Neanderthal language? Just-so stories take center stage. *Frontiers in Psychology* 4. doi: 10.3389/fpsyg.2013.00671
- Tchekhoff, C. (1973). Some verbal patterns in Tongan. *The Journal of the Polynesian Society*, 82.3, 281-292.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Dediu, D., & Levinson, S. C. (2013). On the antiquity of language: The reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology*, 4, 397. doi: 10.3389/fpsyg.2013.00397.
- Finlayson, C. (2009). *The Humans who Went Extinct: Why Neanderthals Died out and we Survived*. Oxford: Oxford University Press.
- Gil, D. (2005). Isolating-monocategorial-associational language, in H. Cohen and C. Lefebvre (eds.), *Handbook of Categorization in Cognitive Science* (pp. 347–379). Amsterdam: Elsevier.
- Harris, E. E. (2015). *Ancestors in our Genome: The New Science of Human Evolution*. Oxford: Oxford University Press.
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Progovac, L. (2009). Sex and syntax: Subjacency revisited. *Biolinguistics* 3.2-3: 305-336.
- Progovac, L. (2015). *Evolutionary Syntax*. Oxford: Oxford University Press.
- Progovac, L., Ofen, N., Crabtree, M. R., Angell, W., & Liddane, R. In preparation. Neural correlates of syntax and proto-syntax: An fMRI study. *Journal of Neurolinguistics*.
- Stone, L., & Lurquin, P. F. (2007). *Genes, Culture, and Human Evolution: A Synthesis*. Blackwell Publishing.
- Stringer, C. (2007). The origin and dispersal of *Homo sapiens*: Our current state of knowledge, in P. Mellars et al. (eds.), *Rethinking the Human Revolution: New Behavioral and Biological Perspectives on the Origin and Dispersal of Modern Humans* (pp. 15-20). University of Cambridge: McDonald Institute for Archeological Research.
- Van Leynseele, H. (1975). Restrictions on serial verb constructions in Anyi. *Journal of West African Languages*, X, 189–217.

different populations in Africa, and that, as these different populations migrated to different parts of the world, they brought with them these diverse hierarchical grammars. Stringer (2007, 17) mentions that there might have been an African version of multiregionalism.

ACTIVE CONTROL OF COMPLEXITY GROWTH IN NAMING GAMES: HEARER'S CHOICE

WILLIAM SCHUELLER, PIERRE-YVES OUDEYER

Flowers Team, Inria Bordeaux Sud-Ouest

ENSTA ParisTech

william.schueller@inria.fr, pierre-yves.oudeyer@inria.fr

How do linguistic conventions emerge among a population of individuals? A shared lexicon can self-organize at population level through local interactions between individuals, what has been shown in the Naming Games computational framework. However, the dynamics of the convergence towards this shared convention can differ a lot, depending on the interaction scenario. Infants, who acquire social conventions really fast, control actively the complexity of what they learn, following a developmental pathway. Adults also adapt the complexity of their linguistic input when speaking to language beginners. We show here that such active learning mechanism can improve considerably the speed of language formation in Naming Game models. We compare two scenarios for the interactions: either the speaker exerts an active control, or the hearer does. The latter scenario shows faster dynamics, with more robustness.

1. Motivations

How does language emerge, evolve and gets transmitted between individuals? What mechanisms underly the formation and evolution of linguistic conventions, and what are their dynamics? Computational linguistic studies showed that local interactions in groups of individuals (e.g. humans or robots) can lead to self-organization of lexica associating semantic categories to words (Steels, Kaplan, McIntyre, & Van Looveren, 2002). However, it still doesn't scale well to complex meaning spaces and a large number of possible word-meaning associations, implying high competition among lexical conventions.

In statistical machine learning and in developmental sciences, it has been argued that an active control of the complexity of learning situations can have a significant impact on the global dynamics of the learning process (Gottlieb, Oudeyer, Lopes, & Baranes, 2013; Lopes & Montesano, 2014; Kaplan, Oudeyer, & Bergen, 2008). This approach has been studied mostly for single robotic agents learning sensori-motor affordances (Oudeyer, Kaplan, & Hafner, 2007; Moulin-Frier & Oudeyer, 2013), but active learning might represent an evolutionary advantage for language formation at the population level as well (Oudeyer & Smith, 2014; Steels, 2004).

Naming Games are a computational framework, elaborated to simulate the

self-organization of lexical conventions in the form of a multi-agent model (Steels, 2001). Through repeated local interactions between random couples of agents (designated *speaker* and *hearer*), shared conventions emerge. Interactions consist of uttering a word - or an abstract signal - referring to a topic, and evaluating communication success or failure.

However, a lot of processes involved in these interactions are random choices, especially the choice of a communication topic. Some preliminary work on the introduction of active learning algorithms in these models already shows significant improvement of the convergence process towards a shared vocabulary, but only with the speaker actively controlling vocabulary growth (Oudeyer & Delaunay, 2008; Schueller & Oudeyer, 2015; Cornudella, Van Eecke, & Van Trijp, 2015).

Memorization skills of infants are improved through active query of lexical knowledge (Partridge, McGovern, Yung, & Kidd, 2015), and experiments with children learning tasks in a social context suggest that this active behavior may also be part of the mechanisms used naturally in an interacting population of human learners (Vredenburg & Kushnir, 2015). In this work, we adapt the existing algorithms to a variant of the Naming Games where the hearer is actively controlling the complexity growth of the shared lexicon.

2. Methods

2.1. Interactions

The exact interaction process used in this work can be described as follows: Among the population, two agents are randomly picked and designated as speaker and hearer. A topic is chosen within the set of possible meanings (either randomly or actively by one of the agents), and the speaker utters the word associated (in its own vocabulary) to this meaning. The hearer then guesses the meaning of the word, and compares it to the actual topic. If the two meanings match, the interaction is successful. Otherwise, the communication is a failure, and both agents have the opportunity to update their vocabularies. We distinguish here three different scenarios: the topic is chosen randomly (like in the original models), either by the speaker, or the hearer (figure 1).

Vocabularies are represented as binary matrices, rows and columns being respectively meanings and words. The sets of meanings \mathcal{M} and words \mathcal{W} are finite (cardinalities M and W) and symbolic (no grounded meaning or word). All N agents of the considered population start with empty vocabularies (all-0 matrices).

In Wellens (2012), a classification of Naming Games interaction types is proposed. We will employ one of the vocabulary update method described there, the *Imitation Strategy* (figure 2, choice explained in section 4). An agent always adds to its vocabulary the meaning-word association used in the interaction, erasing all potential synonyms or homonyms. This ensures a maximum of one associated word per meaning.

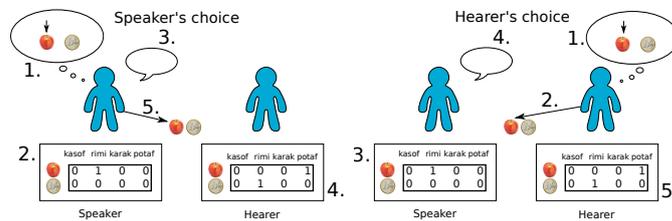


Figure 1. Interaction process for both active scenarios considered in this work. Beforehand, two individuals have been randomly selected among a population, an designated as speaker (S) and hearer (H). Speaker's choice: 1. S chooses a topic, 2. S checks its vocabulary to find or invent an associated word, 3. S utters the word, 4. H guesses the intended meaning, 5. S indicates the intended meaning. Hearer's choice: 1. H chooses a topic, 2. H indicates the intended meaning, 3. S checks its vocabulary to find or invent an associated word, 4. S utters the word, 5. H checks its vocabulary for a meaning associated to the uttered word. In both cases, if all meanings match, the interaction is considered a success, otherwise a failure. After the process, both agents can update their vocabularies.

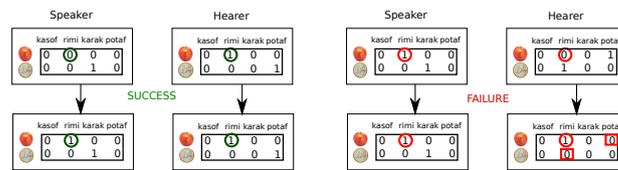


Figure 2. Vocabulary update policy: in the two interactions described here, topic was the object apple, and uttered word was *rimi*. When the speaker is referring to an unlabeled (for him) meaning, it creates an association during the update. When an interaction fails, the hearer adds the used association, and erases any conflicting homonym or synonym.

2.2. Strategies for the active choice of topics

During an interaction (figure 1), one of the involved agents (speaker or hearer) chooses the topic, i.e. the meaning inferred by the speaker. This choice is done with only local information; in other words an agent doesn't access the memory of the others. The way of picking of the topic is called a strategy. A uniform random choice over the set of meanings is called the naive strategy.

The three active strategies used in this work (figure 3) cover a certain diversity of active mechanisms. This work analyzed 3 different strategies, each of which balances two types of behaviors: choosing an unlabeled meaning (to increase the vocabulary size) or choosing an already labeled one (in order to consolidate the existing vocabulary). In the latter case, the agent employs a confidence function to discriminate among known meanings those which would benefit consolidation in priority.

Naive (random)	Success Threshold	Minimal counts	Info. Gain softmax
$m \leftarrow \text{random}(\mathcal{M})$	if $\text{mean}\left(\frac{\text{succ}(i)}{\text{succ}(i)+\text{fail}(i)}\right)_{i \in \mathcal{LM}} \geq \alpha$: $m \leftarrow \text{random}(\mathcal{UM})$ else : $m \leftarrow \text{argmin}_{i \in \mathcal{LM}}\left(\frac{\text{succ}(i)}{\text{succ}(i)+\text{fail}(i)}\right)$	if $\forall i \in \mathcal{LM} \text{ succ}(i) < \mathbf{n}$: $m \leftarrow \text{random}(\mathcal{UM})$ else : $m \leftarrow \text{argmin}_{i \in \mathcal{LM}}(\text{succ}(i))$	if $P_{exp}(\mu, \beta) \geq \text{random}([0, 1])$: $m \leftarrow \text{random}(\mathcal{UM})$ else : $m \leftarrow \text{random}(\mathcal{LM})$
\mathcal{M} : all meanings, \mathcal{LM} : labeled meanings, \mathcal{UM} : unlabeled meanings, μ : vocabulary size (# word-meaning associations) succ : # successful interactions per meaning, fail : # failed interactions per meaning, P_{exp} : equations 1, 2 and tabular 4			

Figure 3. Strategies: Choice of meaning m (by speaker or hearer, depending on the scenario)

Speaker:

$$P_{exp}(\mu, \beta) = \frac{e^{\frac{G(+1)_\beta p(+1)}{\beta}}}{e^{\frac{G(+1)_\beta p(+1)}{\beta}} + e^{\frac{G(-1)_\beta p(-1)}{\beta}}} \quad (1)$$

Hearer:

$$P_{exp}(\mu, \beta) = \frac{e^{-\frac{G(-1)_\beta p(-1)}{\beta}}}{e^{\frac{G(+1)_\beta p(+1)}{\beta}} + e^{-\frac{G(-1)_\beta p(-1)}{\beta}}} \quad (2)$$

Speaker's choice			Hearer's choice		
$\Delta\mu$	Info. Gain G_μ (bits)	Probability p_μ	$\Delta\mu$	Info. Gain G_μ (bits)	Probability p_μ
+1	$\log_2(W - \mu)$	$\frac{(M-\mu)(W-\mu)}{MW}$	+1 (explore)	$\log_2(W - \mu)$	$\frac{W-\mu}{W}$
0	0	$\frac{\mu(1+M+W-2\mu)}{MW}$	0 (explore)	0	$\frac{\mu}{W}$
			0 (check)	0	$\frac{W-\mu+1}{W}$
-1	$-\log_2(W - \mu + 1)$	$\frac{\mu^2 - \mu}{MW}$	-1 (check)	$-\log_2(W - \mu + 1)$	$\frac{\mu-1}{W}$

Figure 4. In both studied scenarios, hearer's possible outcomes with associated gains and probabilities, when having μ associations in the vocabulary \mathcal{V} . There are M meanings and W words. To determine the probabilities, each agent assumes the other agent's vocabulary to be a permutation of its own or in other words, that they share the same μ value. Information measure (defining the gain) is introduced in section 2.3.

2.2.1. Success threshold

The Success Threshold strategy was first introduced in Oudeyer and Delaunay (2008). The confidence function is the percentage of successful interactions, computed for each meaning. If its average value exceeds a given threshold (parameter α), the agent will choose an unlabeled meaning. Otherwise, the meaning with the lowest confidence value is chosen (randomly in the original definition).

2.2.2. Minimal count of successes

This strategy is defined by a confidence function being the sum of successful interactions, per meaning. If it is higher than a given minimum value (parameter n) for all labeled meanings, the agent will choose an unlabeled meaning. Otherwise, the meaning with the lowest value is chosen.

2.2.3. Information gain soft maximization

This strategy is an extension of the Information Gain maximization strategy introduced in Schueller and Oudeyer (2015). The agent chooses between exploring or teaching/checking depending on the expected outcome of the interaction in each case, always for the hearer, quantified as information gain (see 2.3). The choice is done according to a soft-maximization probability distribution (parameter β). Those expected values are computed following a simple assumption: the other agent involved in the interaction has a vocabulary of the same size, but completely independent. In other words, it is supposed to be a random permutation of the binary matrix representing the vocabulary of the decision-making agent. A speaker, if making a decision, will prefer to maximize its outcome for each given hearer. A hearer on the other hand will prefer to check, while minimizing the information loss and avoiding early exploration (see equations 1 & 2).

2.3. Measure

To compare the strategies, we will use the measure introduced in Schueller and Oudeyer (2015). It describes convergence towards a state where all agents have an identical vocabulary, hence sharing a common lexicon. Computed over a population of agents, it takes values in range $[0, 1]$. When the measure equals 0, no agent shares any word-meaning association with any other (maximum distance to converged state), whereas a value of 1 means the population has converged. The exact definition of the measure is the normalized quantity of shared information between 2 agents' vocabularies, averaged over all possible couples in the population. All strategies used, including the naive one, do converge in finite time. The only constraint is $M \leq W$. A proof can be found in Schueller and Oudeyer (2015).

3. Results

3.1. Parameters

To set the parameter of the strategies, we will use the method introduced in Schueller and Oudeyer (2015) and retain values yielding fast convergence dynamics. We ran simulations for different values of the parameter spanning the possible interval, and plotted a snapshot of the status of all simulations after a given number of interactions. We then inferred a best value for the parameter, while having an idea of the robustness of this choice. In all the simulations of this paper, we used $M=W=N=20$.

In all cases we found parameter values yielding convergence in less than 10.000 interactions. According to figure 5, for speaker's choice: $\alpha = 85\%$, $n = 20$ and $\beta = 0.2$; for hearer's choice: $\alpha = 85\%$, $n = 0$ and $\beta = 0.35$. The flatness of the curves indicates that hearer's choice parameters are more robust to a change in value than in the speaker's choice scenario.

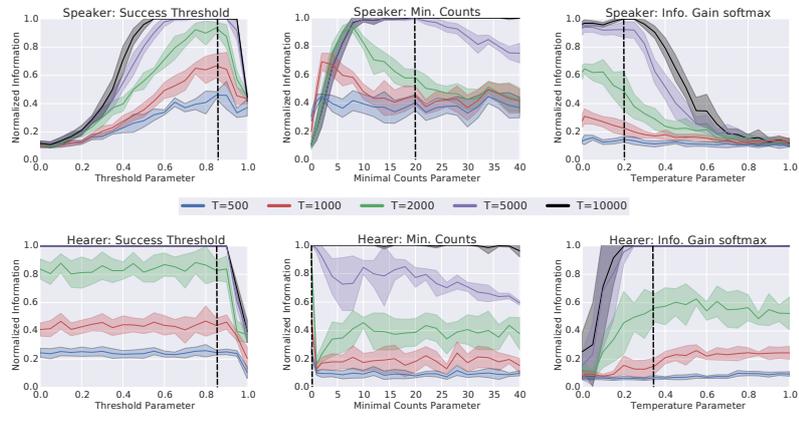


Figure 5. Convergence speed dependance on strategy parameters, for 3 strategies and 2 active interaction scenarios (see section 3.1, strategies described in section 2.2). In all cases hearer’s choice scenario parameters are more robust to change in value. Snapshots are taken for concurrent strategies spanning a relevant parameter interval, at different time steps (500, 1000, 2000, 5000, 10.000 interactions). Vertical lines show parameter values chosen for the comparisons in figure 6. ($M=W=N=20$, averaged over 8 trials)

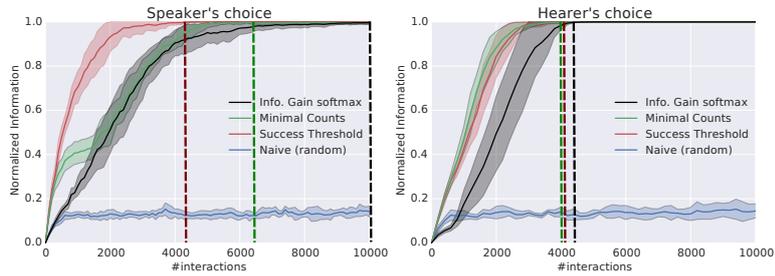


Figure 6. Strategy comparisons in both active scenarios. Naive strategy converges slowly (after 1.000.000 interactions – not depicted here). Hearer’s choice policy is more efficient for all active learning strategies. Last 5% of information are acquired slower when the speaker is choosing. Vertical lines show full convergence time for each strategy. ($M=W=N=20$, averaged over 8 trials)

3.2. Comparison

As we can see in figure 6, active learning strategies were shown to speed up convergence significantly, for all tested strategies. Convergence to a shared lexicon over the whole population was reached between 4000 and 10.000 interactions,

which was a significant improvement compared to the naive strategy (converging after 10^6 interactions).

In the hearer's choice scenario, convergence process is faster for all active learning strategies. However the success threshold strategy shows similar dynamics in both scenarios for the first 95% of normalized information. The last 5%, for all strategies, are slowly acquired when the speaker is choosing, compared to the dynamics of the first 95%. When the hearer is choosing, the dynamics of the last 5% stays comparable to the overall dynamics. Information gain maximization is slower than the other two strategies. Minimal counts strategy shows a shift in convergence speed when speaker is choosing and the set count value is reached. When the hearer is choosing, it performs as well as the success threshold strategy.

4. Discussion

It is important to understand that our research was carried out with the assumption of Imitation Strategy. Other possible vocabulary update policies, including "Minimal" and "Lateral Inhibition" (see Wellens (2012)), are likely to exhibit different convergence rates. The information measure from Schueller and Oudeyer (2015) is not yet defined for those cases, but an extension of the study including those alternative vocabulary update policies is planned.

In this work, we have shown that active information request by the hearer can be a more efficient policy than active information provision by the speaker, in the Naming Games framework. The observed difference between the two policies lies mostly in the acquisition dynamics of the last 5% of information. These findings support that high correlation between vocabularies is best handled by active learning than by active teaching.

Furthermore, active information request is more robust, as a wide range of parameters lead to improved dynamics (compared to random choice), for all studied algorithms. On the other hand, the parameters need to be finely tuned for the speaker's choice policy. This could be understood as a difference in required skill-level between policies. From an evolutionary point of view, it implies that an active information request behavior may be developed faster.

If the dependance of the results on the numbers of meanings, words and agents first needs to be studied, the next logical step would be to mix both policies, by giving agents the opportunity to take turns. Deciding speakers may bring faster dynamics at the beginning, and hearers near the end of the lexicon establishment process. An active choice of the partner, like humans selecting who they may prefer interacting with, could also bring some further improvement.

Source code

The code used for the simulations of this paper was written in Python. It is available as open source software, along with explanatory notebooks, on the Inria Flowers team github: <https://github.com/flowersteam/naminggamesal>

Acknowledgements

We would like to thank all those who provided help and feedback, especially Fabien Benureau, Sébastien Forestier, Florian Golemo and Alexandre Spriet. This research was partially funded by ERC Grant EXPLORERS 240007.

References

- Cornudella, M., Van Eecke, P., & Van Trijp, R. (2015). How intrinsic motivation can speed up language emergence. In *Proceedings of the european conference on artificial life* (pp. 571–578).
- Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, *17*(11), 585–593.
- Kaplan, F., Oudeyer, P.-Y., & Bergen, B. (2008). Computational models in the debate over language learnability. *Infant and Child Development*, *17*(1), 55–80.
- Lopes, M., & Montesano, L. (2014). Active Learning for Autonomous Intelligent Agents: Exploration, Curiosity, and Interaction. *CoRR*, *abs/1403.1*, 40.
- Moulin-Frier, C., & Oudeyer, P.-Y. (2013). Exploration strategies in developmental robotics: a unified probabilistic framework. In *proceedings of the 3rd international conference on development and learning and on epigenetic robotics* (pp. 1–6).
- Oudeyer, P.-Y., & Delaunay, F. (2008). Developmental exploration in the cultural evolution of lexical conventions. In *proceedings of the 8th international conference on epigenetic robotics : modeling cognitive development in robotic systems*.
- Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *Evolutionary Computation, IEEE Transactions on*, *11*(2), 265–286.
- Oudeyer, P.-Y., & Smith, L. (2014). How evolution may work through curiosity-driven developmental process. *Topics Cogn. Sci.*
- Partridge, E., McGovern, M., Yung, A., & Kidd, C. (2015). Young Childrens Self-Directed Information Gathering on Touchscreens. In *Proceedings of the 37th annual meeting of the cognitive science society*.
- Schueller, W., & Oudeyer, P.-Y. (2015). Active Learning Strategies and Active Control of Complexity Growth in Naming Games. In *proceedings of the 5th international conference on development and learning and on epigenetic robotics* (pp. 220–227).
- Steels, L. (2001). Language games for autonomous robots. *Intelligent Systems, IEEE*, *16*(5), 16–22.
- Steels, L. (2004). The Autotelic Principle. *Science*, *313*, 1–16.
- Steels, L., Kaplan, F., McIntyre, A., & Van Looveren, J. (2002). Crucial factors

in the origins of word-meaning. *The transition to language*, 12, 252–271.
Vredenburgh, C., & Kushnir, T. (2015). Young children's help-seeking as active
information gathering. *Cognitive science*.
Wellens, P. (2012). *Adaptive Strategies in the Emergence of Lexical Systems*.

MIND THE GAP: INDUCTIVE BIASES IN PHONOLOGICAL FEATURE LEARNING

KLAAS SEINHORST

*Amsterdam Center for Language and Communication, University of Amsterdam
Amsterdam, The Netherlands
seinhorst@uva.nl*

Although extensive research has been done into the acquisition of non-linguistic feature combinations, empirical evidence about phonological feature learning is scarce. I present results from learning experiments in which participants learnt a data set with the internal structure of a plosive segment inventory. The outcomes suggest that learning biases may indeed play a role in phonological typology, and that learners reduce the cumulative complexity in the data set considerably. These results support the hypothesis that the reduction of complexity is a driving force in the evolution of language.

1. Introduction: feature learning

Recent years have seen an impressive body of research investigating the question how cognitive constraints operate on linguistic typology: considering that language has been passed on between countless generations, any learning biases that humans may possess must have had a profound effect on phenomena that we observe in languages today (Christiansen & Chater 2008, Chater & Christiansen 2010, and many others), and much empirical work has been done to show that some properties of language indeed emerge as a result of the repeated acquisition process.

One striking property of spoken languages is that they seem to prefer sound systems without gaps (De Groot 1931, Martinet 1968, Clements 2003). Surprisingly, very few attempts have been made to explain this observation. Martinet (1968) suggests almost casually that this observation may have something to do with the way we learn a language, but does not expound on this suggestion, and empirical evidence has mostly been lacking. In the last few years, however, some experiments have been done to investigate how humans learn phonological feature combinations. These experiments have been inspired by experiments from cognitive psychology starting in the 1960s, in which participants learnt non-linguistic feature combinations. These stimuli could be

described as a combination of three binary feature values; the complete stimulus set, then, comprised $2^3 = 8$ categories, of which subjects were shown 4. There are six different ways in which four categories can be drawn from the complete data set, nowadays often referred to as the six Shepard types (after the researcher who invented them, cf. Shepard et al. 1961):

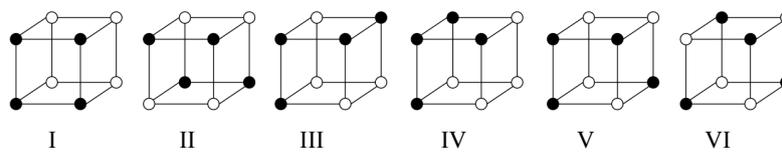


Fig. 1. The six types from Shepard, Hovland & Jenkins (1961).

In memorization and classification tasks, learners of type I did best; learners of type II fared worse, but still better than those trained on types III-V; and type VI proved to be very difficult. Feldman (2000) suggests that the difficulty of a type is correlated with its logical complexity, a measure of the complexity of the internal structure of the type: more compressible inventories have lower complexity indices. Griffiths et al. (2008) carried out a slightly modified version of Shepard et al.'s experiment within the iterated learning paradigm, and it turned out that participants increasingly often selected Type I, the type with the lowest logical complexity. These results suggest that our learning biases cause us to prefer compressible data sets, and indeed Kirby et al. (2015) argue that the reduction of complexity is a major factor in the evolution of language.

2. The acquisition of phonological feature combinations

Both Shepard et al. (1961) and Griffiths et al. (2008) used non-linguistic stimuli in their experiments, but their results may also have implications for our understanding of how humans learn phonological feature combinations. However, so far little research is available on this topic. Moreton et al. (2015) used the Shepard types to investigate the acquisition of different phonological alternation patterns; Pater & Staubs (2013) provided computer simulations suggesting that iterated learning reduces complexity in plosive inventories, and I present empirical evidence for this hypothesis. I have conducted experiments that were inspired by the Shepard types, but applied to phonology, more specifically to sound systems, and in particular to plosive inventories.

All spoken languages that have been described so far make use of plosive segments; the vast majority of languages employ (at least) a three-way place of articulation contrast (labial vs. coronal vs. dorsal), and most of them also

implement an binary voicing contrast (often voiceless vs. voiced). Table 1 lists the six resulting feature combinations that are most common in the world's languages:

Table 1. Common feature combinations in plosive inventories.

	[labial]	[coronal]	[dorsal]
[-voice]	/p/	/t/	/k/
[+voice]	/b/	/d/	/g/

We can capture these $2 \cdot 3 = 6$ categories in a Shepard-type-like representation. Assuming that languages use between three and six feature combinations, the following types can be constructed:

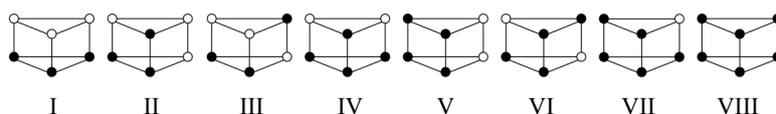


Fig. 2. The eight types that are based on plosive inventories.

In the learning experiment, participants ($n = 96$) were assigned one of the types from Fig. 2. Since the participants were adults and thus had already acquired a phonological system, the experimental stimuli were not segments from spoken language, but from sign language. The data set had the same structure as a plosive inventory: all signs could be described as a combination of a ternary handshape feature and a binary thumb opposition contrast. None of the participants had any knowledge of sign language: this way, it was ensured that participants acquired a feature system *de novo*, similar to the way infants learn a new phonological system (for evidence for feature learning biases in infants, cf. Saffran & Thiessen (2003)).

A male signer was photographed producing the six signs, each eight times, to ensure some phonetic variation between tokens. Participants were exposed to photos of the signs in random order, and each category in the learner's type appeared in the input 24 times (i.e. each photo was shown three times). This means that Type I, II and III learners saw 72 pictures in random order, Type IV, V and VI learners saw 96, Type VII learners saw 120 and Type VIII learners saw 144. Subsequently, participants were asked to indicate with sliders how often they had seen the six possible signs (and two controls). The slider had no ticks, in order to avoid preference for the ticked values; its left end was marked 'not at all', its right end 'very often'. Although participants only saw those two

subjective labels, the left end corresponded to the value 0, the right end to 100; responses were scaled to the highest indicated value. For each type, Table 2 shows the logical complexity and error score (quantified as the average misestimation per category):

Table 2. Logical complexity indices and error scores for the eight types.

	I	II	III	IV	V	VI	VII	VIII
logical complexity	1	3	5	3	1	5	4	1
error score	4	4	9	11	9	24	14	10

A statistically significant effect of type on error score was found ($F(7, 88) = 7.206, p < .001$), as well as a statistically significant correlation between logical complexity and error score ($\rho = .365, p < .001$). These findings are in line with the results from experiments about non-linguistic feature learning.

Knowing what signs participants indicated having seen, we can also interpret their responses in terms of the eight types, i.e. as categorizations: Table 3 shows the probabilities of type A (rows) being categorized as type B (columns).

Table 3. Categorization responses per type.

input	output							
	I	II	III	IV	V	VI	VII	VIII
I	1.0							
II		.92			.08			
III			.92	.08				
IV				.75			.17	.08
V					1.0			
VI						.5	.17	.33
VII							.75	.25
VIII								1.0

The diagonal that runs from the top left to the bottom right contains the “correct” responses. The table reveals interesting patterns: firstly, learners generally reproduce inventories with low complexity faithfully (as was also clear from the low error scores for types I, V and VIII in Table 2); secondly, if learners make an error, they never omit a category that was present in the input, but always fill gaps; thirdly, most errors in the types with more than three categories favour type VIII, one of the types without gaps. Such regularization was reported by a.o. Hudson Kam & Newport (2005) and Ferdinand (2015) as well, and it is likely due to inductive biases that aim to reduce complexity. In

fact, this cohort of learners has reduced the cumulative complexity in the data set by 11.2%.

We can consider the proportions from Table 3 to be transition probabilities, which would make Table 3 a Markov matrix. If we assume that the outputs of these learners serve as inputs to a consecutive group of learners, who have the same inductive biases and whose outputs are fed to a next group, etc., we can set up a Markov chain. Figure 4 shows how the predicted frequencies of the eight types evolve, and reveals that a stable final state emerges in approximately 50 generations.

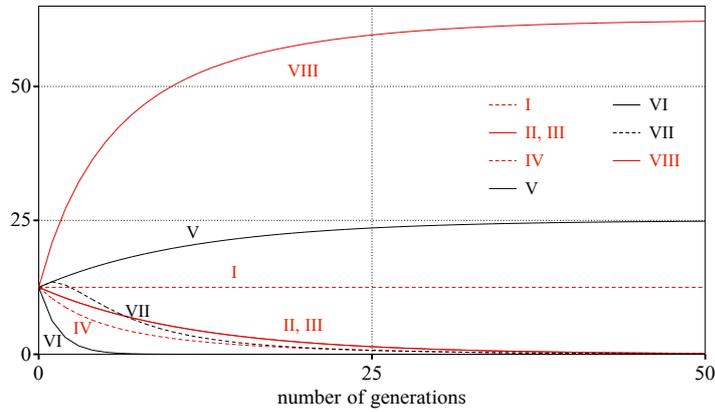


Fig. 4. The development of the relative frequencies of the eight types over 50 generations.

Table 4 lists the predicted frequencies after 50 generations as well as the attested frequencies in P-base, a database of segment inventories (Mielke 2008):

Table 4. Predicted relative frequencies for all types after 50 generations, and their attested relative frequencies in P-base.

frequency (%)	Type							
	I	II	III	IV	V	VI	VII	VIII
predicted	12.5	0.0	0.0	0.0	25.0	0.0	0.0	62.5
attested	20.1	0.3	1.0	2.7	0.5	0.6	10.8	63.9

In this stable state, only the types with the lowest complexity (I, V, VIII) remain. The cumulative logical complexity has been reduced considerably: it is 65.2% lower than in the initial state.

The correlation between the predicted frequencies in the Markov chain and the attested frequencies in P-base, both cautiously treated as ranked variables because of the low numbers of observations for some types, is not statistically significant (Pearson's $r = .625$, $p = .098$). This can largely be ascribed to phonetic factors that play a role in the typology of plosive inventories. For instance, type I languages in spoken language may be more frequent than types II and III because speakers avoid learning an intricate phenomenon like vocal fold vibration until all places of articulation have been used; types II and V may be rare because they only use part of the oral cavity and would force speakers to avoid other regions, which would be articulatorily effortful; /p/ often lenites to /f/, creating a type VII system; /g/ frequently undergoes spirantization, also leaving a type VII inventory. What still stands, however, is the overwhelming majority of type VIII systems, both in the predicted and attested frequencies, and the low frequencies of types II, III, IV and VI.

3. Conclusion

In learning experiments with a sign language that resembles a plosive inventory, logical complexity turned out to be a good predictor of learning success, and a single cohort of participants reduced logical complexity by 11.2%; if we use their categorization proportions iteratively, the reduction increases to 65.2%. The stable state of these iterations correlate fairly well with attested frequencies in spoken language, if we also take perceptual and articulatory factors into account. These results provide empirical support for the hypothesis that the reduction of complexity is a driving factor in language evolution.

References

- Chater, Nick & Morten Christiansen (2010). Language acquisition meets language evolution. *Cognitive Science*, 34, 1131-1157.
- Christiansen, Morten & Nick Chater (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31, 489-558.
- Clements, Nick (2003). Feature economy in sound systems. *Phonology*, 20, 287-333.
- De Groot, Willem (1931). Phonologie und Phonetik als Funktionswissenschaften. *Travaux du Cercle Linguistique de Prague*, 4, 146-147.
- Feldman, Jacob (2000). Minimization of Boolean complexity in human concept learning. *Nature*, 407, 630-633.
- Ferdinand, Vanessa (2015). *Inductive evolution: cognition, culture and regularity in language*. PhD thesis, University of Edinburgh.
- Griffiths, Thomas, Brian Christian & Michael Kalish (2008). Using category structures to test iterated learning as a method for revealing inductive biases. *Cognitive Science*, 32, 68-107.
- Hudson Kam, Carla & Elissa Newport (2005). Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Language Learning and Development*, 1 (2), 151-195.
- Kirby, Simon, Mónica Tamariz, Hannah Cornish & Kenny Smith (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87-102.
- Martinet, André (1968). *La linguistique synchronique: études et recherches*. Paris: Presses Universitaires de France.
- Mielke, Jeff (2008). *The emergence of distinctive features*. Oxford: Oxford University Press. P-base: <http://aix1.uottawa.ca/~jmielke/pbase/index.html>
- Moreton, Elliot, Joe Pater & Katya Pertsova (2015). Phonological concept learning. *Cognitive Science*, 1-66.
- Pater, Joe & Robert Staubs (2013). *Feature economy and iterated grammar learning*. Presentation, 21st Manchester Phonology Meeting.
- Saffran, Jenny & Erik Thiessen (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39 (3), 484-494.
- Shepard, Roger, Carl Hovland & Herbert Jenkins (1961). Learning and memorization of classifications. *Psychological Monographs*, 75 (13).

AGAINST THE EMERGENT VIEW OF LANGUAGE EVOLUTION

MAGGIE TALLERMAN

*Linguistics Section, Newcastle University,
Newcastle upon Tyne, UK
maggie.tallerman@ncl.ac.uk*

Under the emergent view of language evolution (Nóbrega & Miyagawa 2015), it is claimed that the language faculty appeared quite recently in human evolution and had no earlier pre-syntactic stages. Language is said to be an amalgamation of two pre-existing systems that also occur in animal communication: an 'E' system, for 'expressive', which is likened to systems of learned birdsong, and an 'L' system, for 'lexical', which is likened to monkey alarm calls. What occurred solely in the case of humans was the advent of a Merge operation, which integrated the two systems. Here I argue against this 'integration hypothesis'. None of the proposed analogues in animal communication have the critical properties occurring in human language. The syntax of birdsong is unlike the syntax of language in all relevant respects. There are no analogues to the functional elements of the E-system in animal communication. However, under a gradualist rather than an emergent view of language evolution, we can see how functional elements and linguistic morphology evolved.

1. Introduction: the Integration Hypothesis

In a recent paper, Nóbrega & Miyagawa (2015) propose what they term an 'emergent' view of language evolution, which 'assumes that the language faculty emerged late in historical development without any prior pre-syntactic stage'. Under this view, language is the result of Merge operating on two pre-existing systems that also occur elsewhere in animal communication, but are only integrated in humans. Nóbrega & Miyagawa specifically argue against the 'gradualist' view of language evolution, which proposes an asyntactic protolanguage initially consisting of wordlike entities without combinatorial properties, as outlined in the work of Bickerton (1990, 1995 et seq.), Hurford (2012) and Jackendoff (2002), among many others. According to Nóbrega & Miyagawa, and contrary to the cited work on protolanguage, words do not precede syntax, but rather, are derived by it. Here, I outline a number of problems for the Integration Hypothesis (Miyagawa et al. 2013, 2014; Nóbrega & Miyagawa 2015), and argue that in fact there are no evident analogues in animal communication for the two main components of language that the authors propose.

The Integration Hypothesis states that language comprises two main components: sentences are constructed of an inner L-layer, for 'lexical', providing basic lexical meaning, and an outer E-layer, for 'expressive', which

has properties such as Tense, or Question, or Focus (Miyagawa et al. 2013, 2014); the latter are functional elements that lack independent status. Nóbrega & Miyagawa (2015) go further, suggesting that not only phrases and sentences but also simple words consist of these two layers, so that there is an inner L-structure at the core of all expressions, and an outer E-layer, the ‘expression structure’, comprising ‘function elements that give shape to the expression’ (Nóbrega & Miyagawa 2015): these are categorial and inflectional features. For words, the lexical layer is said to be a ROOT, which lacks syntactic category, so is not yet a noun, verb or adjective etc.; it also lacks inflectional information, so has no number, case or tense etc. A root cannot be or become a word until it combines with categorial and grammatical information, provided by the E-layer. Nóbrega & Miyagawa propose that, crucially, this stage in language evolution does not occur until the late appearance of a Merge operation in the grammar (thus, presumably, in human cognition), since Merge integrates the two pre-existing systems, L and E. At word-level, the proposal is that Merge then produces units with syntactic categories and other grammatical features: words as we know them.

Leaving aside the controversial question of whether or not the Merge stage was a recent development, occurring abruptly within the past 100,000 years (though see Dediu & Levinson 2013, Johansson 2013, Tallerman 2014 for alternative arguments), one major problem for the Integration Hypothesis is that it cannot satisfactorily answer the question of where the categorial and grammatical features themselves come from. Miyagawa and colleagues propose that each of the two layers has an antecedent or an analogue elsewhere in animal communication; it is only Merge that is novel. The E-type is claimed to be similar to birdsong, which is said to have ‘syntax without meaning’ (Nóbrega & Miyagawa 2015), while the L-type, the lexical component, is likened to monkey calls, considered to be isolated units with real-world reference. An immediate problem here is that we do not and cannot know what monkey alarm calls refer to, if indeed they *refer* directly to anything in the real world at all, rather than being purely affective, so relating to the caller’s internal state on perceiving a specific predator.

2. Does the E-system have analogues in animal communication?

2.1 *The E-system and birdsong*

Here, I chiefly consider problems with the proposed antecedents for – and the appearance of – the E-layer in language. A variety of parallels between birdsong and language have been suggested in the literature, including the learning of song by juvenile birds of certain species, the appearance of a critical period in both language and learned song, and so on (e.g Bolhuis et al. 2010). Miyagawa

and colleagues, on the other hand, propose that ‘the actual link is between birdsong and the expression structure portion of human language’ (Nóbrega & Miyagawa 2015). This layer is claimed to be directly analogous to birdsong (actually the authors go further, claiming a genetic link between birdsong and language). How robust is this analogy? In birdsong, the E-layer is said to have ‘expressive’ functions such as marking territory or mating availability (Miyagawa et al. 2013, 2014), obviously rather starkly different functions to the linguistic ability to mark a question or focalize an element, also handled by the E-component. More problematically, however, in language these functions are typically signalled by permutations in the syntax, so that for instance focalized constituents move to a special position in a clause – fronted, as in English (*This colour, I really love* __), or an immediately pre-verbal position, as in Basque and Hungarian. Questioned phrases may remain in situ, where they are marked with a *wh*-particle, as in Japanese; but cross-linguistically, *wh*-phrases are often signalled by fronting too. In fact, much of the whole *raison-d’être* of syntax lies in moving elements around to signal different expressive functions, including topicalizing, focalizing, questioning, foregrounding and backgrounding various constituents via passivization (*The window got smashed* __ (*by the football*)), and so on, all to achieve distinct pragmatic effects. Syntax not only varies word order, but crucially also changes the valency of predicates, altering grammatical relations (for instance, the object of an active sentence is the subject of the corresponding passive), all in order to affect the meaning and pragmatic function of sentences.

In contrast, birdsong specialists note that song variation may be extensive, but specifically does *not* signal different meanings: ‘in general the different songs that a bird has convey exactly the same message as each other but do it in a highly varied manner’ (Slater 2012: 97). Strikingly, furthermore, even the more complex bird songs are characterized by highly stereotyped ordering sequences, for instance in the song of the chaffinch. Different syllables (in birdsong, a syllable denotes a precisely repeated sequence of one or more notes; Thompson et al. 1994) may be iterated a different number of times, but their sequence remains fixed. Nightingales have a repertoire of around 100–200 distinct song types (Kipper et al. 2006), but again the sequence of these is fixed. In the zebra finch, the song is also stereotyped, though notes may be skipped or added, or the same note repeated, giving an appearance of variation. Interestingly, it seems that the variations in note order may not even be deliberate, but rather are something akin to ‘performance’ errors in these finches (Hurford 2012: 47f). Even where there are permutations in birdsong syntax, no differentiation occurs in the message: in fact, the song syntax itself seems anything but ‘expressive’, if the variations carry no distinct meanings. It appears, then, that the proposed analogy between an E-layer in language and the syntax of birdsong is at best tenuous.

2.2 *Are there analogues to functional elements in animal communication?*

Turning now to the role of the E-layer below sentence level, recall that this is claimed by Nóbrega & Miyagawa (2015) to turn L-layer roots into words, via Merge, in the course of language evolution, by adding categorial and inflectional information. Certainly, the relevant literature does not suggest that birdsong displays anything remotely similar to the functional elements and grammatical features of language. Where, then, do these elements come from in language, if there are no obvious antecedents in animal communication? We might expect that an investigation of primate communication could shed light on the origins of functional elements, but there is little evidence for this. Nóbrega & Miyagawa speculate that it is the L-layer, not the E-layer, that derives from primate communication: monkey alarm calls are the ‘closest approximation’ to linguistic roots in their view. However, alarm calls seem to have more in common with propositions than roots (Bickerton 2009), and might equally indicate the specific escape strategy required rather than essentially naming the predator. Problematically, also, great apes – despite being phylogenetically much closer to humans – lack such specific alarm calls.

In captive great apes, as we see from ape ‘language’ research, functional elements are generally conspicuous by their absence. Kanzi the bonobo, for instance, is well able to carry out many surprising instructions given in English (e.g. *Can you pour the ice water in the potty? Go get the balloon that’s in the microwave*; Savage-Rumbaugh et al. 1998), to a level of 72% accuracy, but is unable to understand coordination (*Bring me a carrot and a banana*). This may indicate Kanzi’s inability to understand hierarchical structure, or a failure to comprehend a critical functional element such as *and* (the authors also report a problem with modal auxiliary *can*). In the examples above, none of the function words are remotely vital to understanding the meanings (Kirby 2000); even in the case of the relative clause, lexical items plus word order suffice: there is no reason to suppose that Kanzi understands subordination. Kanzi is also unable to understand the indefinite article *a*, so that when asked to give *a carrot / a tomato*, he makes an inappropriate response. It seems unlikely, then, that anything in the E-layer, specifically the ubiquitous functional elements of language, stems from our shared phylogeny with other apes.

There are indeed a few suggestions in the literature that something approaching a functional element – a potential antecedent for the E-layer – occurs in certain monkey calls (e.g. Zuberbühler 2002, Ouattara et al. 2009). In Campbell’s monkeys, a ‘boom’ vocalization preceding (and apparently attenuating) other alarm calls has been interpreted as a ‘syntactic rule’ (Zuberbühler 2002), with the ‘boom’ acting as a kind of modifier. Notably, though, the alarm calls occur about 25 seconds later than the boom, indicating that these are not in fact call combinations after all. Ouattara et al. (2009) also

suggest that Campbell's monkeys use a stem + suffix arrangement to broaden the meaning of several distinct calls, so that for instance the eagle alarm is *hok* while *hok-oo* (i.e. purportedly the *hok* call with an *-oo* suffix) is used for 'general arboreal disturbance'. However, while *-oo* may be described as a suffix if the term is used very loosely, it has little resemblance to affixes and other types of morphology in languages.

3. Functional elements have a purely linguistic development

A wide linguistic literature offers clear views of the sorts of attested pathways that languages follow in the development of independent functional elements and grammatical formatives (e.g. Heine & Kuteva 2007; Carstairs-McCarthy 2010). In essence, inflectional and derivational morphology has a life history, and that history depends exactly on words and morphemes being uttered in closely-adjointing sequences, something that does not occur in primate calls. Functional elements of all kinds in language do not occur spontaneously, but develop from pre-existing lexical items, via various routes of grammaticalization that recur with remarkable frequency cross-linguistically (Heine & Kuteva 2002, 2007). For instance, a 'go (to)' verb often develops a future meaning, or even becomes a future marker, devoid of the semantics of movement, as in English *I'm gonna sit still*; the same process is observed in French, Welsh, Sotho, Zulu etc. Similarly, the Baka adverb *ngili* 'yesterday', is grammaticalized as a verbal tense suffix *-ngi*, marking the near past. Many, though not all, affixes originate as separate words, subsequently morphologized; the history of verbal inflection in the Romance languages is another standard example. New morphological variants (allomorphs) frequently appear as a result of very standard phonological assimilation processes occurring at morpheme and word boundaries yielding distinct forms of the same morpheme; German umlaut illustrates, e.g. *Fuss* [fʊs] 'foot' vs. *Füss-e* (foot-PLURAL) [fʏs-ə] 'feet', where the vowel in the noun stem underwent fronting in the plural, under the influence of what was historically a front vowel in the plural suffix. A gradualist view of language evolution is required in order to understand the kinds of development outlined here, since they are by nature gradually-developing phenomena.

In sum, then, we have a very good idea of where grammatical information and functional elements come from, but nothing in the communication systems of other animals seems to provide a promising candidate for antecedents of the E-layer in language. The L-layer seems equally to be a totally distinct system to primate calls. I therefore suggest that we cannot accept the proposal that Merge simply integrates two pre-existing systems, contrary to the Integration Hypothesis. If the E-system and the L-system were there already before Merge existed, it is because both of them evolved – uniquely – in human cognition.

4. A pre-language stage in evolution

Finally, Nóbrega & Miyagawa's concept of language evolution also raises other questions. If the L-layer existed before Merge, then roots existed. Although the authors deny the existence of a 'structureless, protolinguistic stage', they actually do envisage an earlier stage of language containing pre-words: 'The pre-language stage is composed of root-like elements, each occurring in isolation of the others' (Nóbrega & Miyagawa 2015). These, they say, would have reference but lack syntactic category, and thus would be unable to 'participate in any combinatorial systems'. In fact, such a stage seems to be extremely similar to what is proposed in much of the literature on protolanguage, e.g. Bickerton (1990, 1995), Hurford (2012), Jackendoff (2002), Jackendoff & Wittenberg (2014), Tallerman (2007), and many others. In the Nóbrega & Miyagawa system, such pre-words do not join together to form hierarchical structures, since without syntactic categories a pre-word must lack the properties that words have, such as headedness and taking obligatory complements. There could be, then, no syntax. However, nothing in this system seems to prevent roots being strung together in very short sequences, quite arbitrarily, where their meanings seemed to be related, thus forming short, asyntactic sequences of adjoining roots, exactly as outlined in the cited work of Ray Jackendoff. If some of these pre-words are frequently uttered in close proximity – simply abutting each other, without forming hierarchical structures – they are likely to become loosely associated, as Jackendoff suggests. Such a pre-Merge system can in time lead to formalized relationships between pre-words, so that they start, on a case-by-case basis, to take on the properties of argument structure. For instance, a verb-like root could start to co-occur with a noun-like root as its complement, first very informally, then obligatorily. In this way, the verb-like element becomes a head. As supporting evidence, we already know that two adjoining flat structures can subsequently acquire a hierarchical subordination relationship from studies of the many attested historical changes from paratactic to hypotactic syntax (Heine & Kuteva 2007: ch. 5). In such cases, a formerly independent clause becomes a complement or adjunct to a main clause, via the usual processes of grammaticalization. If these speculations are on the right lines, then 'root-like elements' can become syntactic heads over time, and – rather surprisingly – Nóbrega & Miyagawa (2015) have in fact converged on the gradualist concept of a protolanguage.

References

- Bickerton, D. (1990). *Language and species*. Chicago: University of Chicago Press.
- Bickerton, D. (1995). *Language and human behavior*. Seattle: University of Washington Press.
- Bickerton, D. (2009). *Adam's tongue: how humans made language, how language made humans*. New York: Hill & Wang.
- Bolhuis, J. J., Okanoya, K. & Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, *11*, 747–759.
- Carstairs-McCarthy, A. (2010). *The evolution of morphology*. Oxford: Oxford University Press.
- Dediu, D. & Levinson, S. C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology*, *4*, 1–17.
- Heine, B. & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Heine, B. & Kuteva, T. (2007). *The genesis of grammar: a reconstruction*. Oxford: Oxford University Press.
- Hurford, J. R. (2012). *The origins of grammar: language in the light of evolution*. Oxford: Oxford University Press.
- Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar and evolution*. Oxford: Oxford University Press.
- Jackendoff, R. & Wittenberg, E. (2014). What you can say without syntax: A hierarchy of grammatical complexity. In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring grammatical complexity*, (pp. 65–82). Oxford: Oxford University Press.
- Johansson, S. (2013). The talking Neandertals: what do fossils, genetics, and archaeology say? *Biolinguistics*, *7*, 35–74.
- Kipper, S., Mundry, R., Sommer, C., Hultsch, H. & Todt, D. (2006). Song repertoire size is correlated with body measures and arrival date in common nightingales, *Luscinia megarhynchos*. *Animal Behaviour*, *71*, 211–217.
- Miyagawa, S., Berwick, R. C. & Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology*, *4*, 1–6.
- Miyagawa, S., Ojima, S., Berwick, R. C. & Okanoya, K. (2014). The integration hypothesis of human language evolution and the nature of contemporary languages. *Frontiers in Psychology*, *5*, 1–6.

- Nóbrega, V. A. & Miyagawa, S. 2015. The precedence of syntax in the rapid emergence of human language in evolution as defined by the integration hypothesis. *Frontiers in Psychology*, 6, 1–8.
- Kirby, S. (2000). Review of *Apes, language and the human mind* by S. Savage-Rumbaugh, S. G. Shanker & T. J. Taylor. *Journal of Linguistics*, 36, 190–195.
- Ouattara, K., Lemasson, A. & Zuberbühler, K. (2009). Campbell's monkeys use affixation to alter call meaning. *PLOS One*. DOI: 10.1371/journal.pone.0007808
- Savage-Rumbaugh, S., Shanker, S. G. & Taylor, T. J. (1998). *Apes, language and the human mind*. Oxford: Oxford University Press.
- Slater, P. (2012). Bird song and language. In M. Tallerman & K. R. Gibson (Eds.), *The Oxford handbook of language evolution* (pp. 96–101). Oxford: Oxford University Press.
- Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, 117, 579–604.
- Tallerman, M. (2014). No syntax saltation in language evolution. *Language Sciences*, 46, 207–219.
- Thompson, N., LeDoux, K. & Moody, K. (1994). A system for describing bird song units. *Bioacoustics: The Journal of Animal Sound and its Recording*, 5, 267–279.
- Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, 63, 293–299.

LEARNING TO LEARN FROM SIMILAR OTHERS: APPROXIMATE BAYESIAN COMPUTATION THROUGH BABBLING

BILL THOMPSON, HEIKKI RASILO

Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, Belgium

bill@ai.vub.ac.be

heikki.rasillo@aalto.fi

To emerge and persist over cultural transmission, complex linguistic structures must be learnable from noisy, incomplete linguistic data. Rational statistical inference provides a compelling solution to inductive problems at many levels of linguistic structure. However, from an evolutionary perspective, a core question concerns how language learners could be equipped to approximate rational inferences under limited cognitive resources. We present a computational account of how *self-simulation* of linguistic data can aid otherwise challenging probabilistic inferences during language acquisition – a shortcut made possible because humans acquire language from linguistic data produced by other humans. Through an analogy with a class of computational techniques known as Approximate Bayesian Computation, we show how the capacity to *produce* language data can leverage computationally-cheap inductive leaps that approximate rational inference when *learning* language. We derive an approximate inference model for an idealised problem in the acquisition of speech sounds through babbling – a problem in which the gestural source of those sounds is invisible to the learner but must be reverse-engineered. We simulate the dynamics of cultural transmission under this model, and discuss implications for the evolution of speech and language.

1. Introduction

Language acquisition involves inductive inference under uncertainty: at many levels of language structure, the learner must reverse-engineer a richly structured linguistic model from noisy, incomplete linguistic data. Human learners – but no other learning system – reliably solve this inductive problem during early life. How this is possible remains a core question underpinning inquiry into the evolution of language. Learnability arguments underpin many theories concerning the origins of linguistic structure, from those that appeal to specialised innate *biological* evolutionary innovations (Pinker & Bloom, 1990), to those that cite the structure-forming properties of *cultural* evolution (Kirby, Griffiths, & Smith, 2014); in order to understand how linguistic structures emerge and persist among populations, we must understand the computational principles that underpin acquisition of language from culturally transmitted linguistic data.

A key recent development, with considerable implications for the evolution of language, is the discovery that domain-independent principles of rational statisti-

cal inference appear to provide solutions to many formidable inductive problems in the acquisition of structured linguistic knowledge (e.g. Perfors, Tenenbaum, & Regier, 2011). However, when guided by an evolutionary perspective, a core question arises: how might learners be equipped to implement or approximate the often demanding computations that underpin rational statistical inference? In this paper we argue that, from an inferential perspective, the language learner has a powerful trick up her sleeve: the ability to *self simulate* linguistic data. We present a computational account for how self-simulation of linguistic data can leverage approximate inductive inferences that would otherwise be intractable. The computational solution arises because language is a cultural behaviour, learned from data produced by similar others (Chater & Christiansen, 2010). We draw an analogy with a class of inferential techniques known as Approximate Bayesian Computation (ABC), which solve an analogous problem through analogous means. We focus on a concrete example – acquisition of speech sounds through babbling – and discuss implications for the evolution of speech and language.

2. Language via Rational Inference

Several aspects of linguistic structure traditionally thought to imply language-specific acquisition procedures have been shown to be learnable via rational inference. For example, Perfors et al. (2011) show that, given the statistical properties of linguistic data, an ideal rational learner exposed to these data could make the inductive leap to hierarchical phrase-structure in syntax. Likewise, statistical inference can recover the structure of phonetic categories (Vallabha, McClelland, Pons, Werker, & Amano, 2007), word segmentation boundaries (Goldwater, Griffiths, & Johnson, 2009), and structure in verb classes (Perfors, Tenenbaum, & Wonnacott, 2010), to cite just a few examples.

2.1. An Ideal Rational Learner

Rational inference follows an appealingly simple recipe: when evaluating a hypothesis h as an explanation for observed data d , the learner should update her *prior* belief in that hypothesis (independent of the observed data) $p(h)$ by accounting for the *likelihood* of those data under the hypothesis, $p(d|h)$, in accordance with Bayes' rule:

$$p(h|d) = p(d|h)p(h)/p(d) , \quad (1)$$

where $p(d)$ reflects the likelihood of the data marginalised over all possible hypotheses. This scheme for combining subjective and empirical information aligns with human behaviour in many domains. However, in many real-world learning problems, the computational cost of precisely computing the informational ingredients for this formula – $p(h)$, $p(d|h)$, and $p(d)$, – can be prohibitively demanding, for computer scientists and for human learners. As such, there is considerable interest in understanding how learners might approximate the computations that underpin rational inference under limited cognitive resources.

2.2. Approximating Rational Inferences

A growing body of research aims to describe a class of *rational process models* (see e.g. Griffiths, Lieder, & Goodman, 2015): these models demonstrate how simple algorithms can approximate rational inferences in inductive problems such as causal learning (Bonawitz, Denison, Gopnik, & Griffiths, 2014) and category formation (Sanborn, Griffiths, & Navarro, 2010). Likewise, there have been efforts to demonstrate how classic psychological models, such as exemplar models (Shi, Griffiths, Feldman, & Sanborn, 2010) and neural networks (Abbot, Hamrick, & Griffiths, 2013) can approximate the computations implied by Bayesian models. Several of these discoveries have drawn on analogies between human learning and inference algorithms developed in computer science and statistics. Our proposal for language follows in this vein.

3. Learning by Producing: Approximate Inference via Simulation

A core requirement for rational inference is *likelihoods* knowledge: when reverse engineering the underlying source of our observations, we must evaluate the *likelihood* of those observations under any hypothesis we entertain about their cause. However, in many realistic cases, our generative model for the data is so complex or high-dimensional that we cannot say for certain exactly how likely are our observations under the model: in formal terms, the likelihood distribution – $p(d|h)$ in equation 1 above – over possible observations cannot be efficiently computed. This is a common problem that has hindered model-fitting in a number of disciplines, such as population genetics, systems biology, and economics, in which hypothesised models (e.g. population demographic history) and their resulting data (e.g. gene-sequence data) are high-dimensional, and the data likelihoods intractable. We suggest that – given the complex structures that underpin language, and the open-endedness of linguistic data – the language learner would often find herself in a similar position. While ideal learner models generally have access to these computations, how is the language learner to fare if she does not? The solution we explore here rests on the fact that, in the case of language, the *learner* can also be a *producer*, able to test hypotheses by self-simulation of linguistic data.

3.1. Computation in the Face of Unknown Likelihoods

Advances in computer science and statistics have uncovered methods for performing inference in the case of unknown likelihood quantities. In particular, considerable progress has been made for a special class of cases: those in which, though we may *not* be able to compute the likelihood of our observations exactly, we *are* nevertheless in a position to *simulate data from the model*. Various simple inferential procedures have been developed to perform approximate inference on the basis of data simulation, generally known as Approximate Bayesian Computation (see e.g. Beaumont, 2010, for a review of these methods). We propose that as-

pects of language acquisition fall into this class of problem. Language is culturally transmitted: it is acquired through inference over data generated by other humans. Since teacher and learner share anatomical and cognitive properties, the learner is in a position to simulate linguistic data under a mechanism that approximates the generating mechanism, and could leverage this capacity during inference. ABC methods provide a computational framework to demonstrate *how*, and a plausible model for several inductive problems in language.

3.2. A Simple Rejection-ABC Algorithm

Here we describe a simple ABC algorithm based on rejection sampling: in order to demonstrate the principles behind our proposal, we aim for maximal simplicity throughout the examples in this paper, and discuss more sophisticated extensions in section 5. Given an observed dataset $\mathbf{y} = \{y_1, \dots, y_n\}$ generated by a mechanism with an unknown parameter θ , the inductive problem is to characterise the posterior distribution $p(\theta|\mathbf{y}) = p(\mathbf{y}|\theta)p(\theta)/p(\mathbf{y})$: here $p(\theta)$ is the prior distribution over model parameters θ , $p(\mathbf{y}|\theta)$ is the likelihood function for the data conditional on the parameter θ , and $p(\mathbf{y})$ is a normalising constant. Assuming it is impractical to evaluate the likelihood function $p(\mathbf{y}|\theta)$ directly, but simple to *generate* data \mathbf{x} from the distribution $p(\mathbf{x}|\theta)$, then it is possible to collect samples from the posterior distribution by repeating the following procedure:

1. Generate $\theta_i \sim p(\theta)$
2. Generate $\mathbf{x}_i \sim p(\mathbf{x}_i|\theta_i)$
3. Store θ_i if $\rho(S(\mathbf{x}_i), S(\mathbf{y})) < \epsilon$.

Here $S(\cdot)$ represents summary statistics of the dataset, $\rho(\cdot, \cdot)$ is a difference measure between the simulated and observed datasets, and ϵ is an acceptance tolerance level. The stored values of θ then represents a set of samples that approximates the posterior distribution $p(\theta|\mathbf{y})$. This process is akin to one of trial-and-error, and can be summarised intuitively as follows: keep guessing parameter values, producing data, and storing any guesses that led to data which is close to the original observation. In the next section, we introduce our illustrative inductive problem – speech sound acquisition through babbling – and show how this algorithm can be understood as a model of learning that approximates rational inference in that context.

4. Speech Sound Acquisition through Babbling

One of the first inductive leaps the language learner must make is to learn a structured system of speech sounds. For example, on the basis of the speech sounds it encounters, the child must reverse engineer arrangements of the articulatory apparatus required to produce the system of sounds used in its language.

In traditional terminology, the learner must acquire the *backward mapping* from sounds to articulators. To do this the learner must know something of the *forward mapping* from articulators to sounds - corresponding the likelihood distribution in the Bayesian framework. Babbling is typically thought to allow the child to learn the relationship between articulatory gestures and their acoustic counterparts (Vihman, 1991). Several computational models have been proposed for learning the mappings in a babbling phase, such as the neural network models of Guenther (2006) and Kröger, Kannampuzha, and Neuschaefer-Rube (2009), and these learned mappings are later used to invert observed native language phonemes back to articulation, enabling imitation. Our framework offers a bridge between these implementation-level models and computational-level rational analyses.

4.1. Noise in Production of a Target Speech Sound

Here we describe an idealised model of speech perception which qualitatively captures the nature of the inductive problem: inference of an underlying unknown quantity from a dataset of noisy exemplars. Our strategy is to study a simple, abstract model of inference in which we, the modelers, know the likelihood distribution, but assume the learner does not and must perform inference by data simulation as described above. Assume the speaker aims to produce a target speech sound τ : τ reflects a single phonetic feature on a continuous range, such as voice-onset time or an absolute formant value. Realisations of this speech sound are noisy: when the speaker aims for τ , it produces a sound that is normally distributed with mean τ and variance σ^2 . The learner must reverse engineer the underlying intended target τ from a set of n exemplars $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$.

4.2. Approximate Bayesian Computation through Babbling

Given \mathbf{y} , an ideal rational learner estimating τ would compute the posterior distribution $p(\tau|\mathbf{y}) \propto p(\mathbf{y}|\tau)p(\tau)$. Assume a Gaussian prior - $p(\tau) \propto \mathcal{N}(\tau_0, \sigma_0^2)$ - over the range of *possible* values for τ . The prior may represent any constraints (e.g. anatomical or cognitive constraints) that influence speech production. Likewise allow a likelihood function defined by normally distributed noise around the target - $p(\mathbf{y}|\tau) \propto \mathcal{N}(\bar{y}; \tau, \sigma^2/n)$. Our hypothetical language learner does not know the likelihood function, but can sample from its own data-producing mechanism. Following the procedure outlined in section 3.2, the babbling learner approximates rational inferences as follows:

1. Sample hypothesised target from prior $\hat{\tau}_k \sim \mathcal{N}(\tau_0, \sigma_0^2)$
2. Babble m speech sounds $\mathbf{x} = \{x_1, \dots, x_m\} \sim \mathcal{N}(\mathbf{x}; \hat{\tau}_k, \sigma_{seif}^2)$
3. Store hypothesised $\hat{\tau}_k$ if $|\bar{x} - \bar{y}| < \epsilon$.

This process results in a set of samples that approximates the posterior distribution $p(\tau|\mathbf{y})$. Figure 1 presents a simple visual example. The accuracy of this

approximation depends on a number of factors, including the number of iterations (N), the number of babbled sounds (step 2 - m), the acceptance criteria and tolerance rate (step 3 - ϵ), and how closely the learner's data simulation mechanism ($p(\hat{\tau}_k), \sigma_{self}^2$) matches that which produced the observed data (τ, σ^2).

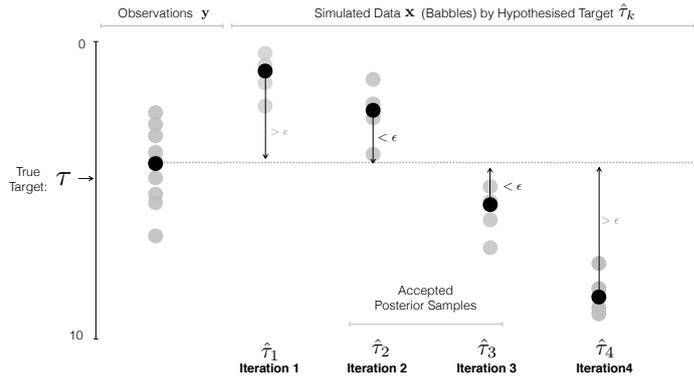


Figure 1. A schematic example of the ABC algorithm as a model of learning: circles represent datapoints – the observed speech sound (leftmost column), and the simulated speech sounds (columns 2-5) – with the sample mean given by the black circle. Only the second and third hypothesised targets τ_2 & τ_3 are accepted as posterior samples, since the mean of the resulting babbles is close to the observed data mean (dotted line).

4.3. Cultural Transmission Among a Population

How does the approximation procedure affect the linguistic structures that emerge over cultural transmission? Existing results for cultural transmission under ideal rational inference (Griffiths & Kalish, 2007) show that the population level distribution of linguistic structures should converge toward the learners' prior distribution: the population level distribution of sound estimates should reflect the prior over τ . By substituting arbitrary parameter values $\tau_0 = 2$ and variance $\sigma_0^2 = 1$ into the Gaussian prior, we can ask how the approximate inference procedure influences this relationship between individuals and populations. Figure 2 shows the mean and variance in the population-level distribution of τ , averaged over 50 simulations of cultural transmission at each point, as a function of the acceptance threshold ϵ . In each of these simulations, 500 generations of single agents in an iterated learning chain (Kirby et al., 2014) transmit τ by listening to the sounds produced by the previous learner, drawing a sample $\hat{\tau}$ from the set of accepted ABC posterior samples, and producing more linguistic data conditional on $\hat{\tau}$ from which the next learner learns. Given the symmetry in the Gaussian representations, the population mean for τ always approximates the prior mean. However,

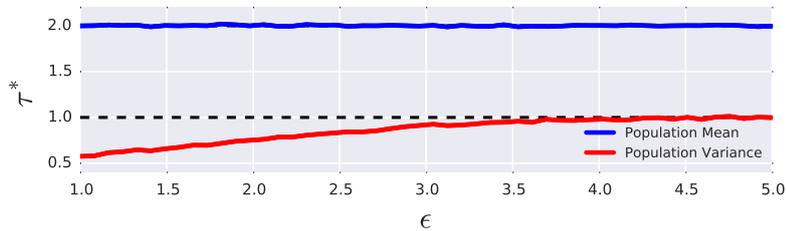


Figure 2. Mean (blue) and variance (red) for τ^* , the population-level distribution of τ , over simulations of 500 generations of learners, average over 50 replications, as a function of ϵ . Dashed lines give the shared prior mean and variance. $n = 20, N = 200, m = 5, \sigma_{self}^2 = 0.1$. See section 4.2 for details of these parameters.

the variance in the population-level distribution of τ only converges to the prior variance under more tolerant acceptance thresholds (higher ϵ). In intuitive terms: if only estimates very close to the target sound are accepted, little cultural evolution occurs; if the tolerance level is weaker, the speech of learners drifts over time to reflect the prior distribution (speech sounds that are easier to produce).

5. Conclusion and Future Directions

We presented a psychologically lightweight computational account for how learners can converge on approximate rational inferences via self-simulation of linguistic data. The fact that learner is also producer broadens the class of learnable linguistic structures that can emerge, persist, and be shaped over cultural transmission: for example, computational principles such as these may make structured systems of speech sounds learnable, and thus evolvable, in the face of unobservable speech-production processes.

The account we developed is maximally abstract: in future work we hope to apply these ideas to specific inductive problems, using articulatory models and acoustic data. Likewise, the ABC procedure we considered is the simplest possible: a fruitful continuation of this research might explore ABC methods which include reinforcement learning, which is known to play an important role during infant speech acquisition (Goldstein, King, & West, 2003), for example. Likewise, the framework provides a natural basis to capture and quantify the effects of *differences* between data generation and simulation mechanisms. For instance, differences between infant and care giver vocal tract morphologies may hinder direct comparison of the spectral features of babbles and adult phonemic categories: reinforcement by caregivers, who know the correspondence between the their own voices and the infant's, may help in category formation, and could be modeled in this framework. A key advantage of our computational proposal is that, when applied to specific problems, existing results for ABC methods make

specific predictions about the trajectory and outcome of learning. Our proposal also predicts, for example, that the order in which articulators develop may influence which speech sounds are acquired first (Esling, Benner, & Moisik, 2015). More generally, the computational principles we have described could apply to any human (or non-human, e.g. birdsong) behaviour that is culturally transmitted via a production mechanism that the learner cannot observe but can simulate herself. Understanding the computational principles that allow data-simulation - in speech or via more general capacities such as pragmatic reasoning - to aid learning from others is an important step toward understanding the evolution of our flagship cultural behaviour, human language.

References

- Abbot, J. T., Hamrick, J. B., & Griffiths, T. L. (2013). Approximating Bayesian inference with a sparse distributed memory system. In M. Knauf, M. Pauen, M. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society*. Berlin: Cognitive Science Society.
- Beaumont, M. A. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 379–406.
- Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: a simple sequential algorithm for approximating Bayesian inference. *Cognitive psychology*, 74, 35–65.
- Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive science*, 34(7), 1131–57.
- Esling, J. H., Benner, A., & Moisik, S. R. (2015). Laryngeal Articulatory Function and Speech Origins. In H. Little (Ed.), *Proceedings of the 18th international congress of phonetic sciences satellite event: The evolution of phonetic capacities*. Glasgow, UK.
- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13), 8030–5.
- Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: exploring the effects of context. *Cognition*, 112(1), 21–54.
- Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive science*, 31(3), 441–80.
- Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, 7(2), 217–29.
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of communication disorders*, 39(5), 350–65.

- Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current opinion in neurobiology*, 28C, 108–114.
- Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C. (2009). Towards a neurocomputational model of speech production and perception. *Speech Communication*, 51(9), 793–809.
- Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, 118(3), 306–38.
- Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, 37(03), 607–642.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13(4), 707–784.
- Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychological review*, 117(4), 1144–67.
- Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic bulletin & review*, 17(4), 443–64.
- Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America*, 104(33), 13273–8.
- Vihman, M. (1991). Ontogeny of Phonetic Gestures: Speech Production. In N. Hillsdale (Ed.), *Modularity and the model theory of speech perception: Proceedings of a conference to honor alvin m. liberman* (pp. 69–84). Erlbaum.

GENETIC DRIFT EXPLAINS SAPIR'S DRIFT IN SEMANTIC CHANGE

IGOR YANOVICH

*Carnegie Mellon University / Universität Tübingen
Pittsburgh, PA, USA / Tübingen, Germany
igor.yanovich@uni-tuebingen.de*

The linguistic notion of “Sapir’s drift” refers to the phenomenon when genetically related languages, long after their separation, undergo the same or very similar linguistic changes. Such “drift” may seem almost magical: given that language change is generally a random process, why would separate linguistic varieties exhibit the same change? There exist possible explanations demystifying Sapir’s drift, including (Joseph, 2012) who argues that if the sister languages all possessed the same variation in a given construction, that variation can serve as the basis for parallel changes long after the languages separate. Here, I propose another, complementary, explanation for Sapir’s drift for changes involving grammatical or semantic reanalysis. The new account is based on evolutionary modeling in the finite-population setting. All finite populations show the effect called *genetic drift* (unrelated to Sapir’s drift) that delays the effect of forces pushing the language in a particular direction. For reanalysis-based changes, this means that even when reanalysis of individual utterances could already occur in the proto-language, the full takeover by the new form may under the right conditions happen many centuries later, in the proto-language’s descendants. Given the introduced model, it would have been surprising if Sapir’s drift never arose, and not the fact that it does.

1. Phenomenon of interest: **motan* modals in the Germanic

Though my argument is general, it is best illustrated with some specific linguistic puzzle. I use for that purpose parallel semantic changes in **motan* modal verbs in Germanic languages.^a Today, this common Germanic word turned into English *must*, Dutch *moeten* and German *müssen*, all of which are necessity modals, roughly synonymous with “it is necessary that...”. But in the early historically attested Germanic languages, **motan* had a very different meaning. The traditional story is that it was then a possibility modal, roughly synonymous with “it’s possible that...” (see (Ono, 1958) for Old English, (Lühr, 1997) for Old High German, (Tellier, 1962) for Old Saxon, a.o.) Meanwhile, (Yanovich, 2015) and to some extent (Standop, 1957) argued instead that **motan* had a special, neither necessity nor possibility, meaning. We do not assume here Standop’s or Yanovich’s specific semantic analyses, but do assume with them that early **motan* must have meant

^aThe word is written with a star because “motan” is the reconstructed form of the verb’s infinitive that was never observed in actual texts. Finite forms of the verb are actually attested.

something special: if it were a regular possibility modal, we would expect a similar development to necessity to happen sometimes to other possibility modals in other linguistic families. But such changes have not been reported.

Here is why the semantic change of **motan* is a Sapir's-drift puzzle. In the early Germanic texts from the 8-9-10th centuries, which already exhibit very different languages (cf. 1), **motan* still was not a necessity modal. In English, it turned completely into such only by the 16th century; in Dutch and High German, even later. Thus the same change eventually occurred to **motan* in those related languages, but it only happened after many centuries of independent development. This is a case of Sapir's drift. Informally, there must have been something about the early **motan* which made it predestined for such a change; however, its effect was delayed for a long time.

(1) **9th century translations of *Pater Noster* into Germanic languages:**

fater unser, du pist in himilum, kauuihit si namo din *Old High German*

ure fæder, þe art on heofonum, sy þin nama gehalgod *Old English*

fadar ūsa <...>, thu bist an them hōhon himila rīkea,
geuuihid sī thīn namo *Old Saxon*

'Our Father, who is in {heaven (OHG, OE) / heaven's high kingdom (OSax)}, let your name be sacred'

We will solve the puzzle of **motan* using explicit evolutionary modeling of semantic change through reanalysis. The crucial assumption deriving differing onsets of change in the different Germanic languages is that the stability of **motan*'s old meaning was due to the forces of *genetic drift*, always observed in finite populations. The role of genetic drift is to delay the progression of an already predestined change. (Sapir's drift and genetic drift are unrelated notions, creating an unfortunate homonymy.)

2. Framework for the evolutionary modeling of semantic change

The items subject to evolution in our model are meaning-form pairs. We will refer to **motan* with the old meaning as A, and to **motan* with the new necessity meaning as B.^b A and B serve the role of alleles of the same gene in biological evolution. At a given discrete stage of evolution, our speaker community produces *n* instances of **motan*, each of them A or B. We look only at this pool of **motan*

^bThis is an oversimplification, though it is unlikely to do harm in our case. In reality, the number of meanings and sub-meanings that **motan* could have is much greater than two (see e.g. (Yanovich, 2015)). In particular, the meaning that seems to have replaced the old meaning was specifically that of *circumstantial* necessity, as opposed to other types of necessity (see (Lühr, 1997) for Old High German, (Yanovich, 2015) for Old and Middle English). To keep modeling simple, we abstract away from the complexities of multiple meanings.

utterances, not at the grammars of individual speakers. (A remark on how to model some grammar effects is made on p. 8.) For simplicity, we assume that n is kept constant across stages. All utterances at the same stage are treated equally, so a given population state may be described simply as a vector $\vec{x} = (x_A, x_B)$, where x_A is the number of As.

An utterance *reproduces*, i.e. gets a descendant at the next stage, through being used as the model for a new utterance. In the normal case, reproduction will be faithful — e.g., an old instance of A will give rise to a new A. But As and Bs are subject to semantic reanalysis, or *mutation*: when a speaker uses some past A-utterance as the model, she may have misanalyzed it as a B-utterance, or vice versa. In this case, an old A may give rise to a new B, or the other way round. The mechanism of semantic reanalysis is by definition necessary for a meaning change to take place, (Eckardt, 2006) a.o.: for A-meaning-form pairs to turn into B-pairs, some speakers must mistake some As for Bs. Furthermore, if As may be mistaken for Bs in some suggestive contexts, some Bs will usually also sometimes be mistaken for As. But the rates of A-to-B and B-to-A reanalysis (in symbols, $r_{A \rightarrow B}$ and $r_{B \rightarrow A}$) are generally not equal.

Finally, types of form-meaning pairs may be associated with different *fitnesses* $f_A(\bar{x}) \neq f_B(\bar{x})$. Fitness is by definition the measure of reproductive success: some types may serve as the model more frequently, and fitnesses capture that. We do not have to take a stand on what causes differences in fitness, but one possible interpretation is as follows: fitness stems from better communicative efficiency. For example, it may be that given the rest of the linguistic system, it is extremely useful to utter some B tokens. In this case, old B tokens, when available, would be used as a model more frequently, i.e. have higher fitness. The mechanism causing utterances with higher fitness to reproduce more frequently is called *selection*.

The introduced framework conceptualizes semantic change as a familiar and reasonably well-studied type of finite-population evolutionary process guided by selection, mutation and genetic drift. Using it, we can apply well-known analytical machinery to puzzles in semantic change, including Sapir's drift of **motan*.

3. When drift is stronger than mutation, Sapir's drift may ensue

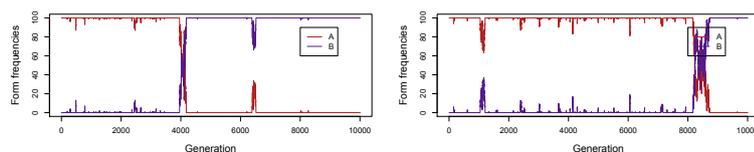
The old meaning of **motan* was relatively stable, as it persisted for many centuries in several independent Germanic languages. Such stability may come from three sources in our model: (i) selection, based on fitness (e.g. communicative efficiency), significantly favored A-**motan* over the new B-**motan*; (ii) B→A semantic reanalysis occurred much more frequently than A→B reanalysis; and (iii) it was B that was favored by mutation and selection, but due to the strength of genetic drift in a relatively small population of utterances, the initial all-A state persisted for a long time before crossing over to all-B.

Reasons (i) and (ii) do not work well for **motan*: under them, we need to stipulate that a great change happened either to the relative usefulness, (i), or to the

reanalysis potential of **motan*, (ii), to explain the subsequent A→B development in later Germanic languages. We do not know, of course, that such drastic changes did not happen. But presumably such changes would be random events. If so, they would be very unlikely to occur independently in long-separated languages. Thus if we explain **motan*'s drift under (i) or (ii), and thus through drastic changes, we then need to explain why those changes were coordinated. In other words, we just kick the explanatory can down the road.

But reason (iii) is different. It effectively says that the A→B change was lurking under the surface in the common Germanic because the possibility for A→B reanalysis of *individual utterances* was already there. But under particular parameters of the process, the eventual takeover by B can be inevitable yet delayed, possibly for a very long time. More precisely, when genetic drift is sufficiently strong relative to mutation and selection, it will take a long time for sporadic A→B reanalysis to cause a systemic change, even though on the level of individual utterances it will always create some B-noise. Example trajectories for such delayed B-takeover are given in 2.

(2) **Sample trajectories for delayed B-takeover** ($r_{A\rightarrow B}=10^{-4}$, $r_{B\rightarrow A}=10^{-5}$)



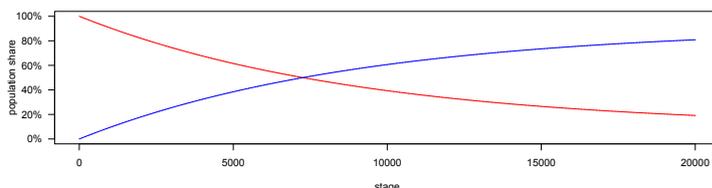
What is *genetic drift* anyway? Imagine an infinite population of A utterances that have a 1% chance of being reanalyzed as B, i.e. $r_{A\rightarrow B} = 0.01$. For simplicity, assume equal fitnesses: $f_A = f_B$. The share of B will then be steadily increasing with each stage, at least until a certain point where back-reanalysis will become strong enough to stop the B-growth.^c As the population is infinite, there will always be A-individuals mutating into B.

Now imagine the same process happening to a population of 10 A-utterances. Most likely, the next stage's population will not feature any Bs at all. Even if it will (which will happen with $p=0.1$), at the next stage the new B can still die out without offspring simply because reproduction is a random process. It is mathematically necessary that in a finite population, uniform states will be more favored. The smaller the population, the more bias towards uniformity there is. The mathematical force causing that is called genetic drift. Importantly, genetic drift by itself does not care about the actual states: A is as good as B for it. Bias for A-to-B shifts is created by asymmetry in mutation, selection or both. Genetic drift, in

^cFor example, with $r_{A\rightarrow B} = 10^{-4}$, $r_{B\rightarrow A} = 10^{-5}$, that point is when the share of Bs reaches $\frac{1}{11}$. It is easy to check that $\frac{1}{11}r_{B\rightarrow A} = \frac{10}{11}r_{A\rightarrow B}$, i.e. the inflow and outflow become equal.

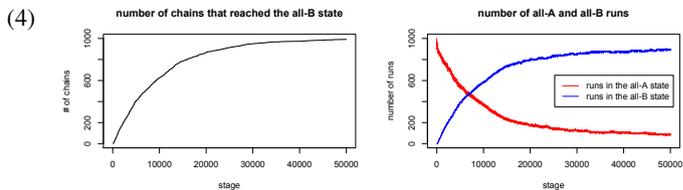
its turn, may delay the effects of that bias.

(3) **Infinite-population trajectories for $r_{A \rightarrow B} = 10^{-4}$, $r_{B \rightarrow A} = 10^{-5}$**



It is instructive to compare the drift-dominated trajectories in 2 with the trajectories we would observe in an infinite population, without stochasticity, 3. In the stochastic case 2, the process stays close to the all-A state, and the jumps between the states are rapid. In the deterministic, infinite-population case 3, the share of Bs starts to gradually rise from the very start. No bounces back into all-A occur, the system just moves gradually and steadily towards the equilibrium point.

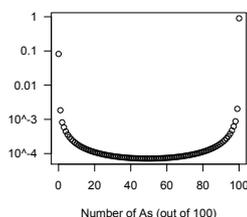
Graphs in 4 show how the same drift-dominated process as in 2 behaves statistically, on the scale of 1000 simulations starting from identical conditions. The left graph shows how many of the runs (*y* axis), all starting in an all-A state, passed through an all-B state by a given stage (*x* axis). To make this more concrete, imagine that the change in **motan*'s meaning proceeded exactly according to the laws of our evolutionary process. English, Dutch and High German would each correspond to a randomly placed horizontal line on the plot. Where that line would cross the graph would determine the timing of a complete switch away from the non-necessity meaning. It is easy to see that the switch would be inevitable in the long run, but would likely happen at quite different times for the three languages.



The right plot in 4 shows the number of runs, out of the same 1000, which happen to be in the all-A (red) or the all-B (blue) state at a given stage. We can note that initially, the share of runs in the all-B state rises sharply, but then starts to level off. The share of runs in the all-A state does not go to 0 either. In fact, both of them tend towards specific *equilibrium values*. As both A-to-B and B-to-A reanalysis is possible, our evolutionary process will not only reach all-B starting from all-A, but will also inevitably return back to all-A after having been in all-B.

The long-term behavior of our process can be studied analytically by using its stationary distribution S . As long as transitions between all states are possible and there is a state which can exactly reproduce itself at the next stage (as all our population states are)^d, by the Perron-Frobenius theorem there is a unique vector S that describes how much time the process will spend in each state in the limit of eternity. Computing S for large n is computationally prohibitive, but for small n , even a personal computer can do the work fast enough.^e The stationary distribution for the same population size and reanalysis rates as we used above are given in 5. The shares of runs in the all-A and all-B states will in the limit tend towards the stationary distribution's shares of all-A and all-B. In our examples, the red and blue lines on the right graph of 4 tend towards 8.2% and 90.0%. The rest 1.8% of the time the process is in the intermediate states.

(5) The states' time shares for $r_{A \rightarrow B} = 10^{-4}$, $r_{B \rightarrow A} = 10^{-5}$, $f_A = f_B$ (log scale).



For cases of Sapir's drift, such long-term behavior is not a good approximation. Examples with constant switching back and forth between the archaic and innovative forms are not generally found in historical-linguistic literature. But this does not undermine the usefulness of our analysis for two reasons. First, in realistic linguistic cases, the system would often switch from a B-dominated state to a further innovative meaning-form type C. If the outflow from all-B states to C, D, and so forth down the reanalysis chain, is strong enough, the process will rarely have time to return to all-A.

Second, cases of constant and rapid back-and-forth reanalysis were not something that historical linguists looked for. The standard thinking about language

^dOur evolutionary process is a Markov chain. The conditions described in the text are irreducibility and aperiodicity.

^eBy Perron-Frobenius, S is the probability left eigenvector of the transition matrix T of the process. Rows of T correspond to the current population state x , columns, to the next population state y . The entries t_{ij} of T state the probability of getting from state x_i to state y_j . Let i and j in t_{ij} refer to the number of A-utterances out of n in the current and next states respectively, and let's call individuals with an A-parent A-offspring. Then:

$$t_{ij} = \sum_{0 \leq k \leq n} p(k \text{ A-offspring in } y \mid x_i) \sum_{0 \leq l \leq j} f_{bin}(l; k, 1 - r_{A \rightarrow B}) f_{bin}(j - l; n - k, r_{B \rightarrow A})$$

where $f_{bin}(a; b, p)$ is the probability of a successes after b trials with probability p of success.

change does not feature a scenario when there is a constant flux between two states (as opposed to cases of stable variation when two forms coexist for a long time). So when linguists observe form B to be frequent in year 1500, and then again in year 1600, the default assumption is that the later uses are direct descendants of the earlier ones. In our evolutionary model, this is not necessarily so. If the process jumps frequently between all-A and all-B, then there could have been a complete absence of Bs for some decades between 1500 and 1600; the new abundance of Bs at 1600 may be unrelated by descent to the old abundance. What enabled both cases of abundance in such a scenario would be the general potential for reanalysis from A to B in the language. Summing up, one reason why we do not find descriptions of erratic back-and-forth trajectories in the historical literature may be the fact that such trajectories have not been conceptualized before as possible for language change.

Let's conclude our analysis. There are three main cases for the behavior of our evolutionary process. (i) When asymmetrical reanalysis rates or asymmetrical fitness strongly dominate genetic drift, we see different languages starting from the same condition following roughly the same time course, with a little noise due to the stochastic nature of the process. This would be "regular", non-delayed linguistic development. (ii) As genetic drift becomes stronger, the time course of a change caused by the same asymmetry will become very different. The change will eventually happen in any language with the same starting conditions, but it may occur at quite different times (recall the distribution on the left plot in 4). Most of the time the modeled language will be in a uniform or almost-uniform state (see 2 for two examples, and 5 for the equilibrium shares of the uniform states). If the time-scale of our observations is comparable with the time scale of the sharply rising part of the curve in the left graph in 4, we will see languages undergoing the same A-to-B change with different timing. This is the condition for Sapir's drift. (iii) Finally, if drift dominates, but the time scale of our observations is such that we can only see the near-equilibrium portion of the curve on the left in 4 rather than its sharply rising portion, we will see rapid shifts between A and B happening from time to time, followed by periods of relative stability. Whether there exists actual changes that follow this pattern is an open empirical question.

4. Conclusion: Towards generalizing the model

We have seen above that a very simple model of language change predicts the existence of Sapir's drift. What we have *not* seen is whether our simple model can be an adequate approximation for actual language change. First, speaker's grammars are conspicuously absent from analysis. Yet we know that however usage-based language evolution may be, utterances are produced by speakers who may have their idiosyncratic constraints as to what to say and how to understand what they hear. Another conspicuous oversimplification is that our model above treats all linguistic utterances as equally accessible for all speakers (who are implicit, but

present in the model). A more realistic model would involve several subpopulations of utterances, corresponding to different speakers or tightly knit speech communities, with some exchange of utterances between them. However, both preliminary theoretical analysis and computer simulations suggest that accounting for those more complex conditions would not make Sapir's drift disappear.

Second, Sapir's drift is predicted to ensue only for particular sets of values of the evolutionary parameters in the model, including the size of the utterance population n , reanalysis rates $r_{A \rightarrow B}$ and $r_{B \rightarrow A}$, fitnesses of forms f_A and f_B , number of stages passed t , and also the mode of reproduction for utterances (i.e. the probability distribution over the number of offspring, see (Der, Epstein, & Plotkin, 2011)). I have not, however, presented here any mapping from real-world linguistic observations to parameters of the model. Thus we know that the model predicts Sapir's drift, but we do not know if it predicts it for reasonable parameter ranges.

One reason I have refrained from providing any specific mapping is that lifting the model to more realistic cases has crucial consequences for the parameter ranges that create Sapir's drift situations. For example, suppose we include into the model the following effect of individual speaker grammars: reanalyzing A as B is easier if the speaker can already produce some Bs than if they have to innovate B without any prior direct evidence of its existence. In this case, the process will become much more strongly biased towards uniform states than by genetic drift alone. Thus very different predictions about the number of stages needed for Sapir's drift's emergence will be made by the simpler and more complex models.

That said, there is hope for getting reasonable empirical estimates for some of the evolutionary parameters. For example, if we equate one evolving pool of utterances with the memory pool of one speaker, we can then use the simple model as it stands to estimate, using both experimental and single-speaker-corpus techniques, the effective size of the utterance pool for real-life speakers (i.e. n), the effective rates of reanalysis of given constructions (i.e. different r s) and/or their relative fitnesses (i.e. f s), and the scale of t . We can then lift those results into a more complex model by looking at the linguistic output of whole speech communities as recorded in corpora and other types of language surveys.

In other words, estimation of the parameters of the model will not be an easy task, but it does not appear *a priori* impossible. If this line of research bears fruit, we will get a set of explicit models of language change that can be used for inferring crucial evolutionary characteristics of diachronic processes, and in effect for probabilistically predicting when and which language change can take place.

Acknowledgements

I am grateful to Adrian Brasoveanu, Gerhard Jäger and audiences at CMU and UCSC for useful discussions, and to the anonymous reviewers for EVOLANG for their valuable comments which made the text better. I also gratefully acknowledge the support of the Alexander-von-Humboldt foundation and of CMU Philosophy.

References

- Der, R., Epstein, C. L., & Plotkin, J. B. (2011). Generalized population models and the nature of genetic drift. *Theoretical Population Biology*, 80, 80-99.
- Eckardt, R. (2006). *Meaning change in grammaticalization. an enquiry into semantic reanalysis*. Oxford: Oxford University Press.
- Joseph, B. D. (2012). A variationist solution to apparent copying across related languages. In *Copies vs. cognates in bound morphology*. Brill.
- Lühr, R. (1997). Zur Semantik der althochdeutschen Modalverben. In G. Fritz & T. Gloning (Eds.), *Untersuchungen zur semantischen entwicklungsgeschichte der modalverben im deutschen* (pp. 159–176). de Gruyter.
- Ono, S. (1958). Some notes on the auxiliary *motan. *Anglica*, 3(3), 64-80.
- Standop, E. (1957). *Syntax und Semantik der modalen Hilfsverben im Altenglischen* magan, motan, sculan, willan. Bochum-Langendreer: Pöppinghaus.
- Tellier, A. (1962). *Les verbes perfect-présents et les auxiliaires de mode en anglais ancien: (viii^e s. - xvi^e s.)*. Paris: C. Klincksieck.
- Yanovich, I. (2015). *Old English *motan, variable-force modality, and the pre-supposition of inevitable actualization*. (to appear in *Language*)

Abstracts

TRIADIC ONTOGENETIC RITUALIZATION: AN OVERLOOKED POSSIBILITY

EKATERINA ABRAMOVA

*Department of Philosophy, Theology and Religious Studies, Radboud University
Nijmegen, The Netherlands
e.abramova@ftr.ru.nl*

This paper aims to contribute to the currently popular theory that language has emerged from manual gestures. I argue that in the discussion of potential mechanisms that could have given rise to a gestural protolanguage, one particular possibility has so far been overlooked, namely *triadic ontogenetic ritualization*.

Existing gestural theories typically propose that a gestural protolanguage was composed of a type of natural pantomimes, i.e. manual movements that resemble some actions or objects. Such accounts argue that pantomimes could have emerged through e.g. bodily mimesis (Zlatev, 2008) or an extension of the mirror neuron system (Arbib, 2012). Importantly, whatever the particular claim, they typically maintain that a sender consciously produces an iconic gesture with an intention to evoke a particular thought in the mind of the receiver and elicit a desired response via a recognition of this intent.

Put shortly, on a currently dominant view, a gestural protolanguage is a system characterized by two features. First, it is characterized by the properties of the signals that comprise it, which are triadic (can iconically stand for actions or objects) and intersubjective. Second, it is characterized by the properties of the cognitive mechanisms underlying pantomime-based communication, in which the ability to produce and comprehend communicative intentions is seen as a central achievement in the evolution towards human language. From this point of view, simpler cognitive mechanisms such as those available to contemporary non-human primates, i.e. mechanisms that subservise ontogenetic ritualization (OR), are seen as insufficient. OR is a process of transformation from recurrent instrumental social actions which are used to affect the behavior of a conspecific into communicative actions (Tomasello & Zuberbühler, 2002). This process is said to yield signals that do not stand for anything, are asymmetric between senders and receivers and produce responses due to over-trained elicitation, thus missing the two properties described above.

In the current paper, following Bar-On (2013), I argue that the traditional view that adopts a Gricean view on communication, as captured by the second feature

of a gestural protolanguage, i.e. communicative intentions, runs into the problem of constructing the gestural protolanguage on a cognitive basis which is itself left unexplained. I also suggest that OR can in fact lead to a language precursor with proto-semantic properties^a – matching the first feature. All that is required is envisioning conditions in which ritualized signals are triadic (involve objects without referring to them) and intersubjective (the signal can be used equally by senders and receivers). Such conditions could be provided by a socio-cultural setting, in which joint action that involves objects, such as tool use, is common^b.

Tool use has been argued to be implicated in language evolution (Stout & Chaminade, 2012) but so far solely as a bridge between hierarchical structure of complex action-sequences and the emerging syntactic structure. What is needed for my proposal is a recognition that interaction around objects is an inherently social activity. Artifacts are often constructed and used together, “in a chain of complementary actions guided largely by anticipation of what the other participant will do” (Reynolds, 1993, p. 412). If at some point in our evolutionary history, interactions around objects became more complex and complementary, ritualized signals could emerge from recurring episodes, where participants learned to anticipate each other’s actions from increasingly abbreviated intention-movements directed towards those objects.

The signals resulting from such a process would be imperative gestures for acting on an object in a specific way. In that sense they would be triadic and modestly referential. Furthermore, if such interactions were to happen among grown-up skilled individuals, the roles could be interchanged. That is, the same individual could have the chance to both produce and respond to a gesture, lending the signals intersubjective quality. Finally, coordination with multiple individuals on different occasions would introduce a pressure to converge on shared forms, establishing a trajectory in which gestures become more symbolic as they become more shared. In such a scenario, postulating a generatively open set of pantomimes, which then needs to be replaced with speech, is not necessary. Rather, the system grows in tandem with a growth of types of interactions available until a point is reached where novel communicative forms can emerge through the processes of cultural creativity and transmission (Brighton, Smith, & Kirby, 2005).

^aThe claim is, that whatever mechanisms are found to underpin *dyadic* OR are also involved in the *triadic* case, albeit perhaps extended in important ways that have to do with other mechanisms involved in joint action. Naturally, empirical studies and computational simulations of TOR would be needed to establish whether, how and under what conditions such mechanisms could lead to an emergence of a gestural proto-language.

^bFor my proposal, any *triadic* activity that takes place regularly would be adequate. The advantage of tool use is the availability of objects that persist across different interactions and can structure them in consistent ways.

Acknowledgements

This work is part of the research project PGW-12-32, which is financed by the Netherlands Organisation for Scientific Research (NWO).

References

- Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis*. Oxford: Oxford University Press.
- Bar-On, D. (2013). Origins of meaning: Must we 'go Gricean'? *Mind & Language*, 28(3), 342–375.
- Brighton, H., Smith, K., & Kirby, S. (2005, September). Language as an evolutionary system. *Physics of Life Reviews*, 2(3), 177–226.
- Reynolds, P. C. (1993). The complementation theory of language and tool use. In K. Gibson & T. Ingold (Eds.), *Cognition, tool use, and human evolution* (pp. 407–428). Cambridge: Cambridge University Press.
- Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1585), 75–87.
- Tomasello, M., & Zuberbühler, K. (2002). Primate vocal and gestural communication. In *The cognitive animal: Empirical and theoretical perspectives on animal cognition* (pp. 293–229). Cambridge: MIT Press.
- Zlatev, J. (2008). From proto-mimesis to language: evidence from primatology and social neuroscience. *Journal of Physiology, Paris*, 102(1-3), 137–151.

**PRE AND POST PARTUM WHISTLE PRODUCTION OF A
BOTTLENOSE DOLPHIN (*TURSIOPS TRUNCATUS*) MOTHER-
CALF DYAD**

AUDRA AMES

*Department of Psychology, University of Southern Mississippi,
Hattiesburg, MS, USA
audra.ames@eagles.usm.edu*

SARA WIELANDT, DIANNE CAMERON

*Six Flags: Discovery Kingdom
Vallejo, California, USA
sWielandt@sfp.com, dCameron@sftp.com*

STAN KUCZAJ

*Department of Psychology, University of Southern Mississippi,
Hattiesburg, MS, USA
s.kuczaj@usm.edu*

Vocal learning has been defined as the process by which an individual modifies its vocal repertoire as a result of the sound environment (Marler, 1976a; McCowan & Reiss, 1997). Vocal learning has been found in a number of species, including the Atlantic bottlenose dolphin (*Tursiops truncatus*) (see review: Janik & Slater, 1997). Dolphins mimic sounds heard in their sound environment, and modify their vocal repertoire based on this mimicry. Much of the research on vocal learning in bottlenose dolphins has focused on whistle development in dolphin calves (Fripp et al., 2005; McCowan & Reiss 1995; Morisaka, Shinohara, & Taki, 1995; Tyack & Sayigh, 1997). In particular, this literature focuses on the sounds calves may select as part of their signature whistles and has suggested that a calf's signature whistle is strongly influenced by the sounds it hears, especially those sounds produced by conspecifics (Bojanowski, Veit, & Todt, 2000; Fripp et al., 2005; McCowan & Reiss, 1995; Sayigh, Tyack, Wells, & Scott, 1990; Sayigh, Tyack, Wells, Scott, & Irvine, 1995). Few studies have addressed the ontogeny of whistle development during a calf's early life (McCowan & Reiss, 1995; Morisaka et al., 1995). Studies that have attempted to discuss the early development of calf acoustic repertoires, rarely address the roles of adult dolphin signature whistles in this development. For example, bottlenose dolphin mothers increase their signature whistle rates surrounding the birth of their calf (Fripp & Tyack, 2008; Gnone & Moriconi, 2010; Mello & Amundin, 2005), and it has been suggested that this increase could serve as a model for dolphin calves in the development of their own signature whistle (Fripp & Tyack, 2008).

Our findings suggest that this may not be the case, and that increased signature whistle rates in bottlenose dolphin mothers must exist for an alternate reason. In this study, we investigate a calf's developing whistle repertoire as a function of the adult whistles it hears. The signature whistles of a group of five adult bottlenose dolphin females in managed care were recorded from a four-month period (two months prior to and two months after the birth of a calf to one of the group members). We gathered video recordings with hydrophone input, which allowed us to observe sound in the environment as it was simultaneously produced with behavior. As has been previously reported for other pregnant dolphins, the mother dolphin exhibited a significant increase ($p < .05$) in whistle production before the birth of the calf. The mother also produced whistles at an additional significant rate ($p < .05$) following the birth of the calf, but this rate decreased over time, consistent with findings from similar studies (Fripp & Tyack, 2008; Gnone & Moriconi, 2010). The remaining four adult females produced whistles at relatively low levels pre-partum, showing an increase in whistle rates as the mother dolphin gradually decreased her vocal production over the first two months of the calf's life. These findings are more supportive of additional hypotheses regarding this phenomenon (e.g., imprinting; Mann & Smuts, 1998; Fripp & Tyack, 2008), not with whistle modeling. We used a discriminate analysis to determine which signature whistles present in the calf's environment, if any, existed in her early repertoire. Parameters from whistles identified as belonging to the calf or as a match to an adult signature contour, were extracted from the data using sound analysis software (RavenPro 1.5). These parameters included the beginning, end, minimum, maximum and delta frequencies of the whistle, along with the whistle duration and inflection points. We did not find that the calf developed a predominant whistle type in the early months of her life, but instead used each of the adults' signature whistles in addition to several whistles that were dissimilar to the adults' signature sounds through the course of the study. Based on our findings, it does not appear that repeated exposure to a sound guarantees that sound's use by the calf. The adult contours most commonly mimicked by the calf were that of a female who produced moderate levels of her signature whistle during the course of the study. This is consistent with findings that suggest a calf may select their signature whistle based on sounds that are not over abundant in their environment (Bojanowski et al., 2000; Fripp et al., 2005). The selection process for sounds that calves include in their vocal repertoire is still largely unknown, but studies that add to the dearth of information regarding this developmental process may help us piece together how dolphins develop their communication system.

Keywords: dolphin communication, vocal learning, signature whistle

References

- Bojanowski, E., Veit, F., & Todt, D. (2000). The development of a bivocal signature whistle in a bottlenose dolphin calf. *European Research on Cetaceans*, 14, 70-74.
- Fripp, D., & Tyack, P. (2008). Postpartum whistle production in bottlenose dolphins. *Marine Mammal Science*, 24(3), 479-502. doi:10.1111/j.1748-7692.2008.00195.x
- Fripp, D., Owen, C., Quintana-Rizzo, E., Shapiro, A., Buckstaff, K., Jankowski, K., Wells, R., & Tyack, P. (2005). Bottlenose dolphin (*Tursiops truncatus*) calves appear to model their signature whistles on the signature whistles of community members. *Animal cognition*, 8(1), 17-26.
- Gnone, G., & Moriconi, T. (2010). Use and function of distinctive whistle-like signals in bottlenose dolphins (*Tursiops truncatus*) mother-calf pair. In A. G. Pearce & L. M. Correa (Eds.) *Dolphins: Anatomy, Behavior and Threats* (pp. 149-167). Hauppauge: Nova Science Publishers.

- Janik, V., & Slater, P. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, 26, 59-99.
- Mann, J., & Smuts, B. B. (1998). Natal attraction: Kidnapping, allocare, and separations among wild bottlenose dolphin (*Tursiops truncatus*) mothers and infants. *Animal Behaviour*, 55, 1097-1113.
- Marler, P. (1976). An ethological theory of the origin of vocal learning. *Annals of the New York Academy of Sciences*, 280, 386-395.
- McCowan, B. & Reiss, D. (1995). Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins, Delphinidae *Tursiops truncatus*: a re-evaluation of the signature whistle hypothesis. *Ethology*, 100, 193-209.
- McCowan, B., & Reiss, D. (1997). Vocal learning in captive bottlenose dolphins: A comparison with humans and nonhuman animals. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 178-207). Cambridge: Cambridge University Press.
- Mello, I., & Amudin, M. (2005). Whistle production pre- and post-partum in bottlenose dolphins (*Tursiops truncatus*) in human care. *Aquatic Mammals*, 31, 169-175.
- Morisaka, T., Shinohara, M., & Taki, M. (2005). Underwater sounds produced by neonatal bottlenose dolphins (*Tursiops truncatus*): I. Acoustic characteristics. *Aquatic Mammals*, 31(2), 248.
- Sayigh, L. S., Tyack, P. L., Wells, R. S., & Scott, M. D. (1990). Signature whistles of free-ranging bottlenose dolphins *Tursiops truncatus*: stability and mother offspring comparisons. *Behavioral Ecology and Sociobiology*, 26(4), 247-260.
- Sayigh, L., Tyack, P., Wells, R., Scott, M., & Irvine, A. (1995). Sex difference in signature whistle production of free-ranging bottlenose dolphins, *Tursiops truncatus*. *Behavioral Ecology and Sociobiology*, 36(3), 171-177.
- Tyack, P. L., & Sayigh, L. S. (1997). Vocal learning in cetaceans. In C. Snowdon & M. Hausberger (Eds.) *Social influences on vocal development* (pp. 208-233). Cambridge: Cambridge University Press.

Noise in Phonology Affects Encoding Strategies in Morphology

DAVID H. ARDELL, NOELLE M. ANDERSON, BODO WINTER

*Quantitative and Systems Biology and Cognitive and Information Science Programs,
University of California, Merced, CA 95306 U.S.A.
dardell@ucmerced.edu*

As with the evolution of a population's genetic variability, the evolution of human linguistic variability must be shaped by multiple interacting forces. The iterated learning paradigm (for overview, see Kirby, Griffiths & Smith, 2014) demonstrates that languages can evolve compositional structure when there is a learning bottleneck: Learners infer a linguistic system from limited input, requiring them to generalize beyond what they observe. Through this, linguistic patterns that are systematically structured become more frequent in the process of cultural evolution.

Besides the 'transmission bottleneck' (Hurford, 2002), the social composition of languages has been argued to be another force acting upon language structure. The 'Linguistic Niche Hypothesis' (Lupyan & Dale, 2010) proposes that morphological complexity is inversely correlated with population size. The mechanism behind this correlation is commonly assumed to be a learning difficulty of adult second language learners in acquiring specifically morphology (Bentz & Winter, 2013; Trudgill, 2011). However, crucially, the major share of evidence for the Linguistic Niche Hypothesis is correlational, leaving the underlying mechanism underspecified (Nettle, 2012).

An additional mechanism explaining the loss of morphological complexity in larger populations may be phonological variability. Adult learners introduce heterogeneity (effectively noise) into the phonological system (Nettle, 2012: 1833-1835). Larger populations harbor more pronunciation variants, paralleling the higher 'noise' present in large populations in the form of stochastic genetic variation. In a large population of speakers, noise is incorporated via contact with other dialects or because of second language learners with different accents. Because morphological markers generally rest on limited phonetic material, they are susceptible to ambiguity if phonological turnover in a population of speakers is high. Using a sequential strategy (i.e., different words/ word order changes) to mark the same contrast in meaning will be a more robust encoding strategy in high-noise signaling channels (Nettle, 2012).

A signal space in an iterated learning framework in principal has multiple dimensions by which they could evolve to preserve the structure of a meaning space. We wish to demonstrate clearly that ILM chains evolve so as to be robust to transmission noise by allocating important differences in meaning to the most reliable dimensions of transmission in signal space. We argue that perhaps the presence of noise causes the self-organization in encoding known as structure-preservation, as is also seen in genetic codes (Sella & Ardell, 2002).

Although effects of dimensionality and noise have been discussed (e.g., Little, Eryilmaz & de Boer, 2015), systematic quantitative study of how meanings get embedded in signal spaces of different sizes and structures in the ILM is still missing.

Integrating ideas from the evolution of the genetic code, we propose a computational architecture that addresses the role of noise in the ILM framework when dimensions of the signal space and population size are modulated. We aim specifically to demonstrate the transition from a morphological/paradigmatic to a syntagmatic/sequential strategy as phonological turnover increases. We predict that within parameter regions without added noise, ILM chains break evenly across these two orthogonal dimensions of compositionality. Under our hypothesis, the introduction of noise into the transmission of one of these dimensions will disrupt the stability of induction and expression and the languages will evolve robustness to this noise. We discuss our hypothesis in light of recent contradictory experimental results (Atkinson, Kirby & Smith, 2015). Through our model, we attempt to demonstrate that noise in phonology biases against paradigmatic systems with morphological markers relying on minimal phonological elements. Rather than contradicting the Linguistic Niche Hypothesis, the proposed results from our study will provide an alternative mechanism for population-dependent effects on the evolution of language structure.

References

- Atkinson, M., Kirby, S. & Smith, K. (2015). Speaker Input Variability Does Not Explain Why Larger Populations Have Simpler Languages. *PLoS One*, *10*, e0129463
- Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics & Change*, *3*, 1-27.
- Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, *2*, 177-226.
- De Boer, B., & Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Advances in Complex Systems*, *15*, 1150021.
- Hurford, J.R. (2002). Expression/induction models of language evolution: dimensions and issues. In Briscoe, E.J. (ed.), *Linguistic evolution through language acquisition* (pp. 301–344). Cambridge: Cambridge University Press.
- Kirby, S., Griffiths, T.L., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108-114.
- Little, H., Eryilmaz, K. & de Boer, B. (2015). Linguistic Modality Affects the Creation of Structure and Iconicity in Signals. In D. C. Noelle, et al. (Eds.), *Proc. of the 37th Ann. Conf. of the Cognitive Science Society* (pp. 1392-1398). Austin, TX
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS one*, *5*(1), e8559.
- Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1829–1836.
- Sella, G. & Ardell, D. (2002). The impact of message mutation on the fitness of a genetic code. *Journal of Molecular Evolution*, *54*, 638–651.
- Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.

EVOLUTION OF LANGUAGE FROM THE APHASIA PERSPECTIVE

ALFREDO ARDILA

Department of Communication Sciences and Disorders
Florida International University
Miami, Florida, USA
ardilaa@fiu.edu

Despite the potential significance aphasia research and knowledge has towards understanding the origin of human language, there is limited interest in the use of the aphasia model to approach language evolution (Code, 2011). Some authors (e.g., Bickerton, 2007) have emphasized that there are two central issues in language evolution: (a) how symbolic units (words or manual signs) evolve, (b) how syntax evolve. It has been suggested that symbolic units (i.e. lexicon) and syntax (i.e. grammar) are the only real novelties in human communication systems, and are therefore the most important points to approach in a theory on language evolution. That is, a theory of language evolution should explain how the lexicon and how the grammar appeared in human history.

Aphasia is generally defined as the loss or impairment of language caused by brain damage (Benson & Ardila, 1996). Different subtypes of aphasia syndromes are often mentioned in neurology and cognitive neurosciences, including: Broca's aphasia, Wernicke's aphasia, conduction aphasia, amnesic aphasia, transcortical aphasia, etc. The exact number of aphasia subtypes depend on the particular classification, but usually ranges between four to seven. Seemingly, this suggested diversity of aphasic syndromes has obscured the major and basic distinction in aphasia: there are only two major aphasic syndromes (Ardila, 2010).

These two fundamental aphasic syndromes are associated with a disturbance at the level of the language elements (lexical/semantic) in the Wernicke's aphasia, or at the level of the association between the language elements (grammatical) in Broca's aphasia. It has been further observed that these two basic dimensions of language (lexical/semantic and grammatical) are related to two basic linguistic operations: *selecting* (that means, the language as a paradigm) and *sequencing* (that means, language as syntagm) (Jakobson & Halle, 1956). Lexicon and grammar not only depend on different brain circuitries and areas (temporal and frontal-subcortical) (Ardila, Bernal & Rosselli, 2016), as well as are impaired by different brain pathologies (Wernicke's and Broca's aphasia), but also are mediated by different types of learning (declarative and procedural) (Ullman, 2004).

During child's initial language development, language appears as a lexical/semantic system. Grammar begins with the ability to combine two words to create a new higher level unit (a syntagm, two or more linguistic elements that occur sequentially in the chain of speech and have a specific relationship). Analyzing language development in children, Brown (1973) proposed that most of the utterances when beginning grammar development could be described by a small set of functional relationships between words, such as "agent + action" (baby kiss), "action + object" (pull car), and "agent + object" (daddy ball).

Essentially, three stages of language evolution could be distinguished: (a) Primitive communication systems similar to those observed in other animals, including non-human primates; (b) initial communication systems using sound combinations (lexicon), probably appearing thousands and even millions of years ago, correlated with the enlargement of the temporal lobe; (c) complex communication systems including not only a lexicon but also word-combinations (grammar) (Ardila, 2015).

Origins of the Lexical/Semantic System. To understand the origins of language, it is crucial to consider the evolution of the brain areas involved in language processing, such as the temporal lobe (lexical/semantic system). It is known that in monkeys, the temporal lobes participate in recognizing the sounds and calls of their own species (Tagliatela et al., 2009). Hence, the temporal lobe plays a crucial role in auditory communication not only in humans but also in nonhuman primates. The increase of the temporal lobe's dimensions may be related to the complexity of the human auditory communication system. It is interesting to note that the temporal lobe directly participates in the recognition of the own species sounds, and the superior temporal gyrus contains neurons that are tuned to species-specific calls. It has been further suggested that temporal lobe differences between humans and nonhuman primates relate to the temporal lobe volume (Rilling & Seligman, 2002).

Origins of the Grammatical System. It has been suggested that verbs, grammar, and speech praxis (generated spatiotemporal specifications for skilled purposeful articulatory movements) appeared simultaneously in history (Ardila, 2009). Interestingly, grammar, speech praxis, and the ability to use verbs are simultaneously impaired in cases of Broca's area damage, suggesting a common neural activity. So, the origin of grammar is directly linked to the ability to use verbs and the ability to produce certain articulatory movements.

Observations with children's language development and experiments with nonhuman primates demonstrate that language initially appears as a lexical/semantic system. Grammar, on the other hand, is correlated with the ability to use verbs and represent actions. This is an ability that depends on, the so-called Broca's area and related brain circuits. Most likely, this last stage in language evolution is only observed in *Homo sapiens*. Grammar probably originated from the internal representation of actions, resulting in the creation of verbs.

Clinical observations demonstrate that the ability to use and understand grammar is impaired in cases of the so-called Broca aphasia (e.g., Papathanasiou, Coppens & Potagas, 2012). This disturbance in the use of grammar is known as **agrammatism**. Agrammatism is also observed in language understanding; these patients have difficulties understanding sentences whose meanings depend on their syntax. We can consequently assume that grammar is supported by certain specific brain areas and brain circuits that are precisely impaired in Broca aphasia.

It has been further suggested that grammar may represent the basic ability for the development of metacognitive executive functions (such as abstracting, problem solving, metacognition, temporality of behavior, etc.) (Ardila, 2008).

References

- Ardila, A. (2008). On the evolutionary origins of executive functions. *Brain and cognition*, 68(1), 92-99.
- Ardila, A. (2009). Origins of the language: correlation between brain evolution and language development. In: S. M. Platek and T. K. Shackelford (eds.), *Foundations of Evolutionary Cognitive Neuroscience*. New York: Cambridge University Press.
- Ardila, A. (2010). A proposed reinterpretation and reclassification of aphasic syndromes. *Aphasiology*, 24(3), 363-394.
- Ardila, A., Bernal, B., & Rosselli, M. (2016). How localized are language brain areas? A review of Brodmann areas involvement in language. *Archives of Clinical Neuropsychology*
- Benson, D.F. & Ardila, A. (1996). *Aphasia: A clinical perspective*. New York: Oxford
- Bickerton, D. (2007). Language evolution: A brief guide for linguists. *Lingua*, 117, 510–526.
- Brown, R. (1973). *A First Language: The Early Stages*. Boston: Harvard University Press.
- Code, C. (2011). Nonfluent aphasia and the evolution of proto-language. *Journal of Neurolinguistics*, 24(2), 136-144.
- Jakobson, R. & Halle, M. (1956). *Two Aspects of Language and Two Types of Aphasic Disturbances*, New York: Mouton.
- Papathanasiou, I., Coppens, P. & Potagas, C. (2012). *Aphasia and Related Neurogenic Communication Disorders*. Jones & Bartlett Learning.
- Ullman, M.T. (2004). Contributions of memory circuits to language: the declarative/ procedural model, *Cognition*, 92(1-2), 231-270
- Rilling, J.K. & Seligman, R.A. (2002). A quantitative morphometric comparative analysis of the primate temporal lobe. *Journal of Human Evolution*, 42(5), 505–533.
- Tagliatela, J.P., Russell, J.L., Schaeffer, J.A. & Hopkins, W.D. (2009). Visualizing vocal perception in the chimpanzee brain. *Cerebral Cortex*, 19(5), 1151–1157.

TOWARDS AN ACTION-ORIENTED APPROACH TO THE EVOLUTION OF LANGUAGE AND MUSIC

RIE ASANO

*Department of Systematic Musicology, University of Cologne,
Cologne, Germany
rie.asano@uni-koeln.de*

Language and music considered as cognitive systems form a mosaic, consisting of multiple components as parts with different evolutionary origins (Boeckx, 2013; Fitch, 2006). From a comparative perspective, some of these components might be shared and based on the same evolutionary genesis, while others might be different and emerged independently in the course of evolution. Till now, theoretical as well as empirical research proposed several candidates for shared and distinct components (Jackendoff, 2009; Koelsch, 2012; Patel, 2008; Peretz, 2013). However, their evolutionary origins and the way how those several components work together remains still unclear. In the current paper, this issue is discussed within an action-oriented framework exploring language and music in terms of goal of action, action planning, motor control, and sensory-motor integration (Asano & Boeckx, 2015). This framework makes it possible to investigate similarities and differences of cognitive systems at the same time.

Current findings suggest that shared structural and processing aspects are not specific to language and music: Complex action—also organized in an asymmetrical hierarchy—comprises temporal integration processes and involves neural resources shared with language and music (Fitch & Martins, 2014; Jackendoff, 2009; Koelsch, 2012). Moreover, action-based research provides an opportunity to consider the issue of cognitive phylogenies (Boeckx & Fujita, 2014; Fitch & Martins, 2014). Action is namely a basic form of hierarchically structured temporal sequence processing involved in ‘higher’ cognitive systems and is shared with non-human primates. An action-based research strategy allows to avoid all-or-nothing contrastive approaches which are investigating only the pinnacles of mental evolution and have been thus criticized repeatedly (De Waal & Ferrari, 2010; Theofanopoulou & Boeckx, 2015). Furthermore, an action-

based research strategy permits to examine ‘humaniqueness’ (Hauser, 2009) and uniqueness of each cognitive system in a more fine-grained way.

In the current paper, I focus on domain-general, action-based neurocognitive mechanisms playing an important role in the evolution of musical rhythm and discuss their relationship to the evolution of speech and language in terms of temporal integration processes. Especially, the basal ganglia, the motor cortico-basal ganglia-thalamo-cortical circuit as well as the dorsal stream including Broca’s area are discussed as a set of domain-general neural structures undergoing significant changes in the course of the evolution in our lineage (Merchant & Honing, 2014; Merchant et al., 2015; Patel & Iversen, 2014; Rauschecker & Scott, 2009). Based on those findings from cognitive and evolutionary neuroscience and within an action-oriented framework, I propose how the cognitive systems language and music emerged on the basis of shared basic neurocognitive mechanisms.

Acknowledgements

Special thanks to Prof. Uwe Seifert, Clemens Maidhof, and three anonymous reviewers for their helpful comments.

References

- Asano, R., & Boeckx, C. (2015). Syntax in language and music: What is the right level of comparison? *Frontiers in Psychology, 6*, 00942.
- Boeckx, C. (2013). Biolinguistics: forays into human cognitive biology. *Journal of Anthropological Sciences, 91*, 1–28.
- Boeckx, C., & Fujita, K. (2014). Syntax, action, comparative cognitive science, and Darwinian thinking. *Frontiers in Psychology, 5*, 627.
- De Waal, F. B. M., & Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Sciences, 14*(5), 201–7.
- Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition, 100*(1), 173–215.
- Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences, 1316*(1), 87–104.
- Hauser, M. D. (2009). The possibility of impossible cultures. *Nature, 460*(7252), 190–6.
- Jackendoff, R. (2009). Parallels and Nonparallels between Language and Music. *Music Perception, 26*(3), 195–204.
- Koelsch, S. (2012). *Brain and Music*. Chichester, West Sussex; Hoboken, NJ: Wiley-Blackwell.
- Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., & Fitch, W. T. (2015).

- Finding the beat: a neural perspective across humans and non-human primates. *Phil. Trans. R. Soc. B*, 370, 20140093.
- Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, 7(274).
- Patel, A. D. (2008). *Music, language, and the brain*. Oxford, New York: Oxford University Press.
- Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, 8, 57.
- Peretz, I. (2013). The Biological Foundations of Music: Insights from Congenital Amusia. In D. Deutsch (Ed.), *The Psychology of Music* (3rd ed., pp. 551–564). London: Academic Press.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.
- Theofanopoulou, C., & Boeckx, C. (2015). Cognitive phylogenies, the Darwinian logic of descent, and the inadequacy of cladistic thinking. *Frontiers in Cell and Developmental Biology*, 3(October), 1–6.

**ON A MUSIC-READY BRAIN: NEURAL BASIS, MECHANISMS,
AND THEIR CONTRIBUTION TO THE LANGUAGE
EVOLUTION**

RIE ASANO

*Department of Systematic Musicology, University of Cologne,
Cologne, Germany
rie.asano@uni-koeln.de*

EDWARD RUOYANG SHI

*Department of General Linguistics, Universitat de Barcelona
Barcelona, Spain
edwardshiruoyangend@gmail.com*

In the past decades, the concept of a ‘language-ready’ brain has stimulated language research from the perspective of evolution and development. However, parallels in research on the music-ready brain is still in its infancy (Arbib & Iriki, 2013; Seifert & Kim, 2006). To promote a comparison at the level of music- and language-readiness, we suggest beat induction as a promising starting point because it seems to be an important feature for every musical culture (Nettl, 2000) and this capacity is already available in newborns and therefore might be innate (Honing et al., 2009; Trainor & Corrigan, 2010; Winkler et al., 2009; Zentner & Eerola, 2010), but it is less pronounced in other primates (Honing et al., 2012; Zarco et al., 2009). The current paper is divided into two sections. In the first section, we propose four fundamental mechanisms for beat induction, namely hierarchical structure processing, auditory-motor coupling, prediction, and social interaction. Thus, we regard beat induction as an active process rather than a mere perception of auditory beats. In the second section, we discuss two approaches investigating evolutionary origins and neurocognitive mechanisms of beat induction. Their relation to components of a language-ready brain is discussed in terms of two (out of seven) criteria introduced by Arbib (2005). One approach emphasizes the role of basal ganglia and dorsal pathway as well as the motor cortico-basal ganglia-thalamo-cortical circuit which give rise to the domain-general properties involved in beat induction and vocal learning, for

example, prediction (Vuust & Witek, 2014) and auditory-motor coupling (Merchant & Honing, 2014; Merchant et al., 2015; Patel & Iversen, 2014; Patel, 2006). Concerning a component of the language-ready brain, these neural mechanisms are hypothesized to be involved in temporal sequence processing, i.e. mapping hierarchical structure to temporal order.

The other approach stresses mechanisms of social interaction as central to investigate the nature of beat induction (Fitch, 2012). We propose to extend the social approach by pointing out the relevance of social learning (Tomasello, 1996) with our Social Learning Hypothesis which claims that imitation-based social learning mechanisms which emerged on the scaffolding of mirror neuron systems shared with monkeys and apes are involved in beat induction. This hypothesis is in line with the view regarding music as primarily social domain (e.g. Overy and Molnar-Szakacs, 2009; Cross, 2011, 2012; Seifert et al., 2013) and is supported by studies of children in social contexts (e.g. Kirschner and Tomasello, 2009) as well as studies of non-human primates in interactive contexts (e.g. Konoike et al., 2012; Nagasaka et al., 2013). Moreover, the generation of auditory temporal prediction during sensorimotor synchronization recruits medial cortical areas (e.g. medial prefrontal cortex) (Pecenka et al., 2013), which is involved in social neural networks (Vogele & Roepstorff, 2009). Concerning another component of the language-ready brain, complex imitation and its neural correlates in connection with mirror neuron systems is suggested to get more attention in future research on beat induction.

It is worth noting that our social learning hypothesis doesn't deny the biological basis of the beat induction claimed by Honing (2013). Rather, it provides a new possibility to investigate biological foundations of beat induction and their relation to those of speech or language in the light of social cognition. An integrative approach of biological and social perspectives introduced in our paper provides important implications for the growing field of social cognitive neuroscience as well as cultural neuroscience (Han et al., 2013; Knoblich & Sebanz, 2008; Lieberman, 2007; Vogele & Roepstorff, 2009), playing a significant role in research of language and music evolution.

Acknowledgements

Special thanks to Prof. Uwe Seifert and three anonymous reviewers for their helpful comments.

References

Arbib, M. A. (2005). From monkey-like action recognition to human language:

- An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28(02), 105–124. <http://doi.org/10.1017/S0140525X05000038>
- Arbib, M. A., & Iriki, A. (2013). Evolving the Language- and Music-Ready Brain. In M. A. Arbib (Ed.), *Language, Music, and the Brain* (pp. 481–497). Cambridge, MA: The MIT Press.
- Cross, I. (2011). The meanings of musical meanings: Comment on “Towards a neural basis of processing musical semantics” by Stefan Koelsch. *Physics of Life Reviews*, 8(2), 116–9. <http://doi.org/10.1016/j.plrev.2011.05.009>
- Cross, I. (2012). Cognitive science and the cultural nature of music. *Topics in Cognitive Science*, 4(4), 668–77. <http://doi.org/10.1111/j.1756-8765.2012.01216.x>
- Fitch, W. T. (2012). The biology and evolution of rhythm: unravelling a paradox. In P. Rebuschat, M. Rohrmeier, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 73–95). Oxford, New York: Oxford University Press. <http://doi.org/10.1093/acprof>
- Han, S., Northoff, G., Vogeley, K., Wexler, B. E., Kitayama, S., & Varnum, M. E. W. (2013). A cultural neuroscience approach to the biosocial nature of the human brain. *Annual Review of Psychology*, 64, 335–359. <http://doi.org/10.1146/annurev-psych-071112-054629>
- Honing, H. (2013). Structure and Interpretation of Rhythm in Music. In D. Deutsch (Ed.), *The Psychology of Music* (Third Edit, pp. 369–404). London: Academic Press. <http://doi.org/10.1016/B978-0-12-381460-9.00009-2>
- Honing, H., Ladinig, O., Háden, G. P., & Winkler, I. (2009). Is beat induction innate or learned? Probing emergent meter perception in adults and newborns using event-related brain potentials. *Annals of the New York Academy of Sciences*, 1169(Illc), 93–6. <http://doi.org/10.1111/j.1749-6632.2009.04761.x>
- Honing, H., Merchant, H., Háden, G. P., Prado, L., & Bartolo, R. (2012). Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PloS One*, 7(12), e51369. <http://doi.org/10.1371/journal.pone.0051369>
- Kirschner, S., & Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, 102(3), 299–314. <http://doi.org/10.1016/j.jecp.2008.07.005>
- Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1499), 2021–2031. <http://doi.org/10.1098/rstb.2008.0006>
- Konoike, N., Mikami, A., & Miyachi, S. (2012). The influence of tempo upon the rhythmic motor control in macaque monkeys. *Neuroscience Research*, 74(1), 64–7. <http://doi.org/10.1016/j.neures.2012.06.002>

- Lieberman, M. D. (2007). Social Cognitive Neuroscience: A Review of Core Processes. *Annual Review of Psychology*, 58(1), 259–289.
<http://doi.org/10.1146/annurev.psych.58.110405.085654>
- Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Phil. Trans. R. Soc. B*, 370, 20140093.
<http://doi.org/10.1098/rstb.2014.0093>
- Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, 7(274).
<http://doi.org/10.3389/fnins.2013.00274>
- Nagasaka, Y., Chao, Z. C., Hasegawa, N., Notoya, T., & Fujii, N. (2013). Spontaneous synchronization of arm motion between Japanese macaques. *Scientific Reports*, 3, 1151. <http://doi.org/10.1038/srep01151>
- Nettl, B. (2000). An Ethnomusicologist Contemplates Universals in Musical Sound and Musical Culture. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 463–472). London: The MIT Press.
- Overy, K., & Molnar-Szakacs, I. (2009). Being Together in Time: Musical Experience and the Mirror Neuron System. *Music Perception*, 26(5), 489–504. <http://doi.org/10.1525/mp.2009.26.5.489>
- Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, 24(1), 99–104. <http://doi.org/10.1525/mp.2006.24.1.99>
- Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, 8, 57.
<http://doi.org/10.3389/fnsys.2014.00057>
- Pecenka, N., Engel, A., & Keller, P. E. (2013). Neural correlates of auditory temporal predictions during sensorimotor synchronization. *Frontiers in Human Neuroscience*, 7(August), 380.
<http://doi.org/10.3389/fnhum.2013.00380>
- Seifert, U., & Kim, J. H. (2006). Musical meaning : Imitation and empathy. In *Proceedings of the 9th International Conference on Music Perception & Cognition* (pp. 1061–1070). Bologna: ICMPC & ESCOM.
- Seifert, U., Verschure, P. F. M. J., Arbib, M. A., Cohen, A. J., Fogassi, L., Fritz, T., ... Rickard, N. (2013). Semantics of Internal and External Worlds. In M. A. Arbib (Ed.), *Language, Music, and the Brain* (pp. 203–229). Cambridge, MA: The MIT Press.
- Tomasello, M. (1996). Do Apes Ape? In C. M. Heyes (Ed.), *Social Learning in Animals: The Roots of Culture* (pp. 319–346). London: Academic Press.
- Trainor, L. J., & Corrigall, K. A. (2010). Music acquisition and effects of musical experience. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Music acquisition and effects of musical experience*.
- Vogele, K., & Roepstorff, A. (2009). Contextualising culture and social

- cognition. *Trends in Cognitive Sciences*, 13(12), 511–516.
<http://doi.org/10.1016/j.tics.2009.09.006>
- Vuust, P., & Witek, M. a. G. (2014). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, 5(October), 1–14.
<http://doi.org/10.3389/fpsyg.2014.01111>
- Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America*, 106(7), 2468–71.
<http://doi.org/10.1073/pnas.0809035106>
- Zarco, W., Merchant, H., Prado, L., & Mendez, J. C. (2009). Subsecond Timing in Primates: Comparison of Interval Production Between Human Subjects and Rhesus Monkeys. *Journal of Neurophysiology*, 102(6), 3191–3202.
<http://doi.org/10.1152/jn.00066.2009>
- Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13), 5768–73. <http://doi.org/10.1073/pnas.1000121107>

ADULT LANGUAGE LEARNING AND THE EVOLUTION OF LINGUISTIC COMPLEXITY

MARK ATKINSON

*Behaviour and Evolution Research Group,
Psychology, School of Natural Sciences
University of Stirling, FK9 4LA, UK
mark.atkinson@stir.ac.uk*

KENNY SMITH, SIMON KIRBY

*Centre for Language Evolution,
School of Philosophy, Psychology and Language Sciences,
University of Edinburgh, EH8 9AD, UK
kenny@ling.ed.ac.uk, simon@ling.ed.ac.uk*

The pressures shaping languages may differ in different physical, demographic, and sociocultural environments. In other words, non-linguistic factors may systematically determine linguistic features (Lupyan & Dale, 2010; Trudgill, 2011; Nettle, 2012). Identification of such factors, and the mechanisms by which they operate, will provide valuable insights into how languages evolve to exhibit differing degrees of grammatical complexity, and also shed light on the structural properties of the earliest languages (Wray & Grace, 2007). We present three experiments which investigate the mechanisms linking sociocultural factors and linguistic structure: specifically, we attempt to explain why languages of small social groups tend to be morphologically complex and opaque, while the languages of larger groups tend to be morphologically simpler, more regular and transparent (Wray & Grace, 2007; Lupyan & Dale, 2010).

An explanation of how group size could have such an effect is yet to be convincingly demonstrated (Nettle, 2012). One possible mechanism is the different learning biases and abilities of adults and children (Wray & Grace, 2007; Lupyan & Dale, 2010; Nettle, 2012; Atkinson, Kirby, & Smith, 2015): languages with more speakers tend to have a greater proportion of non-native speakers, and these adult learners may acquire and subsequently transmit simplified morphological systems. We assess this claim experimentally, exploring not only the generation of simplified morphology by individuals, but also the potential for such simplifications to influence group-level language characteristics.

In Experiment 1, we trained adult participants on a morphologically-complex

miniature language with semantically-redundant suffixes, and then tested them on their ability to reproduce those languages. We found that ‘non-native’ speakers (participants given less input data, which we use as a proxy for the reduced exposure characterising adult learning) produced simpler morphological systems.

In Experiment 2 we focus on how such simplifications could spread through a population. We used the languages produced by our participants in Experiment 1 as the input for a fresh set of learners, manipulating (i) whether this second generation received input drawn from few or many speakers, and (ii) whether those model speakers had received a lot or a little exposure to the original complex target language (representing input from native or non-native speakers). We found that mixing input from multiple speakers nullified the simplifications introduced by non-native learners in Experiment 1. While the output of individuals may be simplified, such simplifications tend to be idiosyncratic, and therefore mixing the output of one or more non-native speakers with the output of native speakers yields a system which is itself complex and variable, and which did not lead to the acquisition of a simpler language by our second generation.

In Experiment 3, we focus on language *use*, and the consequences of the interaction between native and non-native speakers: after training on a miniature language (which includes irregular forms for native learners but only regular forms for non-native learners), participants interact in pairs, taking turns to describe scenes for each other. Our results show that a speaker who acquires a more complex language than their partner may simplify their output in interaction, in particular by using (inferred) regular forms instead of irregulars.

We suggest that adult learning is a plausible explanation for why languages spoken by more people have simpler morphology, but that native speaker accommodation to non-natives is a key linking mechanism: idiosyncratic simplifications by non-natives alone does not offer a complete explanation.

References

- Atkinson, M., Kirby, S., & Smith, K. (2015). Speaker input variability does not explain why larger populations have simpler languages. *PLoS ONE*, *6*, e0129463.
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, *5*, e8559.
- Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *367*, 1829–1836.
- Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Wray, A., & Grace, G. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*, 543–578.

THE EVOLUTION OF WHAT?

CHRISTINA BEHME

*Department of Philosophy, Mount Saint Vincent University,
Halifax, Canada
ChristinaBehme@msvu.ca*

Abstract: In spite of impressive accomplishments there remain significant gaps in our understanding of language evolution. It will be argued that striving for more conceptual clarity will assist empirical research, eliminate fruitless debates, and (re)focus attention away from pseudo-problems. Attention will be paid to analogies that construct powerful conceptual tools in support of specific hypotheses. It will be shown that this use of analogies is not unproblematic and suggested that analogical reasoning should be used more often to challenge one's own hypothesis.

1. Introduction

Noam Chomsky is often credited with reawakening language evolution research from the 'sleeping beauty' existence it had led for more than a century. Decades ago he proposed that the "idea of regarding the growth of language as analogous to the development of a bodily organ is ...quite natural and plausible" (Chomsky, 1975, 11) and that "the language faculty may be regarded as fixed function, characteristic of the species, one component of the human mind" (Chomsky, 1977, 63). These proposals came at a time when Chomsky's generative grammar had "become the conventional wisdom [in linguistics]" (Searle, 1973, 8) and his influence had extended into psychology, philosophy, anthropology, and other fields. Many non-linguists accepted the assumption that language was a biological organ, and attempted to address the question of how such an organ might have evolved. Today we find a thriving language evolution research community that includes experts from anthropology, archaeology, cognitive neuroscience, computer science, genetics, linguistics, neurophysiology, philosophy, physics, primatology, and psychology. These experts have generated an impressive body of work. However, in spite of impressive accomplishments there remain significant gaps in our understanding of language evolution. As observed by one leading language evolution researcher: "many different scholars have reached valid insights about human language ... but no one scholar or discipline has yet achieved an adequately comprehensive overview of this complex system" (Fitch, 2010, 2). Not undeservedly, language evolution has been called "the hardest problem of science" (Christiansen & Kirby, 2003).

2. What is language?

It seems that before debating about the most promising routes towards substantial progress researchers need to acknowledge that there is no broadly accepted consensus about the ‘object’ of evolution: language and to specify which definition of language they subscribe to. While virtually everyone assumes something in our biology accounts for our ability to use language, the exact nature of this putative “language organ” or “language instinct” remains a matter of controversy and many questions about the ontological status of language and the exact relationship between language and biology still await satisfactory answers. I am of course not the first to notice that terminology is used ambiguously in the language evolution literature. It has been suggested that language is not a ‘monolithic whole’ but “a complex system made up of several independent subsystems” (Fitch 2010, 17) and that the “unspecified use of ...the word ‘language’ ...is probably best avoided” (Fitch 2010, 24).

This cautionary note is often ignored and many researchers either do not define what they mean by language or they focus on specific aspects of a very complex phenomenon without suggesting how their results are relevant to other aspects or could be integrated into a more complete picture. One example, representative for many articles on language evolution is an excellent survey article by Tallerman (2007). She contrasts synthetic and holistic approaches to language evolution. Synthetic approaches assume that humans used single words first and “that an evolving syntax takes single items and forms structure by combining them” (Tallerman 2007, 580). Holistic approaches on the other hand assume “that protolanguage consists of ‘a small inventory’ of indivisible utterances” (Tallerman 2007, 580), and that words emerge later via a process of fracturing. This seems to suggest that what evolves is the system of words and their possible combinations. Protolanguages are systems that are simpler than and/or lack some of the components of fully evolved languages. However, throughout the article Tallerman also discusses cognitive and brain evolution. Yet, it is often not clear if brain structures are relevant for producing language sounds or if they also underwrite grammatical relations between words. She refers to “a fully modern UG or innate predisposition towards learning language” (Tallerman, 2007, 589) and “a fully modern linguistic brain” (Tallerman, 2007, 594) but it is not clear exactly what is innate on her view: cognitive capacities required to learn and use language, specific information related to grammatical structure, or computational mechanisms underwriting language use. Furthermore, it is also not clear whether ‘innate predispositions’ have to be specific to language or could be domain general.

One might argue that terminological imprecision is unproblematic because individual researchers know what they mean by ‘language’ and the time required to conceive of more precise definitions is better spent conducting empirical research. However, the current situation has several undesirable consequences. First, researchers often get entangled in fruitless debates because they start from

different assumptions about 'language'. In one extreme case it has been suggested that the majority of language evolution researchers hold a 'non-existence thesis' regarding language.

"It is widely held that "There are well-developed gradualist evolutionary arguments that language is entirely grounded in a constellation of cognitive capacities that each -- taken separately - - has other functions as well," in which case language exists only in the sense that today's weather exists: a constellation of factors that have independent functions, not an object of serious scientific inquiry in itself ... All of these approaches share the conception of half a century ago that there is no such thing as language in any serious sense ... under the "non-existence" assumptions, the approaches to UG that have so far led to serious results would have to be reshaped. How, it is idle to speculate without credible proposals" (Chomsky 2011, 34).

Chomsky implies here that the majority of language evolution researchers is committed to the ridiculous view that 'there is no such thing as language in a serious sense'. How could anyone plausibly suggest that virtually the entire language evolution-community studies the evolution of something they do not believe to exist? Upon closer reading it becomes clear that Chomsky's implication is based on two assumptions: (i) that 'language' is equivalent to Universal Grammar (a hypothesized construct of Chomsky's linguistic theorizing) and (ii) that anyone who denies that this Universal Grammar exists holds that language is 'not an object of serious scientific inquiry'. Unfortunately, this uncharitable interpretation of the work of others is aided by the often cavalier use of 'language'. Adopting an 'everyone knows what I mean by 'language' attitude is not productive when there is very good evidence that not everyone knows (or wants to know).

3. Helpful analogies?

Humans are the only animals that use language. This makes it impossible to conduct direct comparative research between human language and animal communication systems or to draw inferences about human language based on animal experiments. This situation invites the use of analogical reasoning in language evolution theorizing. It is worthwhile remembering that "[r]easoning by analogy involves identifying a common relational system between two situations and generating further inferences driven by these commonalities" (Gentner & Smith, 2012, 130). These analogies construct powerful conceptual tools but they do not establish that underlying cognitive phenomena are similar.

However, some of the frequently employed analogies hinder instead of furthering progress. That is because researchers often assume as default what needs to be established: that two phenomena (e.g. human language and monkey alarm calls) are similar in certain aspects. Based on this assumption they focus on similarities between these phenomena and set aside differences.

... if Everett were right, it would show that Piraha doesn't use the 'resources that Universal Grammar makes available'. But that's as if you found a tribe of people somewhere who crawled instead of walking. They see other people crawl, so they crawl. It doesn't show that you can't walk. It doesn't show that you're not genetically programmed to walk [and do walk, if you get the relevant kind of input that triggers it and are not otherwise disabled] (Chomsky 2012, 30)

Chomsky's analogy (between language and walking/crawling) nicely illustrates multiple problems with reasoning by analogy. First, in order for the analogy to be valid the underlying processes (language and the ability to walk being genetically pre-programmed) would have to be the same. But, Everett challenges the claim that language is based on an innate Universal Grammar of the kind proposed by Chomsky. In order to address this challenge Chomsky can not assume what needs to be established (that there is a Chomskyan UG) but needs to provide independent arguments in support of his hypothesis. Second, an analogy should only be used to draw conclusions about the underlying process in two domains when the effects of this process are the same as (or at least similar to) the effects in both domains.

When comparing crawling/walking to language development it becomes clear that the domains are quite different. Most infants go through a crawling stage before they learn how to walk. The transition requires some practice from the first tentative and often wobbly steps to confident walking. But once humans have acquired the skill they hardly 'look back' and one finds rarely any adult regularly crawling. Looking now at the acquisition and use of recursion one notices important differences. First, there seems to be no phase during which children acquire and practice recursion¹. Second, even after children acquire the ability to produce recursive structures they continue to produce sentences not

¹ In early language acquisition children move from a stage in which they use single-word utterances to a stage in which they use multi word utterances. But since Piraha children also go through these stages it cannot be argued that the ability to use multi-word utterances is disputed by Everett (for detailed discussion of this ongoing dispute see Everett (2009) Nevins et al. (2009), Adger (2015), Behme & Evans (2015)). More importantly, the ability to string together several meaningful units is shared by several species and could not be considered a species specific property.

containing such structures. So, unlike in the crawling/walking case there is no (virtually complete) replacement of one behavior by another.

Another difference between the two cases is noteworthy. Anyone hypothesizing that our genetically underwritten 'walking faculty' enables us to add one step to another and repeat this operation indefinitely, could find some empirical confirmation for such a hypothesis: humans are indeed able to perform walks consisting of hundreds or thousands of steps. Looking at the case of language one should find a similar ability. Chomsky insists that "the language faculty is that it is a system of discrete infinity. Any such system is based on a primitive operation that takes n objects already constructed, and constructs from them a new object ... call that operation merge... With Merge available, we instantly have an unbounded system of hierarchically structured expressions. (Chomsky 2005, 11). Allegedly this computational operation was installed in human brains by a single mutation: "The simplest account of the 'Great Leap Forward' in the evolution of humans would be that the brain was rewired, perhaps by some slight mutation, to provide the operation Merge" (Chomsky 2005, 11-12). If this ability to generate an unbounded system of hierarchically structured expressions is indeed the essence of language one would expect that humans use massively long and multiply recursively structured sentences. Yet, the vast majority of sentences actually produced and/or used by humans is pathetically short (compared to a hypothetical discrete infinity): the vast majority of sentences contains fewer than 88 words and a 40,004-word sentence would be neither produced spontaneously nor could it be understood by any normal human.

As the forgoing has shown, in many performance aspects there are considerable differences between walking and language. Hence there do not seem to be (strong) analogies between the two domains. It remains of course possible that in spite of these differences the mechanisms underwriting walking and language are of similar nature. But in order to establish such similarities researchers would need to pursue strategies different from analogical reasoning. In fact, the strong dis-analogies between the two domains could suggest that the working hypothesis (that both language and walking follow innately pre-determined developmental patterns) needs to be re-evaluated.

Reasoning by analogy is widely used in the language evolution literature. Especially when comparing communication systems of non human species and language researchers focus on similarities and support their arguments with analogies. Given that on some level of analysis many phenomena are similar to each other, it is not surprising that those looking for similarities will find them. Many evolutionists prefer gradualist accounts and study the mental mechanisms underlying the communication of our closest relatives, non-human primates. Researchers attempt to establish which aspects of human language are within the non-human primates' capabilities and which are not. When researchers find

similarities between the use of animal vocalizations and human language they sometimes theorize that underlying mechanisms might be similar or that animal calls could be (evolutionary) precursors of language. It has been known for some time that vervet monkeys (*Chlorocebus pygerythrus*) produce predator specific alarm calls for leopard, martial eagle and python. Different alarm calls seem to evoke different responses. Leopard alarm calls prompt the monkeys to climb into trees, eagle alarm calls result in monkeys hiding in nearby bushes, and Python alarm calls prompt the monkeys to cautiously survey the ground. It is well established that the alarm calls convey specific information about approaching predators to other members of the group. Yet, it was assumed that the calls were instinctual, and did not involve vocal learning. Further calls seemed a direct response to a visual stimulus (predator) and not intended to convey benefits to other group members (Seyfarth et al. 1980). These features make alarm calls fundamentally different from human language.

Recent work on other species has attempted to establish closer similarities to human language. For example, 'Contest hoots' are acoustically complex vocalisations produced by adult and subadult male bonobos (*Pan paniscus*). These calls are often directed at specific individuals and regularly combined with gestures and other body signals. They provoke a social reaction in the targeted individual and may function to assert social status. The intentional use of multi-modal sequences to initiate social interactions with important group members could indicate more cognitive complexity than previously attributed to bonobos. (Genty et al. 2013). Since it is assumed that human language requires a great degree of cognitive sophistication one might argue by analogy that cognitive complexity in bonobos indicates that they are cognitively 'language ready'. Further corroboration for the intentionality of primate vocal communication comes from research on chimpanzees (*Pan troglodytes*). In an experiment designed to test for higher order intentionality (the animal knowing what another animal likely knows), researchers presented wild chimpanzees with a python model and found that alarm calls depended on the presumed knowledge level of receivers. The researchers argue that "alarm calls were: (i) socially directed and given to the arrival of friends, (ii) associated with visual monitoring of the audience and gaze alternations, and (iii) goal directed, as calling only stopped when recipients were safe from the predator" (Schel et al., 2013, 1). These findings have been taken to suggest that cognitive capacities required for human language use (the understanding of both one's own and others' mental states and a desire to modify another's mental states) are present in some non-human primates. Because researchers discover analogies between different species, their findings are taken to indicate that the cognitive gulf between humans and other primates may not be as wide as previously assumed.

If there are (strong) language relevant cognitive similarities between humans and non-human primates one has to wonder why we are the only 'fully

linguistic' species. It has been hypothesized that, unlike humans, non-human primates are not able to produce a wide variety of distinct sounds and thus lack an essential prerequisite for spoken language (for overviews see Fitch 2010, Lieberman, 2013). On the other hand, songbirds are able to produce a wide variety of distinct sounds, and it has been shown that they are capable of complex vocal learning, an ability also required for language acquisition. For example, European starlings (*Sturnus vulgaris*) can be trained to recognize acoustic patterns of a complexity comparable to a recursive, self-embedding, context-free grammar. They are able to classify new patterns defined by the grammar and reliably exclude ungrammatical patterns (Gentner et al. 2006). Recent work on Zebra finches has shown that these songbirds can learn to recognize affixations. Affixes have grammatical function in language and recognizing them is one of the many abilities needed in language acquisition. Finches learn to recognize different affix-patterns and show preference for prefixes (Chen et al., 2014). The claim is not that finches recognize affixes as affixes, far less that they are aware of grammatical function. Rather, they are capable of very fine-tuned discrimination, an ability considered crucial for human language production and comprehension. Extensive work on African grey parrots (*Psittacus erithacus*) has shown that these birds can acquire a large vocabulary, learn to differentiate meaning and rudimentary syntax, and can engage in a simple conversation with a human trainer (Pepperberg, 2008). In addition to vocal ability some corvids also show impressive social intelligence, suggesting that these birds have complex cognitive abilities. For example, Scrub jays (*Aphelocoma coerulescens*) remember not only numerous sites where they cached food but also where conspecifics have cached. They pilfer those sites when given the opportunity. It was shown that jays with prior experience of pilfering another bird's caches subsequently re-cached food in new cache sites when they had been observed by other birds caching. This suggests that jays relate information about their previous experience as a pilferer to the possibility of future stealing by another bird, and modify their caching strategy accordingly (Emery & Clayton 2001). Similarly complex abilities have also been confirmed in ravens (*Corvus corax*) (Bugnyar & Heinrich 2005), and New Caledonian crows (*Corvus moneduloides*) (Taylor et al., 2012).

Taken together, the work on non-human animals suggests that several species possess rich cognitive resources underwriting the ability to interpret (and to a lesser degree) produce meaningful acoustic signals. Reasoning by analogy suggests that some of the mechanisms involved in animal cognition are similar to those involved on human language. However, there are also important dis-analogies and no other species has a communication system rivaling human language. Possible reasons for this disparity could be (i) that in spite of similarities in cognitive sophistication the underlying mechanisms are very different or (ii) that only some of the many cognitive and physiological resources needed for language are accessible to members of other species while

only humans have access to the full suite of resources, or (iii) it could be the case that non-human animals are unable to overcome specific barriers to achieving a linguistic communication system. “How humans overcame (or side-stepped) these limitations [remains] a central question for theories of language evolution” (Fitch 2010, 202).

4. Conclusions

Language evolution research conducted in recent years has generated a wealth of exciting results. Nevertheless, there remain significant gaps in our understanding of language evolution and no model that could address the evolution of the full spectrum of human language is currently on the horizon. Language evolution researchers could profit from paying closer attention to conceptual issues. Striving for more conceptual clarity will assist empirical research, eliminate fruitless debates, and (re)focus attention away from pseudo-problems. While it may be neither feasible nor beneficial that all language evolution researchers adopt the same definition of ‘language’ it would be desirable for them to explicitly state which definition they adopt. Analogies used by language acquisition researchers can construct powerful conceptual tools but they do not establish that underlying cognitive phenomena are similar. Given our tendency to interpret the world in ways that are consistent with our hypotheses (confirmation bias), researchers need to resist the temptation to focus on analogies between phenomena they hypothesize to have similar underlying mechanisms. Instead they have to be explicit about the assumptions they are making and consider whether the evidence they gather is inconsistent with those beliefs. This process can be aided if analogical reasoning is used more often to challenge one’s own hypothesis.

References

- Adger, D. (2015). Mythical myths: Comments on Vyvyan Evans’ “The Language Myth” *Lingua*, 158, 76-80.
- Behme, C. & V. Evans. (2015). Leaving the myth behind: A reply to Adger (2015), *Lingua*, 162, 149-159.
- Bugnyar, T., & B. Heinrich. (2005). Food-storing ravens differentiate between knowledgeable and ignorant competitors. *Proceedings of the Royal Society of London, Series B* 272, 1641-1646
- Chen, I., N. Jansen, & C. Ten Cate. (2014). Zebra finches can learn to recognize affixations. *The Evolution of Language Proceedings of the 10th International Conference EVOLANG 10*, 411-412, Vienna, Austria.
- Chomsky, N. (1975). *Reflections on Language*. London: Temple Smith.
- Chomsky, Noam. 1977. On the nature of language. In: *Essays on form and interpretation* (pp. 63 – 80). New York: Elsevier.

- Chomsky, N. (2005). Three Factors in Language Design. *Linguistic Inquiry*, 36, 1-22
- Chomsky, N. (2012). *The Science of Language*. Cambridge: Cambridge University Press.
- Christiansen, M. & Kirby, S. (Eds.), (2003). *Language Evolution*. Oxford: Oxford University Press.
- Emery, N. J. & N. S. Clayton. (2001). Effects of experience and social context on prospective caching strategies by scrub jays. *Nature*, 414, 443 – 446.
- Everett, D. (2009). Pirahã Culture and Grammar: A Response to Some Criticisms. *Language*, 85, 405-442.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.
- Gentner, D. & Smith, L. (2012). Analogical reasoning. In V. S. Ramachandran (Ed.) *Encyclopedia of Human Behavior* (2nd Ed.). pp. 130-136. Oxford, UK: Elsevier.
- Gentner, T. Q., K. M. Fenn, D. Margoliash, & H. C. Nusbaum. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, 440, 1204-1207
- Genty, E., Z. Clay, C. Hobaiter, & K. Zuberbuehler. (2013). Multi-Modal Use of a Socially Directed Call in Bonobos. *PLoS ONE* 9(1): e84738. doi:10.1371/journal.pone.0084738
- Lieberman, P. (2013). *The unpredictable species*. Princeton, NJ: Princeton University Press.
- Nevins, A., C. Rodrigues, & D. Pesetsky. (2009). Pirahã Exceptionality: a Reassessment. *Language* 85.2, 355-404.
- Pepperberg, I. M. 2008. *Alex & Me*. New York: Harper Luxe.
- Savage-Rumbaugh, Sue, & Roger Lewin. (1994). *Kanzi: The Ape at the Brink of the Human Mind*. New York: Wiley
- Schel, Anne Marijke, Simon W. Townsend, Zarin Machanda, Klaus Zuberbuehler, & Katie E. Slocombe. (2013). Chimpanzee Alarm Call Production Meets Key Criteria for Intentionality. *PLoS ONE* 8(10): e76674. doi:10.1371/journal.pone.0076674.
- Searle, John. (1973). Chomsky's revolution in linguistics. In: G. Harman (ed.) *On Noam Chomsky: critical essays*. (pp. 2 – 33). New York: Anchor Books.
- Seyfarth, Robert M., Dorothy L. Cheney, & Peter Marler. (1980). Monkey responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication. *Science* 210, 4471, 801-803.
- Taylor, Alexander. H., Miller, R., & Gray, R. D. (2012). New Caledonian crows reason about hidden causal agents. *Proceedings of the National Academy of Science U S A*, 109 (40), 16389-16391
- Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, 117, 579–604
- Tomasello, M. (2009). *Origins of communication*. Cambridge, MA: MIT Press.

SPONTANEOUS DIALECT FORMATION IN A POPULATION OF LOCALLY-ALIGNING AGENTS

RICHARD A BLYTHE, ALISTAIR H JONES, JESSICA RENTON

*School of Physics and Astronomy, University of Edinburgh
Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK
r.a.blythe@ed.ac.uk, alistair.jones@kcl.ac.uk, s1003707@sms.ed.ac.uk*

A language can be viewed as a shared set of conventions that relate meanings (functions) to signals (forms). These conventions however emerge from episodes of language learning and use between individual members of a group. One important characteristic of these interactions is a ‘principle of density’ advocated by Bloomfield (1933) who states that “every speaker is constantly adapting his speech habits to those of his interlocutors” (quoted by Labov, 2001, p19). By now the question of how local alignment on different conventions can develop into a convention shared by an entire group is well understood, both from modeling (Castellano, Fortunato, & Loreto, 2009; Smith, 2014) and experiment (Centola & Baronchelli, 2015).

A robust prediction of many models is that a state of global consensus—a single set of conventions for the whole group—is almost always reached (Castellano et al., 2009; Blythe, 2015). This therefore precludes the emergence of multiple dialects within a mixing population, by which we mean the existence of multiple equivalent forms for the same function (whether at the phonological, lexical or some other level of linguistic structure) being used by agents who interact. These dialects are clearly a feature of human language, and arguably exist also in certain bird, dolphin and other animal communication systems, perhaps because they have evolutionary utility in allowing kinship or group membership to be discriminated (Fitch, 2000).

There are a few counterexamples to the prediction of certain consensus. If biases that favor a different specific convention within each subgroup pre-exist, a correlation between group membership and linguistic behavior can stably persist (Kandler & Steele, 2008). In this case, dialect formation is not spontaneous: the dialects simply mirror a set of pre-existing biases whose origins remain mysterious. One mechanism that *is* spontaneous, in the sense that individuals have no *a priori* bias towards one behavior over another, arises when individuals preferentially interact with those group members who exhibit similar cultural behaviour (Axelrod, 1997). Then, the network of social influence adapts to reflect differ-

ences in linguistic behavior. Intuition suggests that the converse process, whereby stable differences in linguistic behavior stem from (potentially static) variation in the degree of influence from different speakers, should also occur. This situation corresponds more closely to what is typically understood to be implied by Bloomfield's principle of density (Labov, 2001), but so far coexistence of multiple dialects has been seen only in the extreme case of highly heterogeneous network structures (Dall'Asta, Baronchelli, Barrat, & Loreto, 2006). Blythe and Croft (2012) have proposed a mechanism by which this might occur under more general conditions: namely, a situation whereby agents *acquire* a bias in favor of behavior that correlates with that of influential agents they have previously encountered.

In this work, we show that this mechanism does indeed allow for spontaneous dialect formation under quite general conditions. We demonstrate this in a computational model with two contrasting types of influence structure in the network. When the network is divided into groups within which agents strongly influence one another but between which influence is weak, small initial differences in behavior between the groups can be amplified into distinct stable dialects, as long as the disparity in influence between and within the groups is large. This is consistent with what was already observed by Dall'Asta et al. (2006). However, we find that multiple dialects also emerge when agents are embedded in space and interact only with their neighbors, even if the degree of influence between interacting agents is uniform.

A linear stability analysis can be applied to determine the conditions under which an instability to the emergence of multiple dialects is present; this also shows that the state with multiple dialects is stable to small perturbations. This includes the fluctuations that are a consequence of linguistic interactions comprising a finite amount of data or agents having a finite memory. This result stands in contrast to game-theoretic models where dialects are neutrally stable (Zollman, 2005). In the case where the state of consensus is also stable, and cannot be escaped if reached, it is inevitably the case that a state with multiple dialects will eventually be destroyed. However, we will argue that the fact that speakers are discrete entities with variable behavior implies that multiple dialects may persist for many generations. A further feature of this model is that the bias derived from a historical average of interlocutors' behavior, weighted by influence, can be reinterpreted as a bias in favor of the local majority convention. This *regularization of variation* is typically found to be present in experiments involving linguistic and non-linguistic variation (e.g. Hudson Kam & Newport, 2005; Reali & Griffiths, 2009; Smith & Wonnacott, 2010), and is suggested to derive from such purely cognitive factors as memory limitations or a prior expectation that the world is predictable. This work suggests that Bloomfield's principle of density provides alternative origin of such an effect, in which regularization is a side-effect of variation in social influence between speakers.

References

- Axelrod, R. (1997). The dissemination of culture a model with local convergence and global polarization. *Journal of Conflict Resolution*, *41*, 203–26.
- Bloomfield, L. (1933). *Language*. New York: Henry Holt.
- Blythe, R. A. (2015). Hierarchy of scales in language dynamics. *European Physical Journal B*, *88*, 295.
- Blythe, R. A., & Croft, W. (2012). S-curves and the mechanism of propagation in language change. *Language*, *88*, 269–304.
- Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, *81*, 591–646.
- Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *PNAS*, *112*, 1989–94.
- Dall'Asta, L., Baronchelli, A., Barrat, A., & Loreto, V. (2006). Nonequilibrium dynamics of language games on complex networks. *Physical Review E*, *74*, 036105.
- Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*, 258–67.
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*, 151–195.
- Kandler, A., & Steele, J. (2008). Ecological models of language competition. *Biological Theory*, *3*, 164–73.
- Labov, W. (2001). *Principles of linguistic change: Social factors* (Vol. 2). Oxford: Wiley-Blackwell.
- Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*, 317–28.
- Smith, A. D. M. (2014). Models of language evolution and change. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 281–93.
- Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 444–9.
- Zollman, K. J. S. (2005). Talking to neighbors: The evolution of regional meaning. *Philosophy of Science*, *72*, 69–85.

HOW THE BRAIN GOT GRAMMATICALIZED: GLOBULARIZATION AND (SELF-)DOMESTICATION

CEDRIC BOECKX

*ICREA & Dept. of General Linguistics, Universitat de Barcelona, Barcelona, Spain
cedric.boeckx@ub.edu*

CONSTANTINA THEOFANOPOULOU

*Dept. of General Linguistics, Universitat de Barcelona, Barcelona, Spain
constantinaki@hotmail.com*

ANTONIO BENÍTEZ-BURRACO

*Dept. of Spanish and its Didactics, Universidad de Huelva, Huelva, Spain
antonio.benitez@dfesp.uhu.es*

1. Introduction

This paper seeks to explore a potential connection between two evolutionary hypotheses recently put forward linking the language phenotype and genotype. Both are aimed to cast light on the pre-requisites for more complex communication systems; in the case of our species, grammatical systems. One (Boeckx and Benitez-Burraco 2014 et seq., Theofanopoulou 2015) is the idea that the globularization of the braincase that characterizes our species is the reflex of a genetically-regulated specific brain growth pattern that provided the neural scaffolding for "cognitive modernity", most distinctively our 'language-readiness'. This distinctive growth process is well established (Hublin et al. 2015), and clinical evidence suggests that deviations from this growth trajectory entail cognitive/language deficits (see, e.g., Knight et al. 2014). The other idea concerns (self-)domestication. As Thomas 2014 discussed extensively, self-domestication in our species can prove extremely valuable in understanding the central role played by cultural learning, giving rise to the grammaticalization of our mind. It is now clear that cultural learning appears to be key in capturing all the grammatical paraphernalia that was usually (and misleadingly) assigned to the "Universal Grammar". But as Thomas 2014 points out, a major problem

facing any attempt to account for language structure through a cultural mechanism is that the required processes are only possible if we assume the existence of a range of preconditions, which we may call the “cultural niche” (the ‘cooperative’ niche, as Tomasello would call it). Thomas 2014 thinks that this niche may have been formed by the behavioral, cognitive and temperamental outcomes of self-domestication. Interestingly, Wilkins et al. (2014) have recently put forth the hypothesis that the hypofunction of the neural crest cells (NCCs) during embryonic development in response to external stimuli may result in a constellation of distinctive traits (the “domestication syndrome”).

2. Our hypothesis

Our hypothesis is that the genetic changes that have been claimed to bring about globularization affected the NCC too, thereby fueling the emergence of the (self-)domestication syndrome in our species. To test this hypothesis, we did an exhaustive literature search to determine whether (some of) the “domestication syndrome” genes highlighted by Wilkins et al. are also important for globularization and/or have changed in our lineage compared to Neanderthals and Denisovans (see, e.g., Pääbo 2014). We have also proceeded the other way around: we made extensive search of the literature to learn how many of (and to what extent) the candidates for globularization are involved in the development and function of the neural crest and could be also regarded as “neural crest genes”. The intersection of the two set of genes (encompassing *SOX10*, *SOX9*, *SOX2*, *MTIF*, *MAGOH*, *FGF8*, *EDNRB*, *RET*, *TCOF1*, *BMP7*, *BMP2*, *CDC42*, *CTNNA1*, *DLX5*, *DLX6*, *FGFR1*, *PAX6*, *SHH*, *VCAN* among others) strikes us as particularly promising, given that most of these genes have been implicated in aspects of language and cognition.

3. Conclusions

The data we have gathered suggest to us that a globularized brain and brain case may be intimately connected to the developmental/genetic context for a domestic phenotype, which could then have been selected for the reasons Thomas 2014 discussed. Put another way, the language-ready brain which a globularized brain(case) gave rise to led to full-fledged modern linguistic behavior, the grammaticalization and (self-)domestication of mind.

References

- Boeckx, C., & Benítez-Burraco, A. (2014). The shape of the human language-ready brain. *Frontiers in Psychology*, *5*, 282.
- Hublin, J.-J., Neubauer, S., and Gunz, P. 2015. Brain ontogeny and life history in Pleistocene hominins. *Philosophical Transactions of the Royal Society, B*, *370*, 20140062.
- Knight, S., Anderson, V., and Spencer-Smith, M. & Da Costa, A. (2014). Neurodevelopmental outcomes in infants and children with single-suture craniosynostosis: A systematic review. *Developmental neuropsychology*, *39*, 159-186.
- Pääbo, S. (2014). The human condition – A molecular approach. *Cell*, *157*, 216-226.
- Theofanopoulou, C. (2015). Brain asymmetry in the white matter making and globularity. *Frontiers in Psychology*, *6*, 1355.
- Thomas, J. G. (2014). *Self-domestication and Language Evolution*. Ph.D. Thesis. University of Edinburgh.
- Wilkins, A. S, Wrangham, R., W & Fitch, W. T. (2014). The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, *197*, 795–808.

SIGNATURE WHISTLES IN AN INTRODUCTION CONTEXT

MEGAN BROADWAY¹, JAMIE KLAUS², BILLIE SERAFIN², HEIDI LYN¹

*1 Psychology, University of Southern Mississippi
Long Beach MS, USA
megan.broadway@eagles.usm.edu, heidi.lyn@usm.edu*

*2 Institute for Marine Mammal Studies
Gulfport MS, USA*

One key distinction that has been drawn between human language and animal communication concerns the increased flexibility in human language. In general, communicative systems in animals are considered associative and tied to specific contexts (Scott-Phillips, 2015). Specifically, complex and dynamic vocal communication systems are rare in the animal kingdom, being limited primarily to humans, birds, and delphinids (Janik, 2009). Because these flexible systems are so rare, comparisons between the taxa are important to understand the evolutionary pressures that have led to these systems. Traits which are present across species, such as vocal learning or the ability to reference objects such as the self and others, which are traits shared by humans and dolphins, may be key factors in the evolution of more complex communication systems. Most of what we know about the communication system of delphinids comes from the study of bottlenose dolphins (*Tursiops truncatus*). Although researchers have attempted to decode the communication system of these animals for more than 60 years, the discovery of signature whistles (Caldwell, Caldwell, & Tyack, 1990) has been one of the most promising findings. Signature whistles are distinctive calls that are unique for each individual, and like human language, are a product of vocal learning. Signature whistles primarily act as cohesion calls, and are used in sophisticated contexts such as when groups of dolphins encounter one another in the wild (Quick & Janik, 2012). These calls likely developed due to the limited visibility of the underwater environment and the highly social nature of these animals. It has been suggested that signature whistles may be used self-referentially and to reference others (King & Janik,

2013), similarly to how humans use names (see Janik & Sayigh, 2013). If so, this would be one of the only species to use names to identify individuals and would allow researchers to study the conditions under which reference to self and others arise. Still, the nuances of signature whistle usage remain largely unknown with only a few, un-replicated experimental studies.

One context where these whistles are likely to be used is during the introduction of a new dolphin to an established group. For this study, a new dolphin was introduced to two established residents over an extended period of time by first adding the new individual to an adjoining pool where he was housed for several months and then allowing all three dolphins to swim together freely. Vocalizations and behavioral data were collected before, during, and after the introduction. Underwater vocalizations were recorded using an array of hydrophones to determine if and when signature whistles were used over the course of the extended introduction period (Fig. 1). These data will later be compared to a follow-up study where an additional dolphin was introduced to this group. Further studies based on context dependent interpretations of signature whistles will help to clarify the social and environmental factors that contribute to the evolution of flexible communication systems like human language.

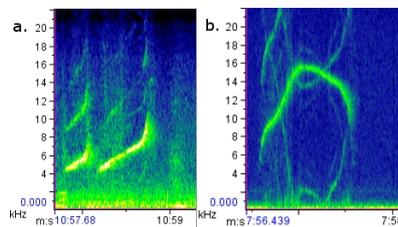


Figure 1. Examples of signature whistles from two of the dolphins.

References

- Caldwell, M. C., Caldwell, D. K., & Tyack, P. L. (1990). Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin. In S. Leatherwood & R. R. Reeves (Eds.), *The bottlenose dolphin* (pp. 199–234). San Diego: Academic Press.
- Janik, V. M. (2009). Acoustic communication in delphinids. *Advances in the Study of Behavior*, *40*, 123–157. [http://doi.org/10.1016/S0065-3454\(09\)40004-4](http://doi.org/10.1016/S0065-3454(09)40004-4)
- Janik, V. M., & Sayigh, L. S. (2013). Communication in bottlenose dolphins: 50 years of signature whistle research. *Journal of Comparative Physiology*, *199*(6), 479–89. <http://doi.org/10.1007/s00359-013-0817-7>
- King, S. L., & Janik, V. M. (2013). Bottlenose dolphins can use learned vocal labels to address each other. *Proceedings of the National Academy of Sciences*, *110*(32), 13216–13221. <http://doi.org/10.1073/pnas.1304459110>
- Quick, N. J., & Janik, V. M. (2012). Bottlenose dolphins exchange signature whistles when meeting at sea. *Proceedings. Biological Sciences / The Royal Society*, *279*(1738), 2539–45. <http://doi.org/10.1098/rspb.2011.2537>
- Scott-Phillips, T. C. (2015). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, *6*(1), 56–80.

HOW DO LAUGHTER AND LANGUAGE INTERACT?

GREGORY A. BRYANT

*Department of Communication Studies, University of California, Los Angeles,
Los Angeles, CA USA
gabryant@ucla.edu*

1. Introduction

Laughter is a universal social vocalization characterized by rhythmic laryngeal and superlaryngeal activity. The sound of laughter varies within and between speakers, but maintains a reasonably stereotyped form, and follows rather specific production rules (Provine, 2000). Acoustic analyses of play vocalizations across several primate species suggest that human laughter is derived from a homolog dating back at least 20 MYA (Davila-Ross, Owren, & Zimmermann, 2009). Human laughter has evolved increased proportions of voiced components, and these features contribute to perceptual judgments of affiliation between speakers and positive affect (Bachorowski & Owren, 2001). But rhythmic characteristics also play an important role in judgments of spontaneity (Bryant & Aktipis, 2014) and playfulness (Nagy & Bryant, 2014). Laughter features might interact in interesting ways with language use that can reveal important aspects of its evolved function.

People laugh in conversation to achieve a variety of pragmatic goals (Flamson & Bryant, 2013), and laughter plays a complex role in negotiating relationships that goes well beyond its connection to humor (Provine, 2000). But the production of spontaneous laughter is likely generated by an emotional vocal system that is separate from the control of articulators during speech production—the so-called dual pathway model of vocal production (e.g., Ackermann, Hage, & Ziegler, 2014; Owren & Rendall, 2011).

2. Studies exploring the interaction of laughter and language

2.1. *Laughter signals play in discourse*

People tend to laugh immediately before and after using indirect speech in which speakers' intentions are not explicitly stated but rich meaning is strategically conveyed. Here I will describe recent research documenting the effect of laughter on the interpretation of verbal irony, a common form of indirect speech. Verbal irony utterances that included adjacent laughter were extracted from natural conversations between friends, and were then manipulated to either include the laughter or not. These utterances were played for listeners (no

listener heard the same utterance twice) and they were asked to rate the indirectness of the speakers' meaning. The presence of laughter increased listeners' judgments of indirectness (Exp. 1). The isolated laughs from these recordings were then played to a different group of listeners and rated for playfulness (Exp. 2). Judgments of playfulness were positively associated with the degree to which laughter increased judgments of indirectness across utterances in the first experiment. These data suggest that spontaneous laughter functions to signal play in social interaction, and sheds light on the relationship between pragmatics and nonhuman animal communication. Play vocalizations induced by tickling in great apes share similarities in both acoustic form and communicative function.

2.2. Laughter and speech production

During conversation, the relationship between interlocutors shapes the way people laugh. For example, the interaction between speech production and laughter production is affected by affiliative status. I will describe recent work using the same corpus of spontaneous conversation recordings showing that, compared to established friends, people who had just recently met embedded laughter into their speech much more frequently (i.e., the laugh bursts constituted speech syllables), suggesting a greater tendency to produce laughs from the speech system as opposed to the phylogenetically older vocal emotion system. The speech system generates laughs with highly recognizable features, and is potentially indicative of social manipulation.

3. Conclusion

These studies represent attempts to explore the role of laughter in signaling social intentions, and potentially cueing social manipulation. The function of human laughter is clearly connected to homologs in other primate species, and its incorporation into human linguistic communication, including pragmatic signaling, provides a fascinating example of how an ancestrally old trait can be integrated with more recent communicative abilities.

References

- Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *37*(6), 529-546.
- Bachorowski, J. A., & Owren, M. J. (2001). Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science*, *12*(3), 252-257.
- Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evolution and Human Behavior*, *35*(4), 327-335.
- Davila-Ross, M., Owren, M., & Zimmermann, E. (2009). Reconstructing the evolution of laughter in great apes and humans. *Current Biology*, *19*, 1106-1111.

- Flamson, T. J., & Bryant, G. A. (2013). Signals of humor: Encryption and laughter in social interaction. In M. Dynel (Ed.). *Developments in linguistic humour theory* (pp. 49-74). Amsterdam: John Benjamins.
- Nagy, M., & Bryant, G. A. (2014). *Laughter as a signal of play in discourse*. Paper presented at the 12th Conceptual Structure, Discourse, and Language (CSDL) conference, Santa Barbara, CA.
- Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, 73(6), 530-544.
- Provine, R. R. (2000). *Laughter: A scientific investigation*. New York, NY: Penguin Press.

CULTURAL EVOLUTION AND COMMUNICATION YIELD STRUCTURED LANGUAGES IN AN OPEN-ENDED WORLD

JON W. CARR, KENNY SMITH

*Centre for Language Evolution, University of Edinburgh
Edinburgh, United Kingdom
j.w.carr@ed.ac.uk, kenny.smith@ed.ac.uk*

HANNAH CORNISH

*Psychology, University of Stirling
Stirling, United Kingdom
hannah.cornish@stir.ac.uk*

SIMON KIRBY

*Centre for Language Evolution, University of Edinburgh
Edinburgh, United Kingdom
simon.kirby@ed.ac.uk*

Language maps signals onto meanings through two types of structure. Firstly, the space of meanings is structured into shared categories. Secondly, the signals employed by a language are structured such that the meaning of the whole is a function of the meanings of its parts and the way in which those parts are combined. Previous work has demonstrated that structured categories (e.g., Xu, Dowman, & Griffiths, 2013) or structured signals (e.g., Kirby, Cornish, & Smith, 2008) can arise through iterated learning. However, the simultaneous emergence of these two types of structure has not been shown experimentally, leading to concerns that one type of emergent structure is simply an artefact of the other.

To explore this issue, we conducted a series of iterated learning experiments using a vast, open-ended, continuous meaning space. The first participant in a transmission chain was trained on 48 randomly generated signals paired with 48 triangles generated by selecting three coordinates at random in a 480×480-pixel box, yielding a space of 6×10^{15} possible stimuli. In a test phase, the participant was asked to label 96 novel triangles, none of which had been seen during training. The output from this test phase became the input to the next generation.

In the first experiment, we ran four chains, each with ten generations. Over time, the space of triangles was categorized into an increasingly small number of discrete regions, which consecutive participants increasingly align on. These

emergent categories, labelled with holistic signals, typically discretized the space of triangles based on their shape and size, ignoring features such as location and orientation (cf., Landau, Smith, & Jones, 1988). There was a cumulative increase in structure, showing that category systems can arise through iterated learning.

Our second experiment used the same experimental design, except at each generation a pair of participants were trained on the language separately and then entered a communication phase in which they took turns to communicate about triangles to each other. One participant was presented with a triangle and was asked to label it for their partner; the communicative partner then had to select the correct target triangle from a selection of six. The output from this communication phase became the input to the next generation in the chain.

The languages in this experiment, where there is a natural pressure for expressivity (Kirby, Tamariz, Cornish, & Smith, 2015), contained more unique signals. Despite these higher levels of expressivity, there was also a cumulative increase in structure in the languages. Furthermore, we found evidence that two of the four chains contained sublexical structure in addition to the categorical structure we observed in the non-communicative experiment. This sublexical structure was driven by shape-based sound symbolism and had morphological features similar to those found in natural languages (such as cranberry morphs; Aronoff, 1976).

Whereas previous iterated learning experiments have been limited to two types of result — categorical structure in meanings or compositional structure in signals — these experiments demonstrate that an alternative is possible. When the space of meanings is open-ended, and lacks clear pre-existing boundaries, then more subtle morphological structure, lacking straightforward compositionality, may evolve as a solution to the joint pressures from learning and communication.

References

- Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the USA*, *105*, 10681–10686. doi: 10.1073/pnas.0707835105
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102. doi: 10.1016/j.cognition.2015.03.016
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321. doi: 10.1016/0885-2014(88)90014-7
- Xu, J., Dowman, M., & Griffiths, T. L. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 1–8. doi: 10.1098/rspb.2012.3073

LASTING IMPACTS OF THE CODE MODEL ON PRIMATE COMMUNICATION RESEARCH

ERICA A. CARTMILL

Department of Anthropology, UCLA, 341 Haines Hall,

Los Angeles, CA 90095-1553, USA

cartmill@anthro.ucla.edu

The ability to mark a particular behavior as a communicative act—instead of relying on a small set of phylogenetically-shaped signals—expands the potential of a communicative system exponentially. Essentially any behavior can be made communicative through ostension. When paired with the ability to infer meaning from novel contexts and behaviors, this generates a powerful communicative engine. Human language is arguably built upon just such an ostensive-inferential engine (Sperber & Wilson, 1995; Origg and Sperber 2000; Scott-Phillips, 2014). The ability to take an action or sound and imbue it with meaning through “performing” it as a signal is undoubtedly an integral part of modern human language. But is it *uniquely* human?

Recently, Scott-Phillips extended the discussion of the O-I system by systematically contrasting the communication systems of great apes and humans with respect to their properties as codes (Scott-Phillips, 2014, 2015). He argues that ape communication is a “natural code,” relying on associative mechanisms and expanded by metapsychological abilities. Human language, by contrast, is a “conventional code,” built upon metapsychological abilities (the O-I system) and made more powerful by associative mechanisms. This contrast between natural and conventional codes makes important predictions about the communicative behavior of apes and humans; in particular, it predicts that only humans possess and recognize *communicative intentions* (an act that provides information that it is communicative, i.e., signaling its own signalhood). This capacity, in turn, lies at the heart of the O-I system. The evidence that ape communication is a natural code (and not based on communicative intentions) comes from the published literature on ape communication. This is reasonable, but could lead to biased conclusions about the nature of primate communicative

systems, because the code model itself has had a dramatic impact on the types of studies that are conducted (and published) in the field of primate communication.

Studies of primate communication search for and highlight predictable forms and contingencies that might be interpretable as codes. Ape gestural communication is less predictable and more flexible than the communication systems of many other animals. While the presence of communicative flexibility is used as evidence of intentionality in the ape gesture literature (Call & Tomasello, 2007), ambiguities in form or use are often discarded or overlooked. In fact, they are largely unanalyzable under this framework, which relies on predictable associations between identifiable forms and functions. Apes use their gestures flexibly, modifying them in response to their communicative partners: they direct their signals towards others, account for their partner's gaze, and wait for a response after gesturing. Yet the majority of published papers focus on predictability of signal to response—because authors are searching for codes.

I will review common data analysis procedures in ape gesture research, like excluding gestures that are not observed frequently enough to analyze (e.g., Cartmill & Byrne, 2010). I then discuss how these procedures influence the perception that gestures are natural communicative codes. I will present video examples of gestures that are typically deemed “unanalyzable.” Many unanalyzable gestures are simply rare or ambiguous gestures that do not show a simple one-form-to-one-meaning mapping and are thus not compelling in a research framework build around the search for codes. Ironically, these examples have the greatest potential to demonstrate ostensive communication in great apes. The theory that human communication is built on a framework of ostension and inference is compelling, but to determine whether humans are unique in these abilities we must assess the lasting impact of the code model framework on studies of primate communication. Primatologists should tackle this challenge head on. Emerging meta-analytic tools may facilitate this analysis by pooling rare events across studies and detecting complex regularities. These approaches would make significant advances in our understanding of the relationship between primate communication and human language.

References

- Call J, Tomasello M (2007) The gestural communication of apes and monkeys. Lawrence Erlbaum Associates, Inc, Mahwah.
- Cartmill, E., and Byrne, R. (2010). Semantics of primate gestures: intentional meanings of orangutan gestures. *Animal cognition*, 13(6), 793-804.

- Origi, G., and D. Sperber. (2000). Evolution, communication and the proper function of language. In *Evolution and the human mind: language, modularity and social cognition*. P. Carruthers and A. Chamberlain, eds. Cambridge: Cambridge University Press. Pp. 140–169.
- Scott-Phillips, T. (2014). *Speaking Our Minds: Why human communication is different, and how language evolved to make it special*. Palgrave MacMillan.
- Scott-Phillips, T. (2015). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, 56(1), 56-80.
- Sperber, D., and D. Wilson. (1995). *Relevance: communication and cognition*. 2nd edition. Oxford: Blackwell.

ARE EMOTIONAL DISPLAYS AN EVOLUTIONARY PRECURSOR TO COMPOSITIONALITY IN LANGUAGE?

FEDERICA CAVICCHIO, LIVNAT LEEMOR,
SIMONE SHAMAY-TSOORY, WENDY SANDLER
Sign Language Linguistics Lab, University of Haifa
Haifa, Israel
Federica.cavicchio@gmail.com, livnatlib@gmail.com,
sshamay@psy.haifa.ac.il, wsandler@research.haifa.ac.il

Compositionality is a basic property of language, spoken and signed, according to which the meaning of a complex structure is determined by the meanings of its constituents and the way they combine (e.g., Jackendoff, 2011 for spoken language; Sandler 2012 for constituents conveyed by face and body signals in sign language; Kirby & Smith, 2012 for emergence of compositionality). Here we seek the foundations of this property in a more basic, and presumably prior, form of communication: the spontaneous expression of emotion. To this end, we ask whether features of facial expressions and body postures are combined and recombined to convey different complex meanings in extreme displays of emotions. There is evidence that facial expressions are processed in a compositional fashion (Chen & Chen, 2010). In addition, facial components such as nose wrinkles or eye opening elicit systematic confusion while decoding facial expressions of disgust and anger and fear and surprise, respectively (Jack et al., 2014), suggesting that other co-occurring signals contribute to their interpretation. In spontaneous emotional displays of athletes, the body – and not the face – better predicts participants' correct assessments of victory and loss pictures, as conveying positive or negative emotions (Aviezer et al., 2012), suggesting at least that face and body make different contributions to interpretations of the displays. Taken together, such studies lead to the hypothesis that emotional displays are compositional - that each signal component, or possibly specific clusters of components (Du et al., 2014), may have their own interpretations, and make a contribution to the complex meaning of the whole. On the assumption that emotional displays are older than language in evolution, our research program aims to determine whether the crucial property of compositionality is indeed present in communicative displays of emotion.

In this study, we aim at finding specific features typical of spontaneous

responses of athletes to victory or defeat. We suggest that these features contribute to the interpretation of the complex emotions experienced in these contexts (Scherer & Ellgring, 2007). In our study, 350 pictures of athletes photographed within moments of winning or losing. We coded facial expressions using FACS (Ekman, Friesen & Hagar, 2002) and body displays using BACS (Cavichio & Marom, 2015). Only the visible face and body features were annotated. As a result, 305 features of facial expression and body posture were coded. We ran a Multiple Component Analysis to identify the features that were more frequently associated with win and loss. We found that a group of face and body features were prototypical of win contexts (e.g., Figure 1a: lowered brows, arms away from body), and others of loss (e.g., 1b: hands covering face/head, torso bent forward). A subset of these prototypical features was shared between the two contexts (e.g., 1c, d: eyes closed). We wish to determine whether the presence of the same signals shared across the two contexts contributes the same 'meaning' to each display – e.g., eye closure to avoid gaze and internalize an event that is too intense, whether joyful or painful. If so, this would suggest that emotional signals are understood compositionally rather than holistically, calling into question the notion of hardwired recognition of discrete emotions. To pursue this hypothesis, we tested participants' interpretation of these displays.

In the first experiment, participants were asked to rate the displays according to whether they portrayed someone who just lost or won on a 0 (clear loss) to 7 (clear win) scale. Participants saw three types of pictures taken from our corpus of 350 pictures: pictures in which facial and body features were frequently associated with win (prototypical win, Fig. 1a), pictures in which displays were frequently associated with loss (prototypical loss, Fig 1b) and pictures in which the displays consist of a mixture of win and loss features. In the mixed group, half of the pictures portrayed athletes who had just won and half athletes who had just lost (mixed win, Fig. 1c, and mixed loss, Fig. 1d). As a result, participants consistently rated as loss (mean=2.5) or win (mean=6) the pictures with the prototypical win and loss body features, but rated the mixed pictures in the middle of the scale (mean=4.5). The differences between conditions (win vs. loss and a mixed vs. prototypical) were all significant.

In the second experiment, we showed participants the same pictures and asked them to rate along a continuum how submissive/dominant, ashamed/proud, angry/not angry, disappointed/not disappointed, not satisfied/satisfied, not frustrated/frustrated or sad/happy the athletes looked. Again, prototypical losing picture scores were at the bottom of each scale and prototypical winning pictures scores were at the top. Mixed win and mixed loss pictures were rated in the middle of each scale. The differences between conditions (win vs. loss and mixed vs. prototypical) were all significant.

From these first results we conclude that different combinations of body and face features can modulate emotion recognition and judgments of win and

loss, laying the groundwork for a theory of compositionality in the expression and interpretation of emotions. Our findings are compatible with a compositional model of communicative emotional displays, and lend credence to the proposal that they are a plausible communicative precursor to language.



Figure 1: (a) Prototypical Win; (b) Prototypical Loss; (c) Mixed Win: eyes closed; (d) Mixed Loss: eyes closed.

Acknowledgements

This research is supported by ERC Grant Agreement 340140 for the Grammar of the Body project. The authors thank Noa Marom for her help with the collection and coding of the pictures.

References

- Aviezer, H., Trope, Y., & Todorov, A., (2012). Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science*, 338 (6111), 1225.
- Chen, M. Y., Chen, C. C., (2010). The contribution of the upper and lower face in happy and sad facial expression classification. *Vision Research*, 50, 1814-1823.
- Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 111(15), 1454-1462.
- Ekman, P., Friesen, W. V., & Hager, J. C., (2002). *Facial Action Coding System: The Manual on CD ROM*. Salt Lake City: A Human Face.
- Jack, R., Garrod, O. G. B., Schyns, P. G., (2014). Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time. *Current Biology*, 24, 1-6.
- Jackendoff, R. (2011). What is the human language faculty? Two views, *Language* 87(3), 586-624.
- Kirby, S. & Smith, K. (2012). Compositionality and linguistic evolution. *Oxford Handbook of Compositionality*. (pp. 493-509). Oxford: Oxford University Press.
- Sandler, W. (2012). Dedicated gestures and the emergence of sign language. *Gesture* 12(3), 265-307.
- Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion* 7(1), 158-17.

FUNCTIONALLY FLEXIBLE VOCALIZATIONS IN WILD BONOBOS

ZANNA CLAY^{1,2}

¹*School of Psychology, University of Birmingham, Birmingham, United Kingdom;* ²*Institute of Biology, University of Neuchatel, Neuchatel, Switzerland*
z.clay@bham.ac.uk

JAHMAIRA ARCHBOLD², KLAUS ZUBERBUHLER^{2,3}

²*Institute of Biology, University of Neuchatel, Neuchatel, Switzerland;* ³*School of Psychology and Neuroscience, St Andrews, UK*
jahmaira.archbold@unine.ch, klaus.zuberbuehler@unine.ch

1. Introduction

A core component underlying the evolution of language and the development of speech in human infants is the emergence of functional flexibility, the capacity to produce vocalizations independent of a fixed function in order to express a full range of emotional content across different situations (Griebel & Oller, 2008). Research has demonstrated that, even before speech emerges in infancy, 3-to-4 month old human infants produce a class of vocalizations—known as ‘protophones’— in functionally flexible ways to express a full range of emotional content (Oller et al., 2013). This finding has been contrasted with an apparent absence of evidence for this capacity in non-human primates, leading to the conclusion that the functional flexibility of human infant vocalizations marks an evolutionary transition between primate vocal communication and human speech (Oller et al., 2013). Here, we addressed this proposed evolutionary discontinuity by examining evidence for functional flexibility in the vocalizations of wild bonobos (*Pan paniscus*), one of our closest living relatives. We focussed on the ‘peep’, a commonly-produced vocalization

specific to bonobos. The ‘peep’ is a closed mouth vocalization that is high in frequency, short in duration and flat in acoustic form.

2. Methods & Results

We conducted behavioral observations and recorded vocalizations of wild adult bonobos at Lui Kotale in DR Congo using focal animal sampling. We analysed the acoustic structure of peeps produced in different behavioural contexts relating to the three principal valence dimensions (positive-neutral-negative) to explore acoustic cues relating to the inferred affective valence. We used Discriminant Function Analyses to examine if peep structure varied across valence contexts and caller identity.

Acoustic analyses revealed that wild bonobos produce a specific call type—the ‘peep’— across the full valence range in every major aspect of their daily lives, including feeding, travel, rest, aggression, alarm, nesting and grooming. Despite differences in the eliciting contexts, peep acoustic structure did not vary between contexts associated with neutral and positive valence. However, peeps produced in negative valence contexts were acoustically distinct, suggesting that vocal flexibility is more constrained by vocal production mechanisms in negatively charged situations. Peeps could be distinguished based on caller identity alone.

3. Discussion

In contrast to earlier conclusions (Oller et al., 2013), our results indicate that functionally flexible vocal signaling is a capacity shared with our closest living ape relatives, demonstrating its deep evolutionary roots. The finding of greater flexibility present in some contexts but not others suggests an evolutionary transition in hominids from functionally fixed to functionally flexible vocalizations. Identifying non-human primate vocalizations that are used in flexible ways, rather than being tied to fixed biological function, can provide relevant insights for the evolution of human speech. We will discuss these results in light of on-going analyses examining the pragmatic responses of receivers to peeps when combined in sequences with other calls.

Acknowledgments

We thank Gottfried Hohmann for supporting our research. We are grateful to the Institut Congolaise pour la Conservation de la Nature (ICCN) for granting permission to conduct research at the Salonga National Park

(MIN.0242/ICCN/DG/GMA/013/2013). Our methods comply with the requirements and guidelines of the ICCN and legal requirements of the DR Congo. We are grateful to Isaac Schamberg for his support, to Barbara Fruth and to all local staff supporting Lui Kotale. We thank members of the Department of Comparative Cognition at the University of Neuchatel, especially Christof Neumann. This research was financially supported by the L.S.B. Leakey Foundation, the National Geographic Society: Committee for Research and Exploration Grant, the British Academy Small Research Grant, the European Union Seventh Framework Programme for research, technological development, and demonstration under grant agreement 283871 and private donors associated with the British Academy and the Leakey Foundation.

References

- Griebel, U., & Oller, D. (2008). Evolutionary forces favoring communicative flexibility. In D. K. Oller and U. Griebel (Eds.), *Evolution of communicative flexibility: complexity, creativity, and adaptability in human and animal communication*, (pp. 9–40). Cambridge: MIT Press.
- Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6318–6323.

**RELATIONSHIP BETWEEN NONVERBAL SOCIAL SKILLS
AND LANGUAGE DEVELOPMENT: COMMUNICATIVE
GESTURES AND ATTENTIONAL SKILLS**

HÉLÈNE COCHET

*Psychology, University Toulouse Jean Jaurès
Toulouse, France
helene.cochet@univ-tlse2.fr*

RICHARD W. BYRNE

*School of Psychology and Neuroscience, University of St Andrews
St Andrews, United-Kingdom
rwb@st-andrews.ac.uk*

Studying the development of communication in infants and children provides information that may help better understand the evolution of human language. Here we focus on some infants' communicative skills such as joint attention and imitation, which have been argued to have a central role in language acquisition, as they offer an effective way to engage in social interactions. Gestural communication may have a particularly important role to play in allowing children to express meanings they are not yet able to verbalize (e.g., Colomesi, Stams, Koster, & Noom, 2010; Rowe & Goldin-Meadow, 2009). The association between gestures and language has also been highlighted in adults through the description of common neural networks in left-lateralized inferior frontal and posterior temporal regions of the brain (e.g., He et al., 2015; Xu, Gannon, Emmorey, Smith, & Braun, 2009).

However, the relationship between non-verbal communication abilities and language development is far from being fully understood, in part due to the difficulty of considering several characteristics of communication all at once, including various functions of children's gestures and both language comprehension and production. Our general objective in the present study was therefore to describe the relationships between early socio-cognitive abilities and

language, and especially to determine whether some gestures might play a stronger role than others in language acquisition. We used semi-structured situations based on the early social communication scale (ESCS: Guidetti & Tourrette, 2009; Mundy et al., 2003) to assess children's communicative profile between 11 and 41 months of age. We described the role played by children in the interactions (i.e., initiation vs. response) and analyzed more specifically the development of gestural communication by recording the number of gestures produced by children, including giving, showing, pointing and symbolic gestures. We also assessed the form (hand shape) and the function of pointing.

Positive correlations were found between language level and scores of joint attention, imitation and gestural communication. In particular, language was strongly correlated to gaze alternation between the communicative partner and the object referred to ($r = .72^{**}$ for language comprehension; $r = .78^{**}$ for language production). Moreover, our results suggest that the role played by children in the interactions influences language development: initiation of joint attention was correlated with both language comprehension ($r = .79^{**}$) and language production ($r = .82^{***}$), whereas response to joint attention was not correlated with either language comprehension or production. This study has also shown that the use of declarative expressive pointing (produced to share some interest with the partner about a specific referent), symbolic gestures and head nods (expressing agreement) was significantly related to language production ($r = .57^*$, $r = .59^*$, and $r = .72^{**}$, respectively), while the use of declarative informative pointing (produced to give helpful information to the partner) was strongly correlated with both language comprehension ($r = .84^{***}$) and language production ($r = .77^{**}$). The production of other gestures such as imperative pointing was not correlated with either language comprehension or production.

We believe that this work can have important implications for researchers studying the development and the evolution of communication, as it may help understand the dynamic processes involved in the emergence of language. Exploring the function of gestures and the role of individuals during the interactions may shed some light on the pressures that might have affected the evolution of communication (e.g., Grosse, Call, Carpenter, & Tomasello, 2015). Finally, this research could also improve language evaluation and intervention programs, especially for children who present communication disorders.

References

- Colonnese, C., Stams, G., Koster, I., & Nool, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review, 30*(4), 352-366.
- Grosse, K., Call, J., Carpenter, M., & Tomasello, M. (2015). Differences in the ability of apes and children to instruct others using gestures. *Language Learning and Development, 11*, 310-330.
- Guidetti, M., & Tourrette, C. (2009). *Echelle d'Evaluation de la Communication Sociale Précoce [Early Social Communication Scale]*. Paris: Eurotests (First edition 1993).
- He, Y., Gebhardt, H., Steines, M., Sammer, G., Kircher, T., Nagels, A., Straube, B. (2015). The EEG and fMRI signatures of neural integration: An investigation of meaningful gestures and corresponding speech. *Neuropsychologia, 72*, 27-42.
- Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). *A manual for the abridged Early Social Communication Scales (ESCS)*. Available through the University of Miami Psychology Department, Coral Gables, Florida (pmundy@miami.edu).
- Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science, 12*, 182-187.
- Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences, 106*(49), 20664-20669.

DWARF MONGOOSES COMBINE MEANINGFUL ALARM CALLS

KATIE COLLIER

*Department of Evolutionary Biology and Environmental Studies, University of Zurich
Zurich, Switzerland
katie.collier@ieu.uzh.ch*

ANDREW N. RADFORD

*School of Biological Sciences, University of Bristol
Bristol, United Kingdom
andy.radford@bristol.ac.uk*

BALTHASAR BICKEL

*Department of Comparative Linguistics, University of Zurich
Zurich, Switzerland
balthasar.bickel@uzh.ch*

MARTA B. MANSER

*Department of Evolutionary Biology and Environmental Studies, University of Zurich
Zurich, Switzerland
marta.manser@ieu.uzh.ch*

SIMON W. TOWNSEND

*Department of Psychology, University of Warwick
Coventry, United Kingdom
simon.w.townsend@warwick.ac.uk*

Syntax, the combination of meaningful words into larger meaningful structures, is a key feature of language that is responsible for much of language's generative power. Comparative data from animal communication studies can help unpack the evolution of syntax, which in turn is a necessary step towards better understanding the evolution of language as a whole. While syntax is present in all human languages, it is rare in animal

communication, though examples of call combinations that can be described as rudimentary syntax exist (Collier et al., 2014). These syntax-like combinations can be compositional, where the meaning of the combination is derived from the meaning of the component calls, as seen in Campbell monkey (*Cercopithecus campbelli campbelli*) alarm calls (Ouattara et al., 2009) or idiomatic or combinatorial, where the meaning of the combination is not related to the meaning of the component calls, as for putty-nosed monkey (*Cercopithecus nictitans*) alarm calls (Arnold & Zuberbühler, 2006).

Whilst comparative data outside of primates remains scarce, it can provide insights into the ecological or social factors that may be important in promoting the emergence of syntax. In this study we investigated alarm call combinations in dwarf mongooses (*Helogale parvula*), small, social living, cooperative breeding mammals. Similarly to other terrestrial mammals, recordings of natural predator encounters and experimental predator presentations suggest that dwarf mongooses produce, among others, one type of alarm call to aerial predators and another structurally distinct variant to terrestrial predators. Interestingly, dwarf mongooses also seem to combine these aerial and terrestrial calls into a third combination alarm, consisting of an aerial alarm followed immediately by a terrestrial alarm. Contextual data suggests these combination alarms are more often produced after the group had already been alerted to the presence of an aerial predator by an aerial alarm and so the function of the combination does not seem to be directly related to the independent functions of both the comprising aerial and terrestrial calls.

In order to verify if the combination alarm really is composed of independent aerial and terrestrial alarms, we first used acoustic analysis to compare whether there were structural differences between the alarm calls occurring alone and those comprising the combination. Secondly, we implemented playback experiments of synthetically constructed combination alarms (aerial + terrestrial alarm) and assessed receiver responses in relation to playbacks of naturally produced combination alarms.

Acoustic analyses revealed that the aerial component of the combination was not structurally different from the independently occurring aerial alarm, whereas the terrestrial component of the combination, on average, differed from the terrestrial alarm. However, the mongooses demonstrated similar behavioural reactions when hearing playbacks of both natural and synthetic combination alarm stimuli.

Thus, dwarf mongooses combine two meaningful alarm calls into a third alarm call whose meaning is not, *a priori*, a function of the meanings of its component calls, making this a potential example of a combinatorial call combination in a non-primate species. Interestingly, we did find acoustic variation between the terrestrial call within the combination alarm and the independent terrestrial alarm. Given these acoustic differences, yet the similar behavioural responses to artificial and natural versions of the combination alarm, it is possible that these subtle structural variations may not be perceived by the mongooses or relevant for their communication. The second half of the combination alarm and the terrestrial alarm call could therefore represent “allomorphs”: two acoustically distinct variants of the same call type that are perceived identically by the mongooses but used in different contexts.

In conclusion, dwarf mongooses combine meaningful alarm calls suggesting that concatenation of semantic units may be more widespread in animal communication than previously thought. Given the relatively large phylogenetic distance between dwarf mongooses and humans, these data can begin to unpack candidate selective pressures driving the emergence of a syntactical combinatorial level.

References

- Arnold, K., & Zuberbühler, K. (2006). Language evolution: Semantic combinations in primate calls. *Nature*, *441*(7091), 303–303.
- Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B., & Townsend, S. W. (2014). Language evolution: syntax before phonology? *Proceedings of the Royal Society B: Biological Sciences*, *281*(1788).
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026–22031.

WORD ORDER UNIVERSALS REFLECT COGNITIVE BIASES: EVIDENCE FROM SILENT GESTURE

JENNIFER CULBERTSON, MARIEKE SCHOUWSTRA, SIMON KIRBY

Centre for Language Evolution

*School of Philosophy, Psychology and Language Sciences, University of Edinburgh
Edinburgh, United Kingdom*

{jennifer.culbertson,marieke.schouwstra}@ed.ac.uk, simon@ling.ed.ac.uk

Research in linguistic typology has identified many cases in which particular patterns appear to be over- or under-represented in the world's languages. The extent to which these so-called typological universals reflect universal properties of human cognition remains heavily debated. In this paper, we provide empirical evidence connecting universals of word order to cognitive biases using a silent gesture experiment. The silent gesture paradigm allows us to capture spontaneous, untrained responses in a modality distinct from participants' previous language experience (Goldin-Meadow, So, Özyürek, & Mylander, 2008).

Our starting hypothesis comes from Greenberg's Universal 20 (Greenberg, 1963), concerning how adjective, numeral, demonstrative and noun are ordered. The distribution of possible orders of these elements is heavily skewed and it has been proposed that the common orders are those which match up transparently with the underlying semantics (Abels & Neeleman, 2012; Culbertson & Adger, 2014). The semantic relations between these elements are shown in Figure 1, and orders which are isomorphic to them can be read off directly without swapping the position of any of the modifiers. For example, N-Adj-Num-Dem and Dem-Num-N-Adj are isomorphic, Num-Dem-N-Adj and Adj-Num-Dem-N are not.



Figure 1. Schematic of semantic relations among elements in the noun phrase. Adj is conceptually closest to N, Dem most distant (Rijkhoff 2004).

Artificial language learning experiments show that English-speaking participants trained only on phrases with single post-nominal modifiers infer relative modifier orders that are isomorphic (e.g., N-Adj-Num-Dem; Culbertson & Adger, 2014). However, this study does not provide unambiguous evidence for a universal isomorphism bias, since English-speakers may have learned it from their L1

(Dem-Num-Adj-N is isomorphic). We use the silent gesture paradigm to test this bias in a modality distinct from participants' previous experience.

Stimuli consisted of simple pictures with a set (4 or 5; Num) of triangles or squares (N), all either spotted or striped (Adj). On each trial a stimulus appeared in one of two positions—proximal or distal to the participant (Dem). Participants were told to use gesture to communicate all the relevant information to another person. There is no training, therefore this is a test of naturalness of ordering.

Participants (N=20; native-English speakers, no experience with sign languages) each provided 32 gestures. Each gesture was coded and then scored according to whether it used post-nominal modifiers, and whether the relative order of the modifiers provided was isomorphic. The former measure, shown in Figure 2A, reveals that in general participants did *not* show a tendency to use English order—most dramatically, almost all gestures provided used a post-nominal adjective. This accords with a typological tendency that English violates (Greenberg, 1963). Critically, gestures also showed a very strong tendency toward isomorphic orders (Figure 2B) among all modifier pairs—whether pre- or post-nominal—supporting the hypothesis that Universal 20 reflects a cognitive bias favoring isomorphism.

This finding provides general support for a connection between statistical language universals and cognitive biases. We will discuss the extent to which our participants gestures reflect purely conceptual or linguistic factors, and the methodological implications of this for the silent gesture paradigm.

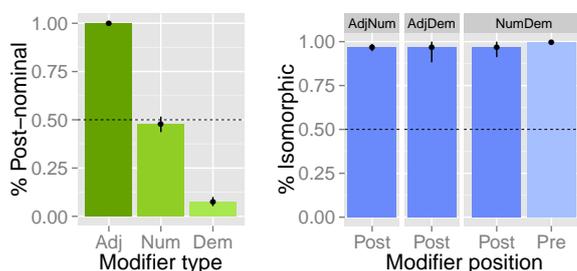


Figure 2. (A) Proportion of responses using post-nominal ordering for each modifier type. (B) Proportion of responses that are isomorphic, for each modifier pair, pre- and post-N.

References

- Abels, K., & Neeleman, A. (2012). Linear asymmetries and the LCA. *Syntax*, 15(1), 25–74.
- Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *PNAS*, 111(16), 5842–5847.
- Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The nat-

- ural order of events: How speakers of different languages represent events nonverbally. *PNAS*, 105(27), 9163–9168.
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (p. 73-113). Cambridge, MA: MIT Press.
- Rijkhoff, J. (2004). *The noun phrase*. Oxford: Oxford University Press.

THE EMERGENCE OF RULES AND EXCEPTIONS IN A POPULATION OF INTERACTING AGENTS

CHRISTINE CUSKLEY^{1,3}, VITTORIO LORETO^{1,2}

¹*Social Dynamics Unit, Institute for Scientific Interchange
Turin, Italy*

²*Dipartimento di Fisica, University of Rome, La Sapienza
Rome, Italy, vittorio.loreto@roma1.infn.it*

³*Centre for Language Evolution, University of Edinburgh
Edinburgh, UK, ccuskley@gmail.com*

Recent studies in language evolution have identified important roles for frequency (Cuskley et al., 2014), phonology (Bybee, 2001), and speaker population (Lupyan & Dale, 2010) in the dynamics of linguistic regularity. We present a model which integrates frequency, phonology, and speaker demographics to investigate how and why regularity and irregularity persist together given the general bias to eliminate unpredictable variation (i.e., irregularity), especially in experimental contexts (e.g., Hudson Kam & Newport, 2005; Smith & Wonnacott, 2010, among others). Kirby (2001) points out that while many models aim to represent how regular structure emerges in language, very few models explain how *irregularity* emerges. Using the iterated learning framework, Kirby (2001) showed that a skewed frequency of meanings and a general pressure for least effort in production can lead to the emergence of both stable regulars and irregulars in a vocabulary.

The current work aims to extend this finding by investigating the role of non-native speakers and phonological similarity in regularity dynamics. A recent study showed that non-native speakers irregularize novel forms more than native speakers. For example, non-native speakers are more likely than native speakers to apply ‘rules’ inferred from existing irregulars with a high token frequency (i.e., to provide the past tense of *spling* as *splung*, as an analogy with *spring* Cuskley et al., 2015). A potential mechanism underlying this result is that native and non-native speakers extend rules in different ways, depending on how rules are represented in their input. In other words, since native speakers have more experience with the ‘long-tail’ of regular verb types (Cuskley et al., 2014), they are more likely to extend the ‘regular’ rule. On the other hand, non-natives’ input is skewed towards irregular types with high token frequency, and thus they are more likely extend

quasiregularity when inflecting novel forms, especially when novel forms exhibit phonological similarity with existing irregulars (Cuskley et al., 2015).

We model the dynamics of regularity in a language evolving among a population of agents engaging in repeated communicative interactions (modelled after the Naming Game, hereafter NG; Loreto & Steels, 2007). The model broadly consists of repeated speaker (S) hearer (H) interactions. Unlike the NG, agents do not evolve labels for meanings, but inflections for forms: instead of naming meanings, the task of the S within the communicative interaction is to inflect an existing form, and success of the interaction is evaluated depending on whether the H shares the same inflection for the same form (see also Colaiori et al., 2015).

Agents begin with no inflections, but have an inventory of shared meanings labelled by strings randomly generated from a set of 10 characters. Meanings are chosen for each interaction based on a skewed, pre-determined frequency distribution. In early interactions, speaker agents choose a random two character string as an inflection; thus, at the outset, success is low, but agents nonetheless store inflections with weighted success (number of interactions/number of successes). Once agents acquire some inflections in their vocabulary as a result of interaction, they choose inflections for uninflected meanings in their vocabulary based on different “native” and “non-native” strategies. Both agent types have a first preference for extending inflections based on phonological similarity above a certain threshold: in other words, if the label for meaning A has a highly weighted inflection and a edit distance ≤ 0.5 away from the label for meaning B, they will generalise the inflection for meaning A to meaning B. Where this strategy fails, natives extend inflections based on type frequency (i.e., apply the inflection used across most items in the vocabulary), while non-natives extend inflections based on token frequency (i.e., apply the inflection from the most frequent item in the vocabulary).

Populations arrive at stable inflectional paradigms which include both regular and irregular forms. By altering the proportion of type and token preference agents in different iterations of the model, we are able to examine how these different strategies affect the structure of language over long timescales, and how changing proportions of token and type extension agents changes languages over time. Results from this framework support recent theories that the relative proportion of native and non-native speakers in a population has the potential to affect the structure of language.

Acknowledgements

This work is supported by the British Academy and by the Kreyon Project, funded by the John Templeton Foundation under contract n.51663.

References

- Bybee, J. (2001). *Phonology and language use*. Cambridge University Press: Cambridge.
- Colaïori, F., Castellano, C., Cuskley, C. F., Loreto, V., Pugliese, M., & Tria, F. (2015). General three-state model with biased population replacement: Analytical solution and application to language dynamics. *Phys. Rev. E*, *91*, 012808.
- Cuskley, C., Colaïori, F., Castellano, C., Loreto, V., Pugliese, M., & Tria, F. (2015). The adoption of linguistic rules in native and non-native speakers: Evidence from a wug task. *Journal of Memory and Language*, *84*, 205-223.
- Cuskley, C. F., Pugliese, M., Castellano, C., Colaïori, F., Loreto, V., & Tria, F. (2014). Internal and external dynamics in language: Evidence from verb regularity in a historical corpus of english. *PLoS ONE*, *9*(8), e102882.
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151-195.
- 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.
- Loreto, V., & Steels, L. (2007). Social dynamics: emergence of language. *Nature Physics*, *3*(11), 758-760.
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, *5*(1), e8559.
- Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444-449.

THE EMERGENCE AND EVOLUTION OF COLLABORATIVE STORIES

CHRISTINE CUSKLEY^{1,3}, BERNARDO MONECHI¹, PIETRO GRAVINO²,
VITTORIO LORETO^{1,2}

¹*Social Dynamics Unit, Institute for Scientific Interchange
Turin, Italy, mone.berna@gmail.com*

²*Dipartimento di Fisica, University of Rome, La Sapienza
Rome, Italy, vittorio.loreto@roma1.infn.it, pietro.gravino@gmail.com*

³*Centre for Language Evolution, University of Edinburgh
Edinburgh, UK, ccuskley@gmail.com*

Studies in literature and narrative have begun to argue more forcefully for considering human evolution as central to understanding stories and storytelling more generally (Sugiyama, 2001; Hernadi, 2002). However, empirical studies in language evolution have focused primarily on language structure or the language faculty, leaving the evolution of stories largely unexplored (although see Von Heiseler, 2014). Stories are unique products of human culture enabled principally by human language. Given this, the dynamics of creativity in stories, and the traits which make successful stories, are of crucial interest to understanding the evolution of language in the context of human evolution more broadly.

The current work aims to illuminate how stories emerge, evolve, and change in the context of a collaborative cultural effort. We present results from a novel experimental paradigm centered around a story game where players write short continuations (between 60 and 120 characters) of existing stories. These continuations then become open to other players to continue in turn. Stories are subject to player selection, allowing for variation and ‘speciation’ of the resulting narratives, and evolve as a result of collaborative effort between players.

The game starts with a seed of over 60 potential stories, and players choose which stories to continue, providing a player-driven story selection mechanism. In this way, stories which are creative, intriguing, and open ended spawn more stories, and eventually lead to longer story paths as play continues. The game also introduces further limitations by constraining a players’ view of the story path: players have access only to a story and its parent, meaning knowledge of the existing narrative is limited. We present data from hundreds of players and

stories, creating large ‘story trees’ which explore the space of different possible narratives which grow out of a confined set of starting points.

This data allows us to investigate several aspects of the growing story trees to illuminate not only what makes a story successful, but how creative stories trigger new stories, and what makes individual storytellers successful. Given the selection mechanism central to game play, we identify the most successful stories by their number of offspring. Particularly successful stories emerge measured both by how many children their stories have spawned, and also how long their story path extends. Coherent stories often emerge, despite the fact that they are authored by several different players, and any given player only sees a limited snapshot of the story path.

We contextualise the results of the game and connect it to language evolution using quantitative and qualitative analysis. We look for detectable triggers of innovation and creativity within the story trees, and identify these as expanding the ‘adjacent possible’ (e.g., new adaptations open the space of other possible adaptations in the future; Tria, Loreto, Servedio, & Strogatz, 2014). We argue that this concept can be extended to stories, using evidence from the game bolstered by evidence from more traditional literature (the Gutenberg Corpus). We frame a more qualitative analysis of the results in terms of recurring themes found in storytelling cross-culturally (Tehrani, 2013). We suggest that the most successful triggers of innovation in stories combine original novelty and a firm grounding in existing recurring story frameworks in human culture. This indicates that much like other cultural and biological systems, stories are subject to competing pressures for stability and conservation on the one hand, and innovation and novelty on the other.

Acknowledgements

This work was supported by the Kreyon Project, funded by the John Templeton Foundation under contract n.51663.

References

- Hernadi, P. (2002). Why is literature: A coevolutionary perspective on imaginative worldmaking. *Poetics Today*, 23(1), 21-42.
- Sugiyama, M. S. (2001). Narrative theory and function: Why evolution matters. *Philosophy and Literature*, 25(2), 233-250.
- Tehrani, J. J. (2013). The phylogeny of little red riding hood. *PLoS ONE*, 8(11), e78871.
- Tria, F., Loreto, V., Servedio, V. D., & Strogatz, S. H. (2014). The dynamics of correlated novelties. *Scientific Reports*, 5890.
- Von Heiseler, T. N. (2014). Language evolved for storytelling in a super-fast evolution. In E. Cartmill, S. Roberts, H. Lyn, & H. Cornish (Eds.), (p. 114-121). Singapore: World Scientific.

EMPIRICALLY ASSESSING LINGUISTIC ABILITY WITH STONE TOOLS

CORY M CUTHBERTSON

*Centre for the Archaeology of Human Origins (CAHO), University of Southampton,
Southampton, United Kingdom
c.m.cuthbertson@soton.ac.uk*

A recent stone knapping experiment is presented which supports that language does indeed fossilize, in the sense that it leaves impressions in stone. It links a) modes of cultural transmission in stone tool manufacture to b) required theory of mind ability to c) correlated linguistic ability. This includes specific syntactic and semantic features such as word reference, mental-state verbs, and complementation.

Discriminating between modes of cultural transmission (CTs) in the archaeological record is important because it identifies cognitive abilities which must be present to operate. Theory of mind (ToM) is the ability to attribute mental states such as knowledge or beliefs. It is the ability to think about thoughts (ToM-1) or the ability to think about thoughts about thoughts (ToM-2), etc. Emulation, imitation and teaching are three CTs that require ToM to different degrees: Emulation (ToM-0) is when an individual copies the end-state or 'goal' that a conspecific has produced, without theorizing about the conspecific's behaviour in production (Huang & Charman, 2005). Imitation (ToM-1) is process-oriented, and copies a conspecific's behaviour in order to produce a goal (Tomasello, 1996). It requires theorizing a relationship between the intended actions and goal. Teaching (ToM-2) is the intentional conferring of knowledge to a conspecific. The author argues that it requires theorizing that another is capable of thinking about the teacher's thoughts.

The relevance for detecting language in stone tools is that ToM and language are intimately linked, both functionally and developmentally, to the point that they are predictive of the other's ability (Miller, 2006); some suggest a likely coevolution (Malle, 2002). Language by its very nature is grounded in ToM, as "we cannot make sense of communicative interactions without presupposing that the interlocutors possess mutual knowledge of relevant beliefs and intentions." (Baldwin & Moses, 1994). ToM makes word reference possible, since it requires a shared knowledge of a linguistic symbol to be anchored to a referent. Identifying ToM-1 in the archaeological record is therefore argued to denote the ability for word reference. Developmentally, humans do not acquire grammatical language without the scaffolding provided by ToM. Mental-state verbs (such as 'know', 'think', 'remember'), as well as the complement structure to organize

mental-state verbs, is key in ToM-2's maturation (de Villers, 2007); therefore, it is argued that ToM-2 in the archaeological record denotes the presence of complementation and lexical items to describe mental states.

In this way, ToM can be used as a proxy for language ability in the archaeological record. An experiment has therefore been constructed to assess the different morphological characteristics of stone tool manufacture transmitted by emulation, imitation, and teaching. Twenty novice knappers replicated handaxes in one of three aforementioned CTs. The 160 resulting handaxes were then 2D scanned and their outlines traced, then uniformly rotated and scaled for shape comparison. The coefficient of variation (CV) was then calculated from the XY data. The CV shows that the shapes were less similar to each other in the emulation group (higher copy error), while the most standardization occurred in the teaching group (lower copy error). This suggests that CTs influence the level of variability in the shape of stone tools. This information can be used to determine what level of standardization can be reached only by higher fidelity CTs, and the ToM it implies.

To summarize, if language ability and ToM ability are correlated, and ToM enables different CTs, and CTs impact rates of copy error in material culture, we can measure stone tool assemblage variability to deduce language ability amongst Palaeolithic stone tool makers. In this way, through its impact on morphological variability, language can in a sense be said to fossilize.

References

- Baldwin, D. A., & Moses, L. J. (1994). Early understanding of referential intent and attentional focus: Evidence from language and emotion. In: C. Lewis and P. Mitchell (Eds.) *Children's early understanding of mind: Origins and development*. (pp. 133-156). Hove: Lawrence Erlbaum Associates.
- de Villers, J. (2007). The interface of language and theory of mind. *Lingua*, 117, 1858-1878.
- Huang, C.-T. & Charman, T. (2005). Gradations of emulation learning in infants' imitation of actions on objects. *Journal of Experimental Child Psychology* 92, 276-302.
- Malle, B. F., (2002). The relation between language and theory of mind in development and evolution. In T. Givón and B. F. Malle (Eds.), *The evolution of language out of pre-language* (pp. 265-284). Amsterdam: Benjamins.
- Miller, C. A. (2006). Developmental relationships between language and theory of mind. *American Journal of Speech Language Pathology*, 15, 142-154.
- Tomasello, M. (1996). Do apes ape? In B. Galef, C. Heyes (Eds.) *Social learning in animals: The roots of culture* (pp. 319–346) New York: Academic Press.

ANATOMICAL BIASING OF CLICK LEARNING AND PRODUCTION: AN MRI AND 3D PALATE IMAGING STUDY

DAN DEDIU, SCOTT R. MOISIK
*Max Planck Institute for Psycholinguistics,
Nijmegen, the Netherlands
dan.dediu@mpi.nl, scott.moisik@mpi.nl*

1. Investigating anatomical biases on speech sound systems

1.1. *Learning to produce clicks: Does vocal tract morphology matter?*

Clicks are among the most rare speech sounds, and their articulation depends upon a rarefaction gesture formed most typically between the tongue and the palate. It has been suggested (e.g., Traunmüller 2003) that click production might be subject to an effort bias associated with palate shape. Some have postulated (e.g. Allot 1994) that such anatomical biases on speech production could contribute to sources of variation driving phonological diversity.

The current paper presents results for data on click learning obtained from a larger imaging study (using MRI and 3D intraoral scanning) designed to quantify and characterize intra- and inter-population variation of vocal tract structures and the relation of this to speech production. The aim of the click study was to ascertain whether and to what extent vocal tract morphology influences (1) the ability to learn to produce clicks and (2) the productions of those that successfully learn to produce these sounds.

1.2. *Why examine clicks?*

Speakers of Khoisan-type languages (which are possibly the origin of click sounds, other non-Khoisan languages having borrowed them from Khoisan languages; see Traunmüller 2003) tend to have smooth palates lacking a prominence along the alveolar process (i.e. an alveolar ridge), as demonstrated by several studies (for a review, see Moisik & Dediu 2015). The suggested bias is that a prominent alveolar ridge requires more distortion of the tongue to form the linguo-palatal seal required for click production. Moisik & Dediu (2015)

show using biomechanical simulations that larger alveolar ridges increase muscular effort and reduce volume change.

2. The imaging study of phonetic learning of click production

To investigate this claim of a palate-shape bias on click production we conducted a study in which 79 participants were taught using auditory and visual aids how to produce (post-)alveolar (IPA notation [!]) and dental (IPA [ɿ]) clicks, during which an audio recording was obtained. These participants were then scanned using both static and real-time MRI sequences and 3D intraoral scanning.

The results indicate that the presence of an alveolar ridge certainly does not prevent an individual from learning to produce click sounds (1). However, the subtle details of how clicks are produced may indeed be driven by palate shape (2), as illustrated in Fig. 1.



Figure 1. Three participants producing [!] (at the moment before release) in a static MRI sequence (10 second sustained articulation). From left to right, palate shape varies from steep to shallow; in correspondence with this, tongue blade placement is increasingly more anterior.

Thus we have some very preliminary evidence that individual articulatory strategies in click production are influenced by anatomical factors. Elucidating the influence of anatomy on phonetics will allow us to better understand the emergence and evolution of speech sound systems and to better infer the speech capabilities of extinct humans from fossil traces of their vocal tract structures.

References

- Allott, R. (1994). Motor theory of language: The diversity of languages. In J. Wind, A. Jonker, R. Allott, & L. Rolfe (Eds.), *Studies in Language Origins* (Vol. 3, pp. 125–160). Amsterdam: John Benjamins Publishing Company.
- Moisik, S. R., & Dediu, D. (2015). Anatomical biasing and clicks: Preliminary biomechanical modelling. In H. Little (Ed.), *The Evolution of Phonetic Capabilities: Causes, constraints, consequences* (pp. 8–13). ICPHS.
- Traunmüller, H. (2003). Clicks and the idea of a human protolanguage. *Reports from the Department of Phonetics*, Umeå University (PHONUM), 9, 1–4.

THE FIDELITY OF ITERATED VOCAL IMITATION

PIERCE EDMISTON, MARCUS PERLMAN, GARY LUPYAN

*Department of Psychology, University of Wisconsin-Madison
Madison, WI, USA*

pedmiston@wisc.edu, mperلمان@wisc.edu, lupyana@wisc.edu

How do spoken words come into existence? Some words at least appear to originate from imitating environmental sounds (Dingemanse, 2012). Presumably, over time and repetition, these imitations become increasingly word-like, as they take on phonological and syntactic properties of their language, becoming less faithful to their original source (Perlman, Dale, & Lupyana, 2015). Yet, this process is not easily observed, and theories of precisely how it happens are largely speculative.

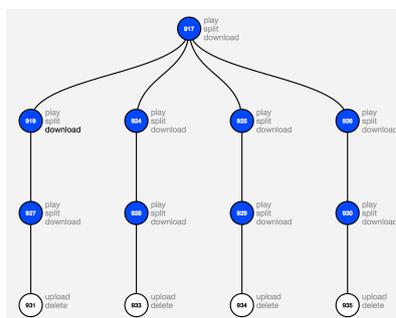


Fig 1. An ongoing vocal imitation game with four parallel chains leading from the same seed message. The recordings can be inspected as the data are collected.

To study spoken word formation, conducted a series of iterative vocal imitation experiments using a web-based application modeled after the game of “telephone.” The game begins with an experimenter setting initial (seed) sounds. Participants listen to each sound and produce an imitation of it that is recorded through their computer microphone. These recordings are passed on to the next participant who imitates the imitation, and so on. We then use match-

to-sample tasks to measure how accurately imitations can be matched to their source. Here we report preliminary findings about the degree of fidelity in vocal imitations over generations. For how many generations are participants able to maintain an imitation that can be traced back to its original source?

To answer this question we recorded people's imitations of several environmental sounds from different categories (a car horn, a telephone ring, a friction sound (ripping paper), and a water sound). New participants were then recruited and asked to guess the source of the imitation given the original (target) sound and several non-target sounds (the other seed sounds used in the experiment). Overall accuracy was relatively high, $M = 53\%$ (chance = 25%), $z = 5.4$, $p < 0.0001$, and surprisingly was not found to decrease significantly over the 11 generations of imitations, $z = -0.77$, $p = 0.44$.

We next increased the difficulty for both imitators and guessers by starting the chains with seed sounds from within the same category (e.g., four different water sounds). Accuracy was above chance after the first generation, $z = 2.2$, $p = 0.03$, but declined somewhat for subsequent generations, $z = -1.6$, $p = 0.09$.

In sum, we report a method of collecting high fidelity vocal imitations online, finding that vocal imitations can preserve correspondences with the original environmental stimulus for at least twelve generations. We are now examining: (1) vocal imitation over a larger range of stimuli under various conditions with the goal of quantifying how repeated imitation can drive vocalizations to become more word-like, as well as to understand the consequences of this transitional (2) the precise ways in which the signals are changing over time. Are the signals indeed becoming more word-like, and if so, what impact do such transitions from analogue imitation to discrete word form have on the ability of the listener to infer the communicative intentions of the speaker?

References

- Dingemans, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6, 654-672.
- Perlman, M., Dale, R., Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, 2: 150152.

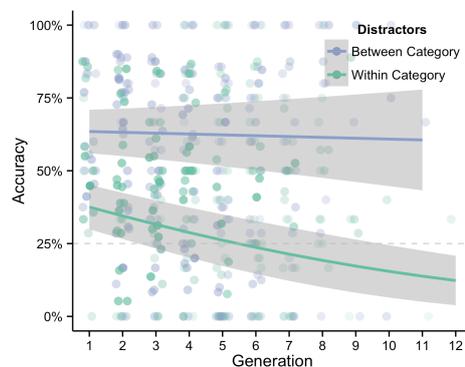


Fig 2. Accuracy in selecting the source of a vocal imitation among three distractors. Distractors were either from different categories or from the same category.

MEANINGFUL CALL COMBINATIONS AND COMPOSITIONAL PROCESSING IN A SOCIAL BIRD

SABRINA ENGESSER

*Department of Evolutionary Biology and Environmental Studies, University of Zurich
Zurich, Switzerland
sabrina.engesser@ieu.uzh.ch*

AMANDA R. RIDLEY

*Centre for Evolutionary Biology, The University of Western Australia
Crawley, Australia
amanda.ridley@uwa.edu.au*

SIMON W. TOWNSEND

*Department of Psychology, University of Warwick
Coventry, United Kingdom
simon.w.townsend@warwick.ac.uk*

A defining feature of language is its generative nature, but elucidating how this capacity evolved is a non-trivial task (Christiansen & Kirby, 2003). Language derives its expressive power from its combinatorial nature: meaningless acoustic elements are phonologically combined into meaningful words, which at a higher syntactic layer can be assembled into phrases, where the meaning of the whole is a product of its parts (Hockett, 1960). While recent work on birds has provided evidence for the phonological level (Engesser et al., 2015; Lachlan & Nowicki, 2015), evidence for basic compositional syntax outside of humans is less clear (Arnold & Zuberbühler, 2008; Hurford, 2011; Ouattara et al., 2009). In particular, experimental data demonstrating a compositional understanding of information are rare (Collier et al., 2014). Here we provide strong evidence for compositionality in the discrete vocal system of the cooperatively breeding pied babbler (*Turdoides bicolor*). Natural observations revealed pied babblers produce acoustically distinct alert-calls in response to close, low urgency threats, and recruitment-calls when recruiting group members during locomotion. Upon encountering terrestrial predators, both vocalizations are combined into a sequence (hereafter 'mobbing-sequence'), potentially to recruit group members in a dangerous situation. To investigate whether babblers process these mobbing-sequences in a compositional way, we conducted systematic playback manipulations, playing back the individual calls in isolation, as well as naturally occurring and artificial sequences. Our results show babbler groups reacted most strongly to mobbing-sequence playbacks, showing a greater attentiveness and a quicker approach to the sound source, compared to individual calls or

control sequences. We conclude pied babbler mobbing-sequences communicate information on both the context and the requested action, with receivers computing the combination of the two, functionally distinct, calls in a compositional way. Given the babblers' constrained vocal repertoire, paired with the extensive number of social and ecological contexts that require communication (Ridley & Raihani, 2007), such compositional production and processing of vocalizations is likely adaptive for pied babblers, allowing them to coordinate key additional events than would otherwise be possible with a non-syntactic system (Arnold & Zuberbühler, 2008). Ultimately, our work indicates that the ability to combine and process meaningful vocal structures, a basic syntax, may be more widespread than previously thought.

- Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*, R202-R203.
- Christiansen, M. H., & Kirby, S. (2003). Language evolution: consensus and controversies. *Trends in Cognitive Sciences*, *7*, 300-307.
- Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B., & Townsend, S. W. (2014). Language evolution: syntax before phonology? *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20140263.
- Engesser, S., Crane, J. M., Savage, J. L., Russell, A. F., & Townsend, S. W. (2015). Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. *PLoS Biology*, *13*, e1002171.
- Hockett, C. F. (1960). The Origin of Speech. *Scientific American*, *203*, 88-111.
- Hurford, J. (2011). *The origins of grammar*. Oxford: Oxford University Press.
- Lachlan, R. F., & Nowicki, S. (2015). Context-dependent categorical perception in a songbird. *Proceedings of the National Academy of Sciences*, *112*, 201410844.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE*, *4*, e7808.
- Ridley, A. R., & Raihani, N. J. (2007). Variable postfledging care in a cooperative bird: causes and consequences. *Behavioral Ecology*, *18*, 994-1000.

USING HMMS TO ATTRIBUTE STRUCTURE TO ARTIFICIAL LANGUAGES

KEREM ERYILMAZ, HANNAH LITTLE, BART DE BOER

*Artificial Intelligence Lab, Vrije Universiteit Brussel
Brussels, Belgium
kerem@ai.vub.ac.be*

Artificial languages have gained popularity as a way of empirically testing hypotheses on language evolution and language change (Scott-Phillips & Kirby, 2010), distancing the participants from the biases of their native languages as much as possible. Literature on artificial languages already contains studies on identifying the building blocks in repertoires of continuous signals. However, these rely on manual methods for identifying subparts of a signal that either use quantitative judgement or enforced boundaries between segments (Verhoef, Kirby, & Boer, 2013; Galantucci, 2005). Since these techniques build on the qualitative judgement of the researcher, they are time-consuming and difficult to replicate. We propose using computational models for this task.

The speech recognition literature uses positional HMMs' states as building blocks that stand in for different positions in a syllable (Rabiner & Juang, 1993), whereas the artificial languages literature tends to use segments of signals as building blocks (see Verhoef et al., 2013 for an example). As an alternative, we propose using computational models trained on an entire repertoire of signals in order to infer building blocks, and not imposing a pre-determined positional structure on the model so that the model structure is learned from the data. The formal process used guarantees reproducibility of the results, and creates an objective basis of comparison for signal repertoires created during artificial language experiments, both of which are improvements to the state of the art in artificial languages research to our knowledge.

We use HMM states as analogues for *areas* in the signal space. This approach is similar to the vowel quality literature which makes heavy use of the vowel triangle on the formant space, on which different areas correspond to different vowels. The extent of each area is defined by the covariance of the area's emission distributions. We use the number of states inferred as a measure of structure in the repertoire being modeled, analogous to phonetic inventory size.

The data is taken from an artificial language experiment, where participants create continuous acoustic signals using hand sensors to label a series of mean-

ings (Little, Eryilmaz, & Boer, 2015). The participants are tested on their recall of the repertoire of signals they created using a forced-choice task at the end of each phase. The signals can be either one- or two-dimensional, with dimensions corresponding to pitch and amplitude of the signal derived from the coordinates of the participant's hand. The meanings are also "dimensional" in that they differ along one or more feature dimensions such as size or colour. The features can be discrete (e.g. simple polygons of different number of vertices) or continuous (e.g. real numbers). The combination of these two spaces uniquely defines a "phase", and the collection of signals that belong to that combination form a "repertoire". The training data comprises a series of real numbers or tuples of real numbers, corresponding to the series of amplitude and frequency values of the tones played to the participants.

The emission distributions of the Gaussian HMMs used are either univariate or bivariate continuous distributions, depending on the dimensionality of the repertoire. Models are trained using the Baum-Welch algorithm, and the number of states are selected by training HMMs of various sizes and picking the one with the lowest Bayesian Information Criterion. The validation of the models is done using linear mixed-effects regression which uses the number of states in the HMM i.e. the "inventory size" of the repertoire, and its interaction with the properties of each phase to predict the recall of the participant at each phase.

The regression model outperformed the baseline, and explained more than half the variation in participant scores ($R^2 = 0.616$), demonstrating the validity of our model. It also matched our predictions that:

1. Continuous meaning spaces are more suitable for iconicity.
2. Discrete meaning spaces are more robust against variation.
3. Uninterrupted strategies outperform interrupted strategies.

While this demonstrates our model captures some aspects of the phenomenon, a third subset of the data encouraging participants to use "duration" as a signal dimension could not be validly modelled. This is most likely due to HMMs' inability to explicitly model time spent in a certain state. To remedy this, we plan to use more advanced latent models such as Explicit Duration HMMs that eliminate self-transitions, or Hidden Semi-Markov Models, which relax the Markovian assumption, in order to estimate duration distributions for the states.

Acknowledgements

Financial support from the ERC starting grant, ABACUS, project number 283435.

References

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, 29(5), 737–767.

- Little, H., Eryilmaz, K., & Boer, B. de. (2015). Linguistic modality affects the creation of structure and iconicity in signals. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. Jennings, & P. Maglio (Eds.), *The 37th annual meeting of the cognitive science society (cogsci 2015)* (p. 1392-1398). Austin, TX: Cognitive Science Society.
- Rabiner, L., & Juang, B.-H. (1993). Fundamentals of speech recognition.
- Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in cognitive sciences*, *14*(9), 411–417.
- Verhoef, T., Kirby, S., & Boer, B. de. (2013). Combinatorial structure and iconicity in artificial whistled languages. In *Proc. cogsci* (Vol. 35).

STICK OR SWITCH: A SIMPLE SELECTION HEURISTIC MAY DRIVE ADAPTIVE LANGUAGE EVOLUTION

NICOLAS FAY

*School of Psychology, University of Western Australia,
Perth, Western Australia
nicolas.fay@gmail.com*

SHANE ROGERS

*School of Psychology and Social Science, Edith Cowan University
Joondalup, Western Australia
shane.rogers@ecu.edu.au*

1. Introduction

If you describe shape (h) from Figure 1 as “the arrow”, but your addressee describes it as the “sleepwalker”, will this information change how you communicate the shape to your addressee? Will you stick with your original shape description, or switch to your addressee’s description? The experiment reported forces participants to choose between the two competing shape descriptions (personal or addressee), and uses participants’ ratings of description informativeness to predict their choice (stick or switch).

Classic theories, which emphasize the role of audience design to effective interpersonal communication, predict that people will adopt their addressee’s perspective. By contrast, minimalist theories suggest egocentric communication is common (for reviews see Brennan & Hanna, 2009; Shintel & Keysar, 2009). Tamariz et al (2014), modeling Fay et al’s (2010) empirical data, show that the spread of communication variants in a population can be explained via the interplay between an egocentric-bias and a content-bias. When people encounter a new sign-to-meaning mapping they tend to reuse the sign they had used before (egocentric-bias) unless the newly encountered sign is perceived to be superior (content-bias).

The present study empirically tests this simple selection heuristic. It also sheds light on the situational factors that cause people to take their addressee’s perspective or communicate egocentrically.

2. Method

Participants (N=112) typed descriptions for 18 abstract geometric shapes (see Figure 1). Next they were presented with their own description for each shape alongside the shape description produced by their addressee. Participants selected which shape description to return to their addressee (own, addressee). Finally, participants were again shown each pair of shape descriptions and indicated if they believed each description would allow a naïve person to identify the intended shape (Yes, No). This is our measure of informativeness.

3. Results & Discussion

Participants' communication behaviour was predicted by how informative they perceived the different shape descriptions to be. When their personal shape description was perceived to be more informative than their addressee's description, there was a strong bias to communicate egocentrically. By contrast, when their addressee's shape description was perceived to be more informative, there was a strong bias to take their addressee's perspective. When the shape descriptions were perceived to be equally informative, there was a moderate bias to communicate egocentrically. This simple selection heuristic may be critical to the adaptive evolution of human communication systems, and cumulative cultural evolution more generally.

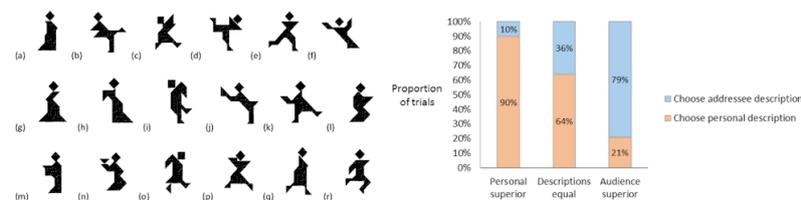


Figure 1. Stimuli used in the experiment (left) and the results (right).

References

- Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science, 1*(2), 274-291.
- Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science, 34*(3), 351-386.
- Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science, 1*(2), 260-273.
- Tamariz, M., Ellison, T. M., Barr, D. J., & Fay, N. (2014). Cultural selection drives the evolution of human communication systems. *Proceedings of the Royal Society B: Biological Sciences, 281*(1788).

PROCESSING PREFERENCES SHAPE LANGUAGE CHANGE

MARYIA FEDZECHKINA¹, BECKY CHU², T. FLORIAN JAEGER²,
JOHN TRUESWELL¹

¹*Department of Psychology, University of Pennsylvania, Philadelphia, USA*
mfedze@sas.upenn.edu, trueswel@psych.upenn.edu

²*Department of Brain and Cognitive Sciences, University of Rochester, Rochester, USA*
bchu@u.rochester.edu, fjaeger@bcs.rochester.edu

Why do languages share structural commonalities? One long-standing tradition has argued that linguistic universals reflect pressures on language use: languages have evolved to better suit the needs of human information processing and communication (Bates & MacWhinney, 1982). By what means these pressures come to shape language evolution, however, remains unknown. In a series of experiments, we explore the possibility that processing pressures operate during language acquisition, biasing learners to deviate from the input they receive, thus changing the input to the subsequent generation of learners and ultimately causing a shift towards a linguistic system that explicitly expresses these biases.

We modeled the situation of language change in the laboratory using a miniature language learning paradigm (Hudson Kam & Newport, 2005; Kirby et al., 2008). In all experiments, we exposed participants (adult monolingual native speakers of English) to miniature languages with several competing forms that expressed the same meaning. In training (administered over 3x1h-sessions on consecutive days), participants heard utterances in a novel language paired with videos depicting simple transitive actions performed by male actors. Participants first learned novel nouns, and then heard sentences using these nouns along with novel verbs. At the end of each session, learners described novel videos in the new language. We studied the deviations from the input in learners' productions.

Psycholinguistic research has linked dependency length minimization to the ease of processing (Grodner & Gibson, 2005) and cross-linguistic investigations have found evidence for this preference at the level of language structure (Futrell, Mahowald, & Gibson, 2015) suggesting a correlation between processing performance and grammar. Prior work, however, has not directly tested the *causality* underlying this correlation. In **Experiment 1**, we directly probe the causal link between processing biases in individual language learners and the preference for shorter dependencies observed cross-linguistically.

Different groups of learners were exposed to two miniature languages that were either head-initial (VSO/VOS word order) or head-final (SOV/OSV word order). All utterances were disambiguated through obligatory case-marking on objects (never subjects). In exposure sentences, subjects and objects were either both long (i.e., modified by a prepositional phrase in the head-initial language or by a postpositional phrase in head-final language, as cross-linguistically common) or both short (no modification). Balanced word order (SO/OS 50/50%) was maintained in all sentence types. During the production test, learners described novel videos, in which only one of the constituents (subject or object) was long (i.e., modified by an adpositional phrase) or both constituents were short. We find that despite receiving only unbiased (short-short, long-long) input, learners of the head-initial language followed the short-before-long ordering ($p < 0.05$), but learners of the head-final language showed the inverse long-before-short preference ($p < 0.001$). These length-based orderings minimize the average length of linguistic dependencies in the two languages. Our results thus suggest that learners are indeed biased towards shorter linguistic dependencies.

Experiment 2 tested the link between biases during learning and another cross-linguistic property linked to constraints on incremental processing – a preference to provide informative cues early. We explored whether learners are biased to provide informative cues early as it permits faster parsing decisions. The two miniature languages in Experiment 2 had SOV/OSV word order variation (50/50%) and optional case-marking (present 67%), but differed in its locus. In the subject-marking language, *subjects* but not objects were optionally case-marked independently of word order. In the object-marking language, *objects* were case-marked independently of word order (never subjects). Thus, the languages differed in the word order that allowed earliest disambiguation in case-marked sentences (SOV in the subject-marking and OSV in the object-marking language). We found that only learners of the object-marking language preferentially used case-marking at the earliest point of disambiguation in OSV sentences ($p < 0.001$). Learners of the subject-marking language marked both SOV and OSV orders equally often ($p > 0.7$) and significantly more frequently than the input on the final day of training ($p < 0.05$). We argue that this behavior is indicative of two preferences influencing language production – a bias to provide informative cues early and a bias to case-mark the less expected (i.e., non-English object-before-subject order) since the two pressures work in the same direction for the object-marking language and in opposite directions for the subject-marking language.

Our results suggest that biases in acquisition are reflected in typologically frequent patterns and can account for cross-linguistic structural similarities in natural languages. At least some of these biases stem from pressures of incremental processing: Even though our languages allowed several alternatives, learners consistently preferred structures that increased processing efficiency.

References

- Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language Acquisition: the State of the Art* (pp. 173-218). Cambridge: Cambridge University Press.
- Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proc Natl Acad Sci USA*, *112*(33), 10336-10341.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science*, *29*(2), 261–290.
- Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.
- Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Lang Learn Dev*, *1*(2), 151-195. doi: 10.1080/15475441.2005.9684215
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc Natl Acad Sci USA*, *105*(31), 10681.

COMMUNICATIVE INTERACTION LEADS TO THE ELIMINATION OF UNPREDICTABLE VARIATION

OLGA FEHÉR, KENNY SMITH

*Centre for Language Evolution, School of Philosophy, Psychology and Language
Sciences, The University of Edinburgh
Edinburgh, United Kingdom
olga.feher@ed.ac.uk, kenny.smith@ed.ac.uk*

NIKOLAUS RITT

*Department of English, Universität Wien
Vienna, Austria
nikolaus.ritt@univie.ac.at*

ELIZABETH WONNACOTT

*Division of Psychology and Language Sciences, University College London
London, United Kingdom
e.wonnacott@ucl.ac.uk*

Certain properties are shared among the world's languages and are assumed to reflect learners' cognitive and linguistic biases. One such feature is that linguistic variation tends not to be unpredictable; instead, the variation present in languages is usually conditioned on grammatical or social context. For example, English exhibits variation in how the plural is marked: it can take different forms besides the regular 's' such as the morpheme '-en' in 'oxen' or nothing as in 'sheep', but these forms are fully predicted by the lexical item they mark, the variation is not free or unpredictable. Studying how learners acquire and use artificial languages exhibiting unpredictable variation can inform us about the biases that shape languages over long time-scales, because certain linguistic "preferences" may show up very quickly during learning and recall. Previous research has shown that children, and adults under certain conditions such as increased task demands, often eliminate variation by regularizing on one of the synonymous forms (Hudson Kam and Newport, 2009). Variation can also be eliminated by processes of transmission as shown by Smith & Wonnacott (2010) using an iterated learning paradigm. Another mechanism by which individual-level processes may shape the properties of languages is *communicative interaction*. We present findings from two experiments showing that

communication leads to the rapid elimination of variation, as interlocutors converge on a shared grammatical system.

We trained pairs of participants on miniature artificial languages that exhibited variation in either the presence or absence of a grammatical marker (Exp. 1) or in word order (Exp. 2), and then allowed them to communicate using the language they learnt. We used semi-artificial languages (pseudo-English words for lexical items and nonce words as grammatical markers) and varied the probability with which the grammatical marker appeared on each training trial in Exp. 1 and the probability of either word order to describe simple visual events in Exp. 2. Participants were randomly placed in one of several conditions (differing only in the statistics of their input language). Regardless of how large the difference was in their initial input language, communicating pairs in both experiments quickly converged on a shared linguistic system, jointly regularizing on a common form (not necessarily reflecting the pattern of their native language, English). Experiment 1 showed that when variable users interact with categorical users (who had been trained on an consistent system), they accommodate their use to match their partners', whereas categorical users do not change their behaviour during interaction. In addition to convergence, Experiment 2 provided evidence for structural priming: participants were more likely to use their preferred word order when their partners used that same order in the trial immediately preceding. This largely automatic priming is a likely mechanism leading to behavioral convergence (Pickering & Garrod, 2004). Moreover, we found a lasting effect of interaction on recall in both experiments as revealed by a post-interaction individual test.

This study shows that moment-to-moment alignment between communicating partners plays an important role in driving convergence to a shared, predictable grammatical system. More generally, we show that communicative interaction leads to the elimination of unpredictable variation, which may be a third important mechanism (besides individual learning and cultural transmission) that may lead to language change and ultimately language design by amplifying learners' weak individual biases.

References

- Hudson Kam, C. L. & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59, 30-66.
- Pickering, M. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169-226
- Smith, K. & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116, 444-449.

WORD LEARNERS REGULARIZE SYNONYMS AND HOMONYMS SIMILARLY

VANESSA FERDINAND

*The Santa Fe Institute
Santa Fe, U.S.A
vanessa@santafe.edu*

MATTHEW SPIKE

*School of Philosophy, Psychology, and Language Sciences, University of Edinburgh
Edinburgh, U.K.
matthew.spike@ed.ac.uk*

Unpredictable variation is rare in language. Explanations for this include a language-specific *regularization bias* (e.g. Reali & Griffiths, 2009), general constraints on memory (e.g. Hudson Kam & Newport, 2005), or both (e.g. Perfors, 2012; Ferdinand, Thompson, Kirby, & Smith, 2013). Experiments on lexical regularization typically study how words in free *synonymous* relationships become increasingly deterministic through use (e.g. some words drop from use). There is also experimental evidence that learners regularize *homonymous* relationships (Vouloumanos, 2010), but to date no experimental design has directly compared the relative regularization of synonyms versus homonyms. This is an important comparison to make because synonyms and homonyms have asymmetrical functional roles in communication (Hurford, 2003) and the jury is still out as to which of these two regularization biases are better for evolving effective communication systems. On one hand, Hurford proposes that there is less bias against homonyms because they are more common in language than synonyms and Piantadosi, Tily, and Gibson (2012) argue for the communicative function of ambiguous lexicons. On the other hand, Doherty (2004) demonstrates children's difficulty in learning homonyms and Spike, Stadler, Kirby, and Smith (2013) show that self-organizing novel lexicons require a bias against homonymy but not synonymy.

We extend the experimental paradigm of Ferdinand et al. (2013) to investigate the relative regularization of synonyms versus homonyms. 128 participants were trained on one of two artificial mini-languages with identically matched distributions of variation. In the synonyms condition, this variation was over word forms and in the homonyms condition it was over referents. Regularization is quantified by the drop in Shannon entropy of the words and referents that participants

produced when tested on their mini-language. Participants regularized 67% of the variation among homonyms ($t(63) = -12.8169, p < .001$) and 56% of the variation among synonyms ($t(63) = -10.5526, p < .001$). However, there was no significant difference between these conditions ($t(126) = 1.3518, p = 0.18$), suggesting that learners compress synonymous and homonymous variation similarly.

This experiment was repeated with non-linguistic stimuli, where participants learned the mappings between marbles and the different containers they were drawn from. Participants also regularized the non-linguistic stimuli, eliminating 42% of the variation among containers ($t(63) = -7.277, p < .001$) and 32% of the variation among marbles ($t(63) = -6.6908, p < .001$), again with no significant difference between conditions ($t(126) = 1.5049, p = 0.13$). This suggests a domain-general component to linguistic regularization. However, participants in the linguistic conditions regularized significantly more than those in the non-linguistic conditions ($F(252) = 11.259, p < .001$). We conclude that regularization results from general-purpose compression during learning, which can be ramped up for effective communication with linguistic stimuli, and operates similarly on synonyms and homonyms.

References

- Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, 31(1), 203–214.
- Ferdinand, V., Thompson, B., Kirby, S., & Smith, K. (2013). Regularization behavior in a non-linguistic domain. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 436–441).
- Hudson Kam, C., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2), 151–195.
- Hurford, J. (2003). Why synonymy is rare: Fitness is in the speaker. In *Proceedings of ECAL 7* (pp. 442–451). Berlin: Springer Verlag.
- Perfors, A. (2012). When do memory limitations lead to regularization? An experimental and computational investigation. *Journal of Memory and Language*, 1–60.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280–291.
- Reali, F., & Griffiths, T. L. (2009, June). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3), 317–28.
- Spike, M., Stadler, K., Kirby, S., & Smith, K. (2013). Learning, feedback and information in self-organizing communication systems. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3442–7).
- Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, 107(2), 729–742.

HUMANS RECOGNIZE VOCAL EXPRESSIONS OF EMOTIONAL STATES UNIVERSALLY ACROSS SPECIES

PIERA FILIPPI^{1,2}, JENNA V. CONGDON³, JOHN HOANG³, DANIEL LIU
BOWLING⁴, STEPHAN REBER⁴, ANDRIUS PAŠUKONIS⁴, MARISA
HOESCHELE⁴, SEBASTIAN OCKLENBURG⁵, BART DE BOER¹, CHRISTOPHER
B. STURDY³, ALBERT NEWEN^{2,6}, ONUR GÜNTÜRKÜN^{2,5}

¹ Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Belgium; ²Center for Mind, Brain and Cognitive Evolution Email: pie.filippi@gmail.com; ³Department of Psychology, University of Alberta, Canada; ⁴Department of Cognitive Biology, University of Vienna; ⁵Department of Biopsychology, Ruhr University, Bochum, Germany; ⁶Department of Philosophy II, Ruhr University, Bochum, Germany;

The perception of danger in the environment can induce physiological responses (such as a heightened state of arousal) in animals, which may cause measurable changes in the prosodic modulation of the voice (Briefer, 2012). The ability to interpret the prosodic features of animal calls as an indicator of emotional arousal may have provided the first hominins with an adaptive advantage, enabling, for instance, the recognition of a threat in the surroundings. This ability might have paved the ability to process meaningful prosodic modulations in the emerging linguistic utterances.

Research has shown that humans are able to recognize different levels of arousal in mammalian calls. However, to our knowledge, no study has ever examined humans' cross-cultural ability to identify different arousal levels in calls of species belonging to several phylogenetically distant taxa, including, for instance, mammals and birds. We addressed this issue by developing a task in which human participants of three different cultures (Canadian, German, Mandarin) listened to ten pairs of vocalizations for each of nine different vertebrate taxa. We used amplitude-controlled calls from the following species: hourglass treefrog, American alligator, black-capped chickadee, common raven, domestic pig, giant panda, African elephant, Barbary macaque, and human.

Calls within each pair differed in arousal level, which was assessed based on the behavioral context of call production (Avey et al., 2011; Bowling et al., 2012; Fischer et al. 1995; Linhart et al., 2015; Reichert 2013; Stoeger et al., 2011, 2012). For each pair of vocalizations, participants were asked to identify the call with the higher level of arousal. Accuracy rate in identifying arousal in each species was higher than expected by chance in all three cultures. No significant differences were observed between cultures. This finding provides empirical support for Darwin's hypothesis on the universality of vocal emotional communication. In order to better understand the mechanisms underlying emotional intensity recognition in our set of calls, we investigated which acoustic parameters correlate with participants' correct responses. We performed this analysis in two steps. First, we identified two acoustic features measurable in all calls of our stimuli set: duration and a frequency related measure, the spectral center of gravity. Second, we calculated the duration and the spectral center of gravity ratio for each pair of calls and correlated these two feature comparisons with the percentage of correct responses across pairs. Our data indicate that the spectral center of gravity is the only feature found across our animal species calls, which significantly correlates with the ability to discriminate high arousal calls. Further work within this research paradigm will provide quantitative data on shared mechanisms involved in emotional vocalizations' production and perception across animal taxa, investigating the perception of arousal in nonhuman species. This may improve our understanding of the semantic value of prosody in animal communication, and of its role in the emergence of human language.

References

- Avey, M. T., Hoeschele, M., Moscicki, M. K., Bloomfield, L. L., & Sturdy, C. B. (2011). Neural correlates of threat perception: neural equivalence of conspecific and heterospecific mobbing calls is learned. *PloS one*, *6*, e23844.
- Bowling, D. L., Sundararajan, J., Han, S. E., & Purves, D. (2012). Expression of emotion in Eastern and Western music mirrors vocalization. *PLoS One*, *7*, e3194.
- Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology*, *288*, 1-20.
- Fischer, J., Hammerschmidt, K., & Todt, D. (1995). Factors affecting acoustic variation in Barbary-macaque (*Macaca sylvanus*) Disturbance Calls. *Ethology*, *101*, 51-66.
- Linhart, P., Ratcliffe, V. F., Reby, D., & Špinka, M. (2015). Expression of Emotional Arousal in Two Different Piglet Call Types. *PloS one*, *10*(8), e0135414.

- Reichert, M. S. (2013). Sources of variability in advertisement and aggressive calling in competitive interactions in the grey treefrog, *Hyla versicolor*. *Bioacoustics*, 22, 195-214.
- Stoeger, A. S., Charlton, B. D., Kratochvil, H., & Fitch, W. T. (2011). Vocal cues indicate level of arousal in infant African elephant roars. *The Journal of the Acoustical Society of America*, 130(3), 1700-1710.
- Stoeger, A. S., Baotic, A., Li, D., & Charlton, B. D. (2012). Acoustic features indicate arousal in infant giant panda vocalisations. *Ethology*, 118(9), 896-905.

Multimodal processing of emotional meanings: A hypothesis on the adaptive value of prosody

PIERA FILIPPI^{1,2}, SEBASTIAN OCKLENBURG³, DAN BOWLING⁴, LARISSA
HEEGE⁵, ALBERT NEWEN^{2,6}, ONUR GÜNTÜRKÜN^{2,3}, BART DE BOER¹

¹Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Belgium; ²Center for Mind, Brain and Cognitive Evolution; Email: pie.filippi@gmail.com; ³Department of Biopsychology, Ruhr University, Bochum, Germany, ⁴Department of Cognitive Biology, University of Vienna, Austria; ⁵Department of General and Biological Psychology, University of Wuppertal, Germany, ⁶Department of Philosophy II, Ruhr University, Bochum, Germany.

Humans combine multiple sources of information to comprehend meanings. These sources can be characterized as linguistic (i.e., lexical units and/or sentences) or paralinguistic (e.g. body posture, facial expression, voice intonation, pragmatic context). Emotion communication is a special case in which linguistic and paralinguistic dimensions can *simultaneously* denote the same, or multiple incongruous referential meanings. Think, for instance, about when someone says “I’m sad!”, but does so with happy intonation and a happy facial expression. Here, the communicative channels express very specific (although conflicting) emotional states as denotations. In such cases of intermodal incongruence, are we involuntarily biased to respond to information in one channel over the other? We hypothesize that humans are involuntary biased to respond to prosody over verbal content and facial expression, since the ability to communicate socially relevant information such as basic emotional states through prosodic modulation of the voice might have provided early hominins with an adaptive advantage that preceded the emergence of segmental speech (Darwin 1871; Mithen, 2005). To address this hypothesis, we examined the interaction between multiple communicative channels in recruiting attentional resources, within a Stroop interference task (i.e. a task in which different channels give conflicting information; Stroop, 1935). In experiment 1, we used synonyms of “happy” and “sad” spoken with happy and sad prosody. Participants were asked to identify the emotion expressed by the verbal content while ignoring prosody (Word task) or vice versa (Prosody task). Participants responded faster and more accurately in the Prosody task. Within the Word task, incongruent stimuli were responded to more slowly and less accurately than congruent stimuli. In experiment 2, we adopted synonyms of “happy” and “sad” spoken in happy and sad prosody, while a happy or sad face was displayed. Participants were asked to identify the emotion expressed by the verbal content

while ignoring prosody and face (Word task), to identify the emotion expressed by prosody while ignoring verbal content and face (Prosody task), or to identify the emotion expressed by the face while ignoring prosody and verbal content (Face task). Participants responded faster in the Face task and less accurately when the two non-focused channels were expressing an emotion that was incongruent with the focused one, as compared with the condition where all the channels were congruent. In addition, in the Word task, accuracy was lower when prosody was incongruent to verbal content and face, as compared with the condition where all the channels were congruent. Our data suggest that prosody interferes with emotion word processing, eliciting automatic responses even when conflicting with both verbal content and facial expressions at the same time. In contrast, although processed significantly faster than prosody and verbal content, faces alone are not sufficient to interfere in emotion processing within a three-dimensional Stroop task. Our findings align with the hypothesis that the ability to communicate emotions through prosodic modulation of the voice – which seems to be dominant over verbal content - is evolutionary older than the emergence of segmental articulation (Mithen, 2005; Fitch, 2010). This hypothesis fits with quantitative data suggesting that prosody has a vital role in the perception of well-formed words (Johnson & Jusczyk, 2001), in the ability to map sounds to referential meanings (Filippi et al., 2014), and in syntactic disambiguation (Soderstrom et al., 2003). This research could complement studies on iconic communication within visual and auditory domains, providing new insights for models of language evolution. Further work aimed at how emotional cues from different modalities are simultaneously integrated will improve our understanding of how humans interpret multimodal emotional meanings in real life interactions.

References

- Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.
- Filippi, P., Gingras, B., & Fitch, W. T. (2014). Pitch enhancement facilitates word learning across visual contexts. *Frontiers in Psychology*, 5, 1-8.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*, 44, 548–567.
- Mithen, S. J. (2005). *The singing Neanderthals: The origins of music, language, mind, and body*. London: Weidenfeld Nicolson.
- Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, 33, 127-146.

- Rigoulot, S., & Pell, M. D. (2012). Seeing Emotion with Your Ears: Emotional Prosody Implicitly Guides Visual Attention to Faces. *PLoS One*, 7(1), e30740.
- Soderstrom, M., Seidl, A., Nelson, D., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49, 249-267.
- Stroop, J. R. (1935). Studies of Interference in Serial Verbal Reactions, *Journal of Experimental Psychology*, 18, 643-662.

DO LAB ATTESTED LEARNING BAISES PREDICT THE STRUCTURE OF A NEW NATURAL LANGUAGE?

ANONYMOUS AUTHOR 1, ANONYMOUS AUTHOR 2, ANONYMOUS AUTHOR3

*University Department, University Name,
City, Country
email1@university, email2@university, email3@university*

Typological analysis clearly shows that the world's languages are not evenly distributed among all logically possible patterns. Recent studies (i.e. Culbertson, Smolensky, Legendre, 2012; Fedzechkina, Jaeger, Newport, 2012; Culbertson & Newport, 2015) on the emergence of language structure in the lab find that the most common typological patterns in languages around the world are generally the patterns adults prefer when learning an artificial language. Accordingly, the researchers conclude that these most common patterns are the product of learner biases (cognitive or communicative) toward certain types of structure. Here we explore this question in a new natural language: Nicaraguan Sign Language (NSL). We investigate whether signers of this new language will use the most typologically common orders for the elements of a noun phrase.

NSL, one of the youngest languages known to science, was born in the late 1970s with the founding of a new school for special education. The first students to enter the school were homesigners: isolated deaf individuals who develop their own gesture systems in order to communicate with the individuals around them. When these homesigners came together in the 1970s, the stage was set for the creation of a new language, and the first cohort of NSL was formed. Though instruction was in written and spoken Spanish, students soon began to communicate with one another manually. As succeeding cohorts of students learn NSL, the language itself is changing rapidly.

Following Culbertson et al. (2012) and Culbertson & Newport (2015), we examine the ordering of noun, adjective, and number elements within noun phrases in NSL. Culbertson and colleagues find that harmonic orders (in which the adjective and number are either both prenominal or both postnominal) were preferred over non-harmonic orders (in which the noun comes between the other two elements), consistent with the typological pattern reported by Dryer (2008). We showed participants a series of cards depicting a set of objects (e.g., dogs or cars); set size varied from 1 to 4, and objects were either large or small. We asked participants to describe the content of each card, and determined the

ordering of noun phrase elements produced by signers in three successive age cohorts of NSL: Cohort 1 (n=9) who came together in the 1970s and formed NSL; Cohort 2 (n=9) and Cohort 3 (n=6) who were exposed to NSL upon school entry between the early 1980s and early 2000s. NSL signers have been shown to build increasingly complex linguistic structure over successive cohorts (Senghas & Coppola, 2001). Data collection was carried out in 2009 and again in 2015. The 2009 data collection included 6 participants (3 from Cohort 1, and 3 from Cohort 2); data was collected from all 24 participants in 2015, including the original 6.

In 2009 (Figure 1), we found that Cohort 1 signers and Cohort 2 signers preferentially produced harmonic orders (either noun-adjective-number or noun-number-adjective), with no significant difference between cohorts ($\beta=-.90$, $p=.45$). This pattern is in keeping with Culbertson et al.'s predictions—individuals creating a new language prefer harmonic orders, potentially reflecting the same biases that have shaped the attested typological pattern.

Interestingly, in 2015 (Figure 2), the pattern we observed was the same for Cohort 1, but not for Cohorts 2 and 3. Cohort 1 signers still preferred harmonic orders. In fact, all three Cohort 1 signers tested at both time points preferred harmonic orders in both 2009 and 2015. However, Cohort 2 signers now more often used non-harmonic number-noun-adjective order, and for Cohort 3 signers this was the most preferred order. The preference for non-harmonic order increases significantly with later cohort ($\beta=-5.24$, $p<.02$). Strikingly, this means that individual signers in the second cohort moved away from the harmonic pattern. Of the 3 Cohort 2 signers tested in both 2009 and 2015, all preferred harmonic orders in 2009 but only 1 of 3 still showed a harmonic preference in 2015.

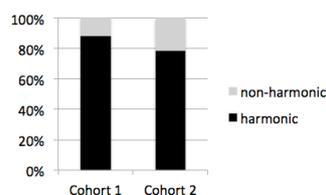


Figure 1. Noun Phrase Ordering in 2009.

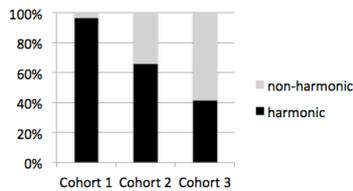


Figure 2. Noun Phrase Ordering in 2015.

We thus see harmonic ordering in the earliest stages of this new language, as the typological and experimental data would predict. The intriguing result is the relatively quick transition from a harmonic pattern to a non-harmonic pattern in Cohorts 2 and 3. Future work is needed to explore pressures leading NSL away from the typologically robust harmonic pattern (e.g., influences from Spanish, which has a non-harmonic pattern, that might be transmitted through co-speech gesture).

Acknowledgements Supported by x (xx), and x (xx).

**COOPERATIVE COMMUNICATION AND COMMUNICATION
STYLES IN BONOBOS AND CHIMPANZEES IN THE WILD:
SAME SAME BUT DIFFERENT?**

MARLEN FRÖHLICH*, PAUL H KUCHENBUCH, GUDRUN MÜLLER, SIMONE
PIKA*

*Humboldt Research Group, Max Planck Institute for Ornithology,
Seewiesen, Germany*

**Correspondence to: mfroehlich@orn.mpg.de, spika@orn.mpg.de*

BARBARA FRUTH

*Department Biology II, Ludwig Maximilian University,
Munich, Germany*

TAKESHI FURUICHI

*Primate Research Institute, Kyoto University,
Kyoto, Japan*

ROMAN M WITTIG

*Department of Primatology, Max Planck Institute for Evolutionary Anthropology,
Leipzig, Germany*

Human language is manifested by fast-paced and extensive social interactions, thereby representing an essentially cooperative endeavour. It has been repeatedly claimed that the cognitive skills related to participation in cooperative communication are unique to the human species (Levinson, 1995; Tomasello, 2008). One way to enable a better understanding of the factors and pressures triggering the evolution language is the comparative approach, which uses empirical evidence from living species to draw inferences about communicative abilities in our ape-like ancestors. Rossano (2013) recently provided evidence that the structure of communicative interactions between mother-infant dyads of captive bonobos is strikingly similar to the sequential structure of social action in human conversation. Using parameters established in human conversation analysis, he found that two dyads frequently established participation frameworks, engaged in cooperative adjacency-pair structures, and communicated at a pace that strongly resembled the timing of ordinary human conversation (Stivers et al., 2009). In the present study, we aimed to investigate and expand some of the parameters used by Rossano (2013) *in situ*, that is in mother-infant dyads of chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) living in their natural environments. Although previous behavioural comparisons of the two sister species revealed a remarkable dichotomy in crucial

aspects of their social matrix, a direct systematic comparison of their communicative skills is to date non-existent.

Since true differences in communicative abilities between two species can only be proposed if within-species variability is taken into account (Boesch, 2007), we compared communicative interactions of 25 mother-infant dyads in two different chimpanzee and two different bonobo communities: *Tai South* in Tai National Park, Côte d'Ivoire (*Pan t. verus*), *Kanyawara* in Kibale National Park, Uganda (*P. t. schweinfurthii*), *Wamba* in the Luo Scientific Reserve, DRC, and *LuiKotale* in Salonga National Park, DRC. We focused on the single communicative function of mother-infant joint travel, since previous studies suggested that this is a fruitful context enabling the observation of frequent communicative exchanges in mother-infant dyads about a distinct goal: leaving a location (Rossano, 2013). The following criteria of human communicative interactions were analysed: (i) formation of participation frameworks before signal production, by analysis of gaze, body orientation and initiation distance, (ii) adjacency pair-like sequences, by analysis of gestural pursuits and response waiting after each pursuit; and (iii) the timing between signal and response. We analysed a total of 415 chimpanzee and 316 bonobos joint travel interactions filmed during 2200 hours of observation. Overall, our results showed that both bonobo and chimpanzee mother-infant dyads showed the capacity and motivation to engage in cooperative communication. Moreover, the two species differed significantly in terms of all three investigated criteria. While gaze, close initiation distance and fast-paced responses were features of bonobo mother-infant interactions, chimpanzees performed a larger number of gestural pursuits, more response waiting and more 'delayed' responses. Notably, none of these findings could be explained by mere within-species variability.

Taken together, we provided compelling evidence that our two closest living relatives differ regarding temporal patterns and styles of their gestural communication. Bonobos seem to anticipate and respond to signals before they have even been entirely executed, while chimpanzees frequently engage in more prolonged communicative negotiations. Nevertheless, both *Pan* species use sequentially organised, cooperative social interactions to achieve a mutual goal: leaving together to another location. Communicative interactions of bonobos and chimpanzees thus reflect crucial features of human social action during conversation, implying that cooperative communication emerged as a means to efficiently coordinate collaborative activities. Our study thus corroborates the hypothesis that the cognitive prerequisites for human language as a collaborative enterprise must have evolved in the primate lineage long before speech arose in modern humans (Levinson, 2006; Seyfarth & Cheney, 2008). Hence, our findings add a crucial facet to the *Pan* dichotomy and, as such, aid in pinpointing some of the crucial factors influencing language evolution.

References

- Boesch, C. (2007). What makes us human (*homo sapiens*)? The challenge of cognitive cross-species comparison. *Journal of Comparative Psychology*, *121*(3), 227-240.
- Levinson, S. C. (1995). Interactional biases in human thinking. In E. N. Goody (Ed.), *Social intelligence and interaction* (pp. 221-260). Cambridge: Cambridge University Press.
- Levinson, S. C. (2006). On the human “interaction engine”. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39-69). Oxford: Berg.
- Rossano, F. (2013). Sequence organization and timing of bonobo mother-infant interactions. *Interaction Studies*, *14*(2), 160-189.
- Seyfarth, R. M., & Cheney, D. L. (2008). Primate social knowledge and the origins of language. *Mind & Society*, *7*(1), 129-142.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., & Heinemann, T. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, *106*(26), 10587-10592.
- Tomasello, M. (2008). *Origins of human communication* (Vol. 2008): MIT press.

INTEGRATION OR DISINTEGRATION?

KOJI FUJITA, HARUKA FUJITA

*Graduate School of Human and Environmental Studies,
Kyoto University, Kyoto, Japan
fujita.koji.3x@kyoto-u.ac.jp, muchtoo79@yahoo.co.jp*

An evolutionary novelty often results from combining preexisting old traits. The non-monolithic, modular architecture of human language, which consists at least of the Conceptual Intentional (CI) system for internalization, the Sensory Motor (SM) system for externalization, and the recursive combinatorial system (syntax) for connecting them (Hauser et al., 2002; see also Berwick & Chomsky, 2016) can be understood as a reflection of its evolution taking place in a similar integrating fashion.

The *Integration Hypothesis* (IH) of human language evolution put forward by Miyagawa and his colleagues (Miyagawa et al., 2014, 2015, Nóbrega & Miyagawa, 2015) is an interesting proposal which is in perfect harmony with the aforementioned general picture of evolution. The IH maintains that human language arose from a combination of two separate systems, each being found in nonhuman animals, the E(xpressive) system of birdsong and the L(exical) system of primate calls. It is also interesting because it immediately raises some important questions, among which are: (i) Do other animals really have E and L systems? (ii) If yes, are these nonhuman versions exactly the same as the human counterparts? (iii) How did these two systems get combined only in the human lineage? In this paper, we challenge the IH by carefully considering and answering these questions. In particular, our answers to the questions (i) and (ii) will be in the negative. As an alternative, we propose the *Disintegration Hypothesis* (DH): animal communication systems do not have a clear distinction between E and L systems, and they become separated only in human language.

N&M (2015) defend the proposed distinction of the two systems by linking them to category-defining functional categories and roots ($\sqrt{\quad}$) with no categorial specification (e.g. $n + \sqrt{\text{CAT}} \rightarrow \text{cat}$), respectively, as defined in the theory of Distributed Morphology (Marantz, 1997, et seq.). Exactly what kind of

information these roots contain (only concepts or also syntactic and phonological features) is a matter of hot dispute, but there is no evidence that animal L systems (to the extent that they exist) are equal to the human L system either qualitatively or quantitatively. Human concepts, for example, are abstract (with no direct reference to the external world) and hierarchically organized, built from atomic components and giving rise to a multitude of derived concepts. Also, animal communication systems do not have human-like lexical categories like nouns and verbs, which suggests at least that E system does not exist there. Animal communications do not distinguish between cognitive, instructional and affective elements. Honeybee dances express information about the found nectar, an instruction to collect it, and a concomitant emotional charge all at the same time, as if everything comes as an inseparable whole. The *separation of affect* (Bronowski, 1977) takes place in human language alone.

Given that questions (i) and (ii) cannot be answered positively, the IH now loses much of its force. Furthermore, Miyagawa et al. (2013) once proposed to derive the nonfinite hierarchical structure of human language from the combination of the two finite systems, but to make this combination possible N&M (2015) now resort to the unbounded Merge operation (an answer to question (iii)). An inconsistent redundancy is obvious; Merge alone should be enough, and the combination of E and L systems is just one example of this general operation, which is found only in human language. Our DH solves these problems by maintaining that the separation of these two systems takes place only in human language, as a reverse effect of Merge (Demerge), which is also an important ingredient of human language that seems to apply, for example, in the mapping from the computational system to the SM interface.

The IH also holds that direct combination of two L layers are impossible. Given that Merge is a free, unconstrained operation (Boeckx, 2015), this restriction should be removed. We examine compounding phenomena and show that in fact L layers can combine with each other (root compounding, such as n [$\sqrt{\text{BIRD}} \sqrt{\text{BRAIN}}$] \rightarrow *birdbrain*). The nature of exocentric compounding (such as *birdbrain* with a metonymic interpretation, together with its implication for language evolution (Jackendoff, 2009, Progovac, 2015) will also be examined. We suggest that a rudimentary form of exocentric compounding may be found in animal communication but is fundamentally different from its human counterpart.

We also argue that Merge is the only new function which is necessary for human language to evolve from preexisting capacities, and that Merge itself evolved from sequential and hierarchical object manipulation typically involved in human and animal tool use.

References

- Berwick, R. C., & Chomsky, N. (2016) *Why only us?: Language and evolution*. Cambridge, MA: MIT Press.
- Boeckx, C. (2015) *Elementary syntactic structures—prospects of a feature-free syntax*. Cambridge: Cambridge University Press.
- Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014) How could language have evolved? *PLoS Biology*, *12*, e1001934.
- Bronowski, J. (1977). *A sense of the future*. Cambridge, MA: MIT Press.
- Fujita, K. (2014) Recursive Merge and human language evolution. In T. Roeper & M. Speas (Eds.), *Recursion: Complexity in cognition* (pp. 243-264). New York: Springer.
- Fujita, K. (2016) On certain fallacies in evolutionary linguistics and how one can eliminate them. In K. Fujita & C. Boeckx (Eds.), *Advances in biolinguistics: The human language faculty and its biological basis* (pp. 141-152). New York: Routledge.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569-1578.
- Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., Chomsky, N., & Lewontin, R. C. (2014) The mystery of language evolution. *Frontiers in Psychology*, *5*, 401.
- Jackendoff, J. (2009) Compounding in the parallel architecture and conceptual semantics. In R. Lieber and P. Stekauer (Eds.), *The Oxford Handbook of Compounding* (pp. 105-128). Oxford: Oxford University Press.
- Marantz, A. (1997). No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics*, *4(2)*, 201-225.
- Miyagawa, S., Berwick, R., & Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology*, *4*, 71.
- Miyagawa, S., Ojima, S., Berwick, R., & Okanoya, K. (2014). The integration hypothesis of human language evolution and the nature of contemporary languages. *Frontiers in Psychology*, *5*, 564.
- Nóbrega, V. A., & Miyagawa, S. (2015). The precedence of syntax in the rapid emergence of human language in evolution as defined by the integration hypothesis. *Frontiers in Psychology*, *6*, 271.
- Progovac, L. (2015) *Evolutionary syntax*. Oxford: Oxford University Press.

EFFECTS OF TASK-SPECIFIC VARIABLES ON AUDITORY ARTIFICIAL GRAMMAR LEARNING AND GENERALIZATION

ANDREEA GEAMBAŞU*

*Leiden University Centre for Linguistics, Leiden University,
Leiden, The Netherlands
a.geambasu@hum.leidenuniv.nl*

MICHELLE J. SPIERINGS*, CAREL TEN CATE

*Institute for Biology, Leiden University
Leiden, The Netherlands
m.j.spierings.2@biology.leidenuniv.nl, c.j.ten.cate@biology.leidenuniv.nl*

CLARA C. LEVELT

*Leiden University Centre for Linguistics, Leiden University,
Leiden, The Netherlands
c.c.levelt@hum.leidenuniv.nl*

Extraction and generalization of rules from stimuli that share an underlying structure is one of the bedrocks of language acquisition. This rule learning ability has been shown in adults using both simple and complex grammars in the auditory, visual, and tactile domains. Understanding the conditions under which simple rule learning can occur and to what extent learning is implicit or explicit is essential for understanding what the fundamentals of language acquisition are and whether language acquisition may have evolved from simpler pattern-extraction mechanisms. Inconsistency in experimental methodology used to show rule learning indicates that it is of interest to explore the precise conditions influencing how well learners perform in such tasks. To this end, we conducted four auditory artificial grammar learning experiments with 12 conditions (n=192) using XYX and XXY grammars. In Experiments 1-3, ten participant groups received passive familiarization with one of the two grammars and were tested with a yes/no paradigm. In Experiment 4, two groups were exposed to one of the two grammars via reinforced training and were tested in a go-left/go-right task. Across these four experiments, we manipulated the following experimental

* Authors share first authorship.

factors: vagueness of instructions, input variety, presence or absence of feedback, and types of testing items.

In Experiment 1, instructions were “undirected,” not directing participants’ attention to the underlying structure. Participants were asked whether the test items are part of the same “language” or “group” as in the listening phase. To study the effect of variety, participants were further divided into groups exposed to either 3 or 15 triplets, for a total of 45 trials in both cases. Test items consisted of a consistent and an inconsistent grammar, each made up of either familiar or novel syllables, constituting “undirected” testing in which test items could not direct participants to what they should be attending to. In Experiment 2, instructions were “directed,” telling participants that the exposure sounds followed a certain “pattern,” and that they should indicate whether test items followed that same pattern. As in Experiment 1, participants either heard 3 or 15 triplets during the familiarization phase, and testing was “undirected,” consisting of both familiar and unfamiliar sounds. In Experiment 3, we again compared the role of instruction and of number of familiarization triplets, but now used “directed” testing, meaning only novel sounds were used (directing participants’ attention away from processing at the sound level). Finally, in Experiment 4, we again varied the number of exposure triplets and used undirected testing, but now exposed participants, without instruction, in a reinforced go-left/go-right task. When they reached criterion, they continued (now without feedback) with the same procedure of categorizing the test items as either a left-side sound or a right-side sound, where each grammar corresponded with one of the sides.

Our results show that participants were able to apply the rule to test items composed of previously heard sounds, independently of our experimental manipulations, discriminating the two grammars significantly above chance in all conditions. However, they were not able to generalize the rule to novel sounds if they were not somehow “directed,” either through directed instruction, directed testing, or feedback training. Notably, variety in number of exposure triplets during familiarization did not affect generalization, with no significant difference in performance between participants exposed to 3 or 15 triplets. It thus seems that in order to generalize simple rules beyond their surface form, participants require their attention to be directed, supporting recent findings in a dual-mechanism account of AGL (Opitz & Hofmann, 2015). These results have implications for the design of future AGL experiments and for theories of implicit vs. explicit AGL. A comprehensive understanding of language learning must integrate the evolution of a primary similarity-detection and an attention-based rule-detection mechanism.

References

- Opitz, B. and Hofmann, J. (2015). Concurrence of rule- and similarity-based mechanisms in artificial grammar learning. *Cognitive Psychology*, 77, 77-99.

INTENTIONAL MEANING OF BONOBO GESTURES

KIRSTY E. GRAHAM

School of Psychology and Neuroscience, University of St Andrews,
St Andrews, UK
keg4@st-andrews.ac.uk

CATHERINE HOBAITER, RICHARD W. BYRNE

School of Psychology and Neuroscience, University of St Andrews,
St Andrews, UK
clh42@st-andrews.ac.uk, rwb@st-andrews.ac.uk

Unlike linguists, animal communication researchers cannot ask their subjects what they mean. Most animal communication is non-intentional, and the function of a signal can be assessed by looking at the outcome (Seyfarth, Cheney, & Marler, 1980). However, great ape gestural communication is intentional. Great apes direct their gestures towards a specific recipient; check the attention of that recipient; wait for that recipient to respond; and, if the recipient does not respond, the signaller persists and elaborates (Call & Tomasello, 2007; Cartmill & Byrne, 2007; Leavens & Hopkins, 1998; Tomasello, George, Kruger, Farrar, & Evans, 1985). These behaviours show that the signaller begins with an intended goal and uses gestures in order to achieve that goal; the signal therefore has meaning, in the sense of Gricean first order intentional meaning (Grice, 1969). To determine a signal's intentional meaning, we cannot just look at the outcome, as with its biological function. Rather, we see which outcome satisfies the signaller, showing that the "apparently satisfactory outcome" (ASO) matched the signaller's original intended goal (Cartmill & Byrne, 2011; Hobaiter & Byrne, 2014). Under the natural conditions likely to elicit a full range of intended meanings, this method of defining the meaning of great ape gestures has only so far been used for wild chimpanzees (Hobaiter & Byrne, 2014). In our current research, we use the same method for gestural communication of wild bonobos. Bonobos are chimpanzees' closest living relatives, having diverged approximately 0.8-0.9 MYA (Becquet & Przeworski, 2007; Won & Hey, 2005). Despite genetic closeness, their social systems are remarkably different. Bonobo females form the centre of parties and high-ranking females outrank males; they engage in frequent genito-genital rubbing and other forms of non-conceptive copulation; and they encounter peacefully with neighbouring communities (Furuichi, 2011; Idani, 1990; Kano, 1980). We already know that the chimpanzee and bonobo gestural repertoires overlap significantly in gesture form (Graham, Hobaiter, & Byrne, 2015), but not whether these shared gestures also have the same meanings. The data of this paper come from 900 hours of focal individual data and 4381 video clips from focal behaviour filming, collected during two six-month field seasons at Wamba, DR Congo. In order to catalogue the bonobo repertoire and examine the meaning of their gestures, we extracted gestures that met criteria for intentionality, in particular those that allowed us to recognize ASOs of the signaller. This paper will concentrate on differences in gesture form and meaning between bonobos and chimpanzees, and relate these differences to the strikingly different social and sexual behaviour of the two species.

Acknowledgements

Funding from the Wenner-Gren Foundation and University of St Andrews.

References

- Becquet, C., & Przeworski, M. (2007). A new approach to estimate parameters of speciation models with application to apes. *Genome Research*, 17(10), 1505–19.
- Call, J., & Tomasello, M. (2007). *The Gestural Communication of Apes and Monkeys*. Lawrence Erlbaum Associates, Publishers.
- Cartmill, E. A. and Byrne, R. W. (2007). Orangutans modify their gestural signalling according to their audience's comprehension. *Current Biology*, 17, 1345-1348
- Cartmill, E. A. & Byrne, R. W. (2011). Addressing the problems of intentionality and granularity in non-human primate gesture. In S. Gale and M. Ishino (Eds.), *Integrating Gestures. The interdisciplinary nature of gesture*, pp.15-26. John Benjamins Publishing Company: Amsterdam
- Furuichi, T. (2011). Female contributions to the peaceful nature of bonobo society. *Evolutionary Anthropology*, 20(4), 131–42.
- Graham, K., Hobaiter, C., & Byrne, R. W. (2015). *Comparison of gestural communication for Pan species*. Paper presented at the meeting of the Association for the Study of Animal Behaviour, Lincoln, UK.
- Grice, H. P. (1969). Utterer's meaning and intention. *The Philosophical Review*, 78(2), 147–177.
- Hobaiter, C., & Byrne, R. W. (2014). The Meanings of Chimpanzee Gestures. *Current Biology*, 24, 1–5.
- Idani, G. (1990). Relations between unit-groups of bonobos at Wamba, Zaire: encounters and temporary fusions. *African Study Monographs*, 11(3), 153–186.
- Kano, T. (1980). Social behavior of wild pygmy chimpanzees (*Pan paniscus*) of Wamba: A preliminary report. *Journal of Human Evolution*, 9, 243– 260.
- Leavens, D. A., & Hopkins, W. D. (1998). Intentional communication by chimpanzees: a cross-sectional study of the use of referential gestures. *Developmental Psychology*, 34(5), 813–822.
- 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.
- 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.
- Won, Y.-J., & Hey, J. (2005). Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, 22(2), 297–307.

PLAIN SIMPLE COMPLEX STRUCTURES: THE EMERGENCE OF OVERSPECIFICATION IN AN ITERATED LEARNING SETUP

STEFAN HARTMANN¹, PEETER TINITS², JONAS NÖLLE³, THOMAS
HARTMANN⁴, MICHAEL PLEYER⁵

¹ German Department, University of Mainz, Jakob-Welder-Weg 18, 55128 Mainz, hartmast@uni-mainz.de; ² Dept. of Linguistics, Tallinn University; ³ Center for Semiotics, Aarhus University; ⁴ Dept. of Electrical Engineering, Karlsruhe Institute of Technology; ⁵ English Department, Universität Heidelberg

Natural languages differ in their degree of overspecification, the extent to which overt semantic markers are required even when irrelevant in the given context. More overspecified languages have been described as typologically more complex based on user effort (Kusters, 2003) and information theory (Dahl, 2004). But how and why does systematic and obligatory overspecification emerge in the first place? While such developments could in principle be seen as mere accumulation of arbitrariness, recent research has emphasized the importance of context in the emergence of different types of language systems (e.g. Winters et al., 2015). To test if subtle contextual pressures can enforce a bias towards overspecification which eventually influences language structure, we designed an Iterated Learning experiment (cf. Kirby et al., 2014) that allowed for variation in the degree of overspecification in the individual participant's output.

205 volunteers were recruited online and allocated into chains of 5 individuals. They were then trained on an artificial language and subsequently asked to point out objects to an alien. The output of generation n was used as input for generation $n+1$. The initial language for all chains consisted of 4 root words (e.g. *meeb* 'ball') and 2 color markers (*pu* 'blue', *li* 'yellow') used systematically and maximally efficiently: The color marker was used only if it was required to disambiguate the object in context. Across conditions, 16 of the 32 trials required disambiguation between two objects of the same type (e.g. a blue cup and a yellow cup, see Fig. 1, left panel, bottom row). In the Control condition, only one single object was displayed in the remaining 16 trials. In the Distractor condition, by contrast, the other 16 trials consisted of pictures showing two different

objects (e.g. cup and pen). Both color (yellow vs. blue) and position (left vs. right) were assigned randomly to the objects in these pictures, such that the color of target and distractor would match in half of the trials (as in Fig. 1) and differ in the remaining trials. As the two types of referential context are very similar in the Distractor condition, while they are quite distinct in the Control condition, we predicted a higher degree of overspecification for the former condition: In these trials, overspecification relieves the language user from the need to distinguish contexts in which the same type of object is displayed twice from contexts featuring two different types of object.

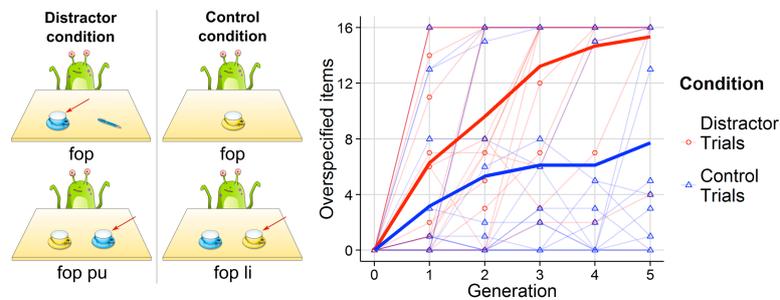


Fig. 1. Left: Experimental design. Right: Number of overspecified color markers (max. 16) for each chain across conditions. 5 chains were omitted due to noncompliance with the instructions (see supplementary materials).

Overspecification increased in both types of trials but proved more pervasive in the Distractor condition. Here, the color marker became fully obligatory in Generation 5 in 17 out of 18 chains, while it was used significantly less in Generation 5 of the control trials (two-sample $t(34)=-4.06$, $p_{\text{two-tailed}}<.001$, $r=.57$, Fig. 1).

These results indicate that contextual pressures can promote the evolution of obligatory semantic markers from adjective-like modifiers in an Iterated Learning setup. Wherever the use of color markers becomes mandatory across contexts, “color” becomes an obligatorily marked semantic distinction. Although the resulting language can be described as more complex on the typological criteria mentioned above, the use of overspecified semantic markers here is a result of strategies facilitating the task at hand in a specific context. This suggests that, given certain contextual pressures, complex structures can sometimes be the simpler solution.

Acknowledgments

The authors wish to thank James Winters for helpful comments on a previous version of the experiment reported on here. In addition, we are grateful to three anonymous reviewers for valuable comments and suggestions on the present paper.

References

- Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated Learning and the Evolution of Language. *Current Opinion in Neurobiology*, 28, 108–114.
- Kusters, W. (2003). *Linguistic Complexity*. Utrecht: LOT.
- Dahl, Ö. (2004). *Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins.
- Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, 7(3), 415–449.

LANGUAGE ORIGINS IN LIGHT OF NEURO-ATYPICAL COGNITION AND SPEECH PROFILES

WOLFRAM HINZEN

*Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra,
Barcelona, Spain
wolfram.hinzen@upf.edu*

JOANA ROSSELLÓ

*General Linguistics Department, Universitat de Barcelona
Barcelona, Spain
joana.rossello@ub.edu*

Crow's (1997) suggestion that 'schizophrenia is the prize that Homo Sapiens pays for language' first linked schizophrenia (SZ) to language evolution. The received view that language is a mere expression of thought and that cognition has primacy over communication, on the other hand, is shared by both lay people and scholars as different as Crow, Frith, Hurford and Fodor. Yet both SZ and ASD may suggest the opposite, as we argue here in line with the un-Cartesian program (Hinzen & Sheehan, 2013). Speech processing as such presents with abnormalities in either condition, which can be derived neither from an intangible thought process separated from language, nor from non-speech related distal problems. These abnormalities include: auditory verbal hallucinations (AVHs) in SZ; the lack of a selective attentional bias for speech over non-speech from birth in ASD (Blasi et al., 2015) many 'first rank symptoms' (thought broadcasting, thought withdrawal, etc.) in SZ; and the reversal of the typical comprehension over production advantage in development in ASD. These abnormalities of speech processing, which go along with atypical cognitive profiles, would be a coincidence according to the received view. If speech processing, instead, differs from thought only insofar as thought (as normally processed) is non-overt essentially, abnormalities in thought can follow from abnormalities in speech processing. Cognition and communication would go hand in hand, integrated but distinguishable in normalcy, and disintegrated in different ways in SZ and ASD.

Sociality is (profoundly) impeded in both SZ and ASD; and also speech, albeit in different ways. Could neuro-atypical speech in both conditions be derived from a social impairment? We suggest it is the other way round. Apart

from the telling case of speechless children with ASD, who are the most impaired socially and cognitively, there are general problems with person deixis in both conditions. But deixis is inherent to and originates in speech. In ASD, the ‘total feedback’ characteristic of speech is difficult to reach in that ‘the so-called internalization of communicative behavior that constitutes a major portion of ‘thinking’ (Hockett, 1960) is overtly precarious (preference for self-reference with proper names or in 3rd person; 1st-2nd person pronoun reversal). In SZ, the deictic frame reached in development in which the world (‘it’), sociality (‘you’) and deictic center (‘I’) are triangulated, breaks down. AVHs concur with a loss of the deictic anchoring of thought and speech, in that thought becomes speech directed at or about the patient. The speech element cannot be subtracted from AVH without losing its substance. It also pervades the rest of reality distortion symptoms to the extent that these, too, adopt the form of an interaction of the self with others. Rational thought may well find its foundation in a healthy speech system deployed in social interaction. A speechless Language of Thought (Fodor 1975) equivalent to its human-specific form has not been documented in any species. Sociality, communication (i.e. speech and co-speech gestures) and thought profiles in humans are inherent facets of natural language, which must inform language evolution.

The role of speech in cognition is also supported by the leading role that speech as such plays in the induction of hemispheric specialization in humans, which is considered of high cognitive import (Hervé et al., 2013) and is abnormal both in SZ and ASD. Brain correlates for either are consistent with the view argued for here but pose a new question: how is it that not all the abnormalities in speech processing (aphasias eg.) give rise to thought disturbances? Our tentative answer targets the insula (Klein et al., 2013), a highly integrative (interoception and exteroception) and phylogenetically novel (present in great apes) cortical structure that is involved both in speech processing (production and perception) and in the dopamine sensitive salience network (with the cingulate), which is differently abnormal in SZ and ASD.

References

- Blasi, A., et al. (2015). Atypical processing of voice sounds in infants at risk for autism spectrum disorder. *Cortex*, 71: 122-133.
- Crow, T. J. (1997). Is schizophrenia the price that Homo sapiens pays for language? *Schizophrenia Research* 28: 127–141.
- Fodor, J.A. (1975). *The language of thought*. Cambridge, MA: Harvard U. Press.
- Hervé, P.-Y et al. (2013). Revisiting human hemispheric specialization with neuroimaging. *Trends in Cognitive Sciences* 17(2): 69–80.
- Hinzen, W. & Sheehan, M. (2013). *The philosophy of universal grammar*. OUP.
- Hockett, C. F. (1960). The Origin of Speech, *Scientific American* 203, 88–111
- Klein, T. A, et al. (2013). Error awareness and the insula: links to neurological and psychiatric diseases. *Frontiers in Human Neuroscience*, 7(February), 14.

EFFORT VS. ROBUST INFORMATION TRANSFER IN LANGUAGE EVOLUTION

T. FLORIAN JAEGER¹, MARYIA FEDZECHKINA²,

¹*Department of Brain and Cognitive Sciences, University of Rochester, Rochester, USA
fjaeger@bcs.rochester.edu*

²*Department of Psychology, University of Pennsylvania, Philadelphia, USA
mfedze@sas.upenn.edu*

In his seminal work, Zipf (1949) popularized the hypothesis that languages are shaped by a trade-off between production effort and robust message transfer. It is hard to overestimate the influence this idea has had in functional linguistics and related approaches (see Piantadosi, 2014 for a comprehensive review). Yet, to this day, there is little direct (rather than correlational) evidence for this trade-off.

Recent large-scale quantitative typological studies have shown that lexicon structure in a variety of languages exhibits properties that are consistent with the hypothesized trade-off (e.g., Ferrer i Cancho et al., 2013; Piantadosi, Tily, & Gibson, 2011). Iterated miniature language learning studies have identified a potential cause for these patterns: biases during learning and communication cause learners to deviate from the input towards languages that conserve effort while still guaranteeing robust communication (Kirby, Tamariz, Cornish, & Smith, 2015).

While this work has identified patterns consistent with the trade-off hypothesis, it has not manipulated effort or the chance of communicative success to directly test the presence of a trade-off. Here we present a crowdsourcing-based miniature language learning experiment that directly assesses whether learners trade off the probability of successful message transmission against the effort associated with producing the message. We ask in particular whether the inverse correlation between word order (WO) flexibility and the presence of a case system in a language is shaped by this trade-off.

In the experiment (administered in 2x45min sessions over 2 consecutive days over Amazon Mechanical Turk), different groups of participants learned miniature artificial languages by watching short videos and hearing their descriptions. All videos depicted human actors performing simple transitive events. Participants first learned the names of the actors and then learned the

grammar through sentence exposure. At the end of each session, participants were shown the entire lexicon of the language at the top of the screen and asked to describe previously unseen scenes by clicking on the corresponding lexical items. All languages had optional case-marking (present on 67% of objects; never on subjects). The languages differed in the amount of WO flexibility: The fixed WO language used SOV 100% of the time, while the flexible WO language used SOV and OSV equally frequently. Thus, the uncertainty about the intended message was low in the fixed and high in the flexible WO language. The critical manipulation was the amount of effort required to produce case. During the production test, participants in the low-effort condition were shown a case-marked and non-case-marked variant of every noun (along with all the verbs). Case production, thus, required the same number of clicks as production of bare nouns. In the high-effort condition, participants saw non-case-marked variants of all nouns along with the two free case-markers. Case production, thus, took 2 additional clicks compared to bare nouns.

While our formalization of effort in terms of additional clicks is an imperfect approximation of production effort in terms of additional syllables, it allows us model the scenarios of case production involving additional effort or not, which would be difficult to tease apart in natural production (there is always additional effort associated with articulating an additional phoneme). If production effort is indeed traded off against robust message transfer, we would expect learners in the high-effort condition to use more case in the language with higher uncertainty about the intended meaning (flexible WO) compared to the language with lower uncertainty (fixed WO). Since there is no difference in effort associated with case use in the low-effort condition, differential case use would not be expected here. The results support our hypothesis. We observed differential case use only in the high-effort condition: Learners tended to maintain case only in the flexible WO language ($p < 0.05$). In contrast, in the low-effort condition, learners of both languages produced the same amount of case, equal to the input proportion ($p > 0.8$). In all conditions, learners matched the input distribution of word order variants.

Our findings suggest that some cross-linguistic patterns are shaped by a trade-off between production effort and robust message transmission. Even though the difference in uncertainty about the message between the flexible and fixed WO languages was equal across the two effort conditions, learners restructured the input language to more closely resemble naturally occurring types only when case production required a substantial effort increase. Our results also highlight the potential of web-based miniature language learning experiments in investigating factors underlying language evolution.

References

- Ferrer i Cancho, R., Hernandez-Fernandez, A., Lusseau, D., Agoramoorthy, G., Hsu, M., & Semple, S. (2013). Compression as a universal principle of animal behavior. *Cognitive Science*, *37*(8), 1565–1578.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87-102.
- Piantadosi, S. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychological Bulletin & Review*, *21*, 1112-1130.
- Piantadosi, S., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proc Natl Acad Sci USA*, *108*(9), 3526. doi: 10.1073/pnas.1012551108
- Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

SIMPLE AGENTS ARE ABLE TO REPLICATE SPEECH SOUNDS USING 3D VOCAL TRACT MODEL

RICK JANSSEN, SCOTT R. MOISIK, DAN DEDIU

*Max Planck Institute for Psycholinguistics,
Nijmegen, the Netherlands*

rick.janssen@mpi.nl, scott.moisik@mpi.nl, dan.dediu@mpi.nl

Many factors have been proposed to explain why groups of people use different speech sounds in their language. These range from cultural, cognitive, environmental (e.g., Everett, et al., 2015) to anatomical (e.g., vocal tract (VT) morphology) properties. How could such anatomical factors have led to the similarities and differences in speech sound distributions between human languages (see Janssen & Dediu (in press) for a theoretical background)?

It is known that hard palate profile variation can induce different articulatory strategies in speakers (e.g., Brunner et al., 2009). That is, different hard palate profiles might induce a kind of *bias* on speech sound production, easing some types of sounds while impeding others. In a population of speakers (with a proportion of individuals) that share certain anatomical properties, even subtle VT biases might become expressed at a population-level (through e.g., *bias amplification*, Kirby et al., 2007). However, before we look into population-level effects, we first have to consider within-individual anatomical factors. For that, we have developed a computer-simulated analogue for a human speaker: an *agent*. Our agent is designed to replicate speech sounds (frequency-domain vowels) using a *production* and *cognition* module in a computationally tractable manner.

Previous agent models have often used more abstract (e.g., symbolic) signals. (e.g., Kirby et al., 2007). We have equipped our agent with a three-dimensional model of the VT (the *production* module, based on Birkholz, 2005) to which we made numerous adjustments. Specifically, we used a 4th-order Bezier curve that is able to capture hard palate variation on the mid-sagittal plane. Using an evolutionary algorithm, we were able to fit the model to human hard palate MRI tracings (see <http://www.mpi.nl/artivark> for our data-collection project), yielding high accuracy fits and using as little as two parameters (Janssen et al., 2015). We can thus use this procedure to import palate measurements into our agent's production module to investigate the effects on acoustics. Furthermore, we also

show that our model's fits are comparable to PCA, but without the reliance on an empirical induction step when *generating* hard palates (Janssen et al., submitted). In effect, we can thus exaggerate/introduce novel biases in order to investigate their effect in the agent model.

Our agent is able to control the VT model using the *cognition* module. Previous research has focused on detailed neurocomputation (e.g., Kröger et al., 2014) that highlights e.g., neurobiological principles, speech recognition performance or time-domain acoustics. However, neither the brain nor temporal dynamics in acoustics are the focus of our current study. Furthermore, present-day computing throughput does not allow for large-scale deployment of these architectures, as required by the population model we are developing. Thus, the question whether a very simple cognition module is able to replicate sounds in a computationally tractable manner, and even generalize over novel stimuli, is one worthy of attention in its own right.

Our agent's cognition module is based on running an evolutionary algorithm on a large population of feed-forward neural networks (NNs). As such, (anatomical) bias strength can be thought of as an attractor basin area within the parameter-space the agent has to explore. The NN we used consists of a triple-layered (fully-connected), directed graph. The input layer (three neurons) receives the formants frequencies of a target-sound. The output layer (12 neurons) projects to the articulators in the production module. A hidden layer (seven neurons) enables the network to deal with nonlinear dependencies. The Euclidean distance (first three formants) between target and replication is used as fitness measure. Results show that sound replication is indeed possible, with Euclidean distance quickly approaching a close-to-zero asymptote.

Statistical analysis should reveal if the agent can also: a) Generalize: Can it replicate sounds not exposed to during learning? b) Replicate consistently: Do different, isolated agents always converge on the same sounds? c) Deal with consolidation: Can it still learn new sounds after an extended learning phase ('infancy') has been terminated? Answering these questions forms the foundation of the investigation of anatomical biases on a population level.

References

- Birkholz P. (2005). *3D-Artikulatorische Sprachsynthese*. Logos Verlag, Berlin.
- Brunner, J., Fuchs, S., & Perrier, P. (2009). On the relationship between palate shape and articulatory behavior. *J. Acoust. Soc. Am.*, 125 (6), 3936-49.
- Janssen, R., Dediu, D. (in press). Genetic biases in language: Computer models and experimental approaches in (Villavicencio, A., Poibeau, T.) *Language, Cognition and Computational Models*. Cambridge University Press.
- Janssen, R., Moisik, S.R., Dediu (2015). Bézier modelling and high accuracy curve fitting to capture hard palate variation. In *Proc. of the 18th International*

- Congress of Phonetic Sciences*. Glasgow, UK. Janssen, R., Moisik, S.R., Dediu, D (submitted). Modelling Human Hard Palate Shape with Bézier Curves. *PLoS ONE*.
- Everett, C., Blasi, D. E., & Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proc. Natl. Acad. Sci. U.S.A.*, 112, 1322-1327.
- Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *PNAS*, 104, 5241-5245.
- Kröger, B.J., Kannampuzha, J., & Kaufmann, E. (2014). Associative learning and self-organization as basic principles for simulating speech acquisition, speech production, and speech perception. *EPJ Nonlinear Biomedical Physics*, 2(1), 1-8.

NONLINEAR BIASES IN ARTICULATION CONSTRAIN THE DESIGN SPACE OF LANGUAGE

RICK JANSSEN¹, BODO WINTER², SEÁN ROBERTS¹,
SCOTT R. MOISIK¹, DAN DEDIU¹

¹*Max Planck Institute for Psycholinguistics,
Nijmegen, the Netherlands*

rick.janssen@mpi.nl, sean.roberts@mpi.nl, scott.moisik@mpi.nl, dan.dediu@mpi.nl

²*Cognitive & Information Sciences, University of California, Merced,
California, United States of America
bodo@bodowinter.com*

In Iterated Learning (IL) experiments, a participant's learned output serves as the next participant's learning input (Kirby et al., 2014). IL can be used to model cultural transmission and has indicated that weak biases can be amplified through repeated cultural transmission (Kirby et al., 2007). So, for example, structural language properties can emerge over time because languages come to reflect the cognitive constraints in the individuals that learn and produce the language. Similarly, we propose that languages may also reflect certain *anatomical* biases. Do sound systems adapt to the affordances of the articulation space induced by the vocal tract?

The human vocal tract has inherent nonlinearities which might derive from acoustics and aerodynamics (cf. quantal theory, see Stevens, 1989) or biomechanics (cf. Gick & Moisiuk, 2015). For instance, moving the tongue anteriorly along the hard palate to produce a fricative does not result in large changes in acoustics in most cases, but for a small range there is an abrupt change from a perceived palato-alveolar [ʃ] to alveolar [s] sound (Perkell, 2012). Nonlinearities such as these might bias all human speakers to converge on a very limited set of phonetic categories, and might even be a basis for combinatoriality or phonemic 'universals'.

While IL typically uses discrete symbols, Verhoef et al. (2014) have used slide whistles to produce a continuous signal. We conducted an IL experiment with human subjects who communicated using a software-implemented slide whistle for which the degree of nonlinearity is controlled. A single parameter (α) changes the mapping from slide whistle position (the 'articulator') to the acoustics. With $\alpha=0$, the position of the slide whistle maps Bark-linearly to the acoustics. As α approaches 1, the mapping gets more double-sigmoidal, creating

three plateaus where large ranges of positions map to similar frequencies. In more abstract terms, α represents the strength of a nonlinear (anatomical) bias in the vocal tract.

Six chains (138 participants) of dyads were tested, each chain with a different, fixed α . Participants had to communicate four meanings (pictographs showing different animals) by producing a continuous signal using the slide-whistle in a ‘director-matcher’ game, alternating roles (cf. Garrod et al., 2007).

Results show that for high α s, subjects quickly converged on the plateaus. This quick convergence is indicative of a strong bias, repelling subjects away from unstable regions already within-subject. Furthermore, high α s lead to the emergence of signals that oscillate between two (out of three) plateaus. Because the sigmoidal spaces are spatially constrained, participants increasingly used the sequential/temporal dimension with higher α s (i.e., more nonlinear mappings). As a result of this, the average duration of signals with high α was ~ 100 ms longer than with low α . These oscillations could be an expression of a basis for phonemic combinatoriality.

We have shown that it is possible to manipulate the magnitude of an articulator-induced non-linear bias in a slide whistle IL framework. The results show that language might indeed come to reflect the nonlinear mapping from the articulators to acoustics. In particular, the signaling systems in our study quickly converged (within-subject) on the use of stable regions. While these conclusions were drawn from experiments using slide whistles with a relatively strong bias, weaker biases could possibly be amplified over time by repeated cultural transmission, and likely lead to similar outcomes.

Future studies could investigate anatomical biasing with more realistic models of the articulators, and address the interaction with other factors (socio-linguistics, environment, etc.). Our model is a deliberate abstraction from reality in order to tightly control experimental conditions. In reality of course, anatomical biases have to be thought of as one factor that shapes human language, but in complex ways.

References

- Garrod S., Fay N., Lee J., Oberlander J., MacLeod T. (2007) Foundations of representation: where might graphical symbol systems come from? *Cognitive Science*, 31, 961-987.
- Gick, B., & Moisik, S. R. (2015). Quantal biomechanics in an embodied phonetics. *JASA*, 137(4), 2301–2301.
- Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *PNAS*, 104, 5241-5245.

- Kirby, S., Griffiths, T.L., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28, 108-114.
- Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *J. Neuroling.*, 25(5), 382-407.
- Stevens, K. N. (1989). On the quantal nature of speech. *J. Phon.*, 17, 3-46.
- Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43, 57-68.

MODELING LANGUAGE CHANGE TRIGGERED BY LANGUAGE SHIFT

ANNA JON-AND

*Centre for the Study of Cultural Evolution, Stockholm University
Stockholm, Sweden*

*School of Languages and Media Studies, Dalarna University
Falun, Sweden*

anna.jon-and@su.se

ELLIOT AGUILAR

*Department of Biology, University of Pennsylvania
Philadelphia, USA*

elliott.g.aguilar@gmail.com

Language shift is widely believed to accelerate change in the target language, an effect which is generally attributed to innovations introduced by new speakers during the second language acquisition (SLA) process (Thomason & Kaufman, 1988). If this hypothesis is correct, then the rate of contact-induced language change in a language shift context should be related to the rate at which second language (L2) speakers enter the population. Unfortunately, little diachronic data exists to test this hypothesis. The aim of the present paper is to model the mechanism that makes SLA accelerate language change on a population level and compare its predictions to a rare diachronic data set from the ongoing language shift in Maputo, Mozambique.

To model linguistic interaction, we adapted Jansson et al.'s model of creole formation (Jansson et al. 2015). At each time step, all speakers met in pairwise interactions and chose to utter one of n variants of a linguistic feature based on their probability distribution of usage. Each agent then modified their distribution of usage based on what they heard by using a linear updating rule with parameter l . After a round of interactions, population turnover occurred with some individuals dying and new L1 and L2 speakers entering the population with rates b and r , respectively. Newborn L1s chose two linguistic 'parents' at random and averaged their usage distributions to initialize their own. L2 individuals started with the population mean frequencies of usage. However, with probability μ , a newly recruited L2 speaker could assign all the probability

mass to a 'mutant' variant. We explored the general behavior of the model in both fixed and expanding populations for 100 years, with 365 rounds of interaction per year. We then ran a specific set of runs parameterized by demographic data (number of L1 and L2 speakers) from Maputo over a thirty-two year period (1975-2007). We compared our model runs with diachronic data on innovative preposition use and reduced verbal morphology in Maputo Portuguese from two time points (1993 & 2007), presuming that the use of the innovative forms was zero in 1975, as the spread of Portuguese through massive L2 acquisition started only after this year. The datasets comprehend 12 hours of recordings with 20 participants in similar circumstances from each time point, where variation between innovative and conservative forms is quantified.

As predicted, our results show that the rate of increase in usage of the novel variant was most strongly dependent on the rate at which L2 speakers entered the population, r , as well as the mutation rate, μ . In the Maputo runs, however, our data points did not fall within the 95% confidence intervals of any of our parameter groupings. We then modified the model to allow the L2 speakers to continue to introduce variation for the first five years they were in the population, to represent the fact that the SLA process occurs over time. Using the same criterion we found agreement between the simulation and the preposition data, while the verb data continued to diverge from model predictions. Importantly, our model assumed neutral evolution of the linguistic features. The departure of the verb data from our model predictions may indicate the presence of selection pressures or biases, for instance, the new verb forms being more economical.

Agent-based models have been successfully used in the field of cultural language evolution for explaining the emergence of linguistic structure (e.g. Kirby 2001), whereas change in already established structures seems to be more difficult to account for. Recent theoretical papers (Blythe & Croft, 2012; Pierrehumbert et al., 2014) have aimed at modeling the propagation of a single innovation (introduced by one speaker) in a population, thus accounting for language change with no pressure from contact. In these models, conditions such as biases and/or innovator network position, are required for the novel variant to be successful. Our simulations demonstrate how with minimal assumptions novel variants can be introduced and spread in a population, due to multiple introductions by different individuals. We thus suggest that this may be a basic typological difference between contact-induced and non-contact-induced language change, which would explain how SLA may increase language change in shift situations.

References

- Blythe, R. A., & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, 88(2), 269-304.
- Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *Evolutionary Computation, IEEE Transactions on*, 5(2), 102-110.
- Jansson, F., Parkvall, M., & Strimling, P. (2015). Modeling the Evolution of Creoles. *Language Dynamics and Change*, 5(1), 1-51.
- Pierrehumbert, J. B., Stonedahl, F., & Daland, R. (2014). A model of grassroots changes in linguistic systems. *arXiv preprint arXiv:1408.1985*.
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Univ of California Press.

THE EVOLUTION OF ZIPF'S LAW OF ABBREVIATION

JASMEEN KANWAL, KENNY SMITH, JENNIFER CULBERSTON, SIMON KIRBY

*Centre for Language Evolution
University of Edinburgh
s1252181@sms.ed.ac.uk*

As Zipf observed in 1935, human languages appear to exhibit an inverse relationship between word length and word frequency; the higher the frequency of a word, the shorter it tends to be. Since then, this inverse relationship (Zipf's Law of Abbreviation, or ZLA) has been observed in a wide range of languages (see, e.g., Sigurd et al., 2004; Piantadosi et al., 2011; Ferrer-i-Cancho & Hernández-Fernández, 2013).

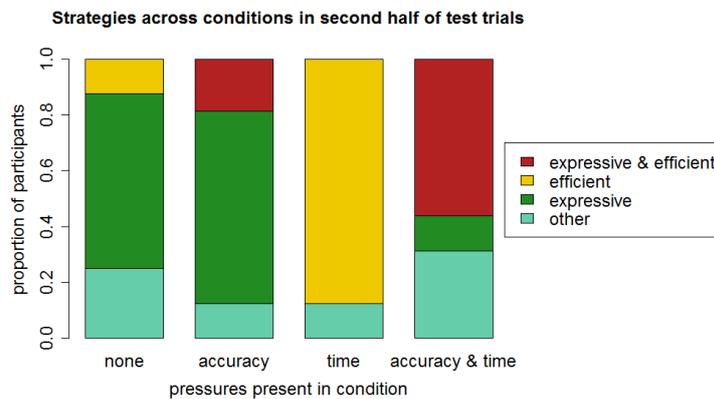
What causes languages to align length and frequency in this way? Zipf hypothesised that speakers operate in accordance with a Principle of Least Effort (PLE). Combined with a pressure to communicate successfully, the pressure to minimise effort for speakers—here, by minimising the length of an utterance—would lead optimally to the shortest forms being mapped to the most frequent meanings, leaving longer forms to describe rarer referents. This hypothesis (H1) relies on the assumption that speakers are sensitive to frequency differences, and can optimise form-meaning mappings accordingly. However, an alternative hypothesis (H2) states that ZLA can be explained purely by invoking cognition-external statistical facts about randomly generated systems (Moscoso del Prado, 2013; Ferrer-i-Cancho & Moscoso del Prado, 2012).

If ZLA is a result of the PLE, as H1 suggests, then we should be able to observe speakers actively optimising form-meaning mappings during communication. We test this by using an artificial language task manipulating the frequency of meanings. We predict, following H1, that speakers will spontaneously map forms to meaning in a way that optimises efficiency.

We trained participants to learn a long name and a 'clipped' name for each of two objects. The clipped name was the same for both objects, while the long names were distinct. Crucially, one object appeared more frequently than the other. In the critical condition, pairs of participants played a communication game in which they took turns transmitting the name of the object they saw to their partner, who then had to guess which of the two objects the first partner was seeing. Participants could choose to transmit either the long or short name to their partner. The longer names took a longer amount of time to transmit, so that greater length

was associated with greater ‘effort’. Pairs were rewarded for completing the task in the quickest time (introducing a pressure to minimise effort), while maximising the number of correct guesses (pressure to communicate successfully). Three control conditions were included for a full 2x2 manipulation of time and communicative pressures.

In the condition with a time pressure but no communicative pressure (N=8), participants mapped both objects to the ambiguous short form (‘efficient strategy’). In the condition with only a communicative pressure (N=8 pairs), most participants retained the unique long forms for each object (‘expressive strategy’). Crucially, in the condition with pressures to communicate both accurately and quickly (N=8 pairs), by the end of test trials, most participants converged on the optimal strategy wherein the most frequent object was mapped to the ambiguous short name, and the infrequent object to its unique long name, making the ‘language’ both efficient and expressive.



Because the optimal strategy was significantly more likely to occur in the critical condition than in any of the controls, we conclude that speakers are sensitive to frequency differences, and when subject to pressures to communicate accurately and efficiently, they actively map forms to meanings in a way that optimises both these factors. These results are proof of concept that the PLE is in fact a plausible explanation for why the inverse relationship between word length and frequency is so widespread in human languages and communication systems; it results from a gradual accumulation over time of speakers’ optimising behaviour.

References

Ferrer-i-Cancho, R., & Hernández-Fernández, A. (2013). The failure of the law

- of brevity in two new world primates. statistical caveats. *Glottology International Journal of Theoretical Linguistics*, 4(1), 45–55.
- Ferrer-i-Cancho, R., & Moscoso del Prado, F. (2012). Information content versus word length in random typing. *JSTAT*.
- Moscoso del Prado, F. (2013). The missing baselines in arguments for the optimal efficiency of languages. In *Proceedings of the 35th annual conference of the cognitive science society* (p. 1032-1037).
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *PNAS*, 108(9), 3526–3529.
- Sigurd, B., Eeg-Olofsson, M., & Van Weijer, J. (2004). Word length, sentence length and frequency-Zipf revisited. *Studia Linguistica*, 58(1), 3752.
- Zipf, G. K. (1935). *The psycho-biology of language* (Vol. ix). Oxford, England: Houghton, Mifflin.

THE SPONTANEOUS EMERGENCE OF LINGUISTIC DIVERSITY IN AN ARTIFICIAL LANGUAGE

DEBORAH KERR, KENNY SMITH

*Centre for Language Evolution, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, EH8 9AD, UK
kenny.smith@ed.ac.uk*

Human languages are socially learnt, and this social learning permits the persistence of linguistic diversity at many levels (language families, languages, dialects, accents, familylects). Language therefore provides an excellent marker of social identity, and this social function for language has been implicated in the generation of linguistic diversity: the linguistic systems of competing social groups will be driven apart due to social selection, resulting in linguistic divergence and the emergence of reliable markers of group identity (Nettle, 1999). Moreover, the evolution of the capacity for vocal learning itself may be driven by the need to acquire ever more subtle social group markers (a possibility known as the *password hypothesis*, see e.g. Fitch, 2000).

Roberts (2011) explored the role of between-group competition in linguistic diversification. He developed an experimental paradigm where sets of participants played a trading game involving requesting objects from each other using an artificial language. He found that password-like social group markers developed in conditions where there was competition between groups and high frequency of interaction within groups; this was driven by the intentional generation of linguistic diversity by his participants. Here we present an experimental paradigm, combining artificial language learning with the Minimal Group method borrowed from social psychology (e.g. Tajfel, 1970), and demonstrate the spontaneous emergence of linguistic diversity despite the absence of functional pressures for social differentiation.

We ran an experiment in which groups of four participants (72 participants total) were trained on a miniature language and then communicated dyadically using that language. After arriving in the lab, participants completed a short-form personality questionnaire. They were then trained on the miniature language: all four participants were seated in a room with a large shared monitor and speakers, saw objects appear on the screen and heard names of those objects (novel words). The target language featured two labels for every object, both used equally frequently during training, and therefore exhibited unpredictable

variation (of a form similar to that used in e.g. Reali & Griffiths, 2009). After training, all four participants sat together around a table and took turns naming objects for the person sitting opposite them (their partner) and selecting objects from an array based on their partner's label. All four participants could therefore hear the labels used by all other participants (and the feedback signal indicating whether the label was correctly interpreted by the partner), but only ever had to make object selections based on the descriptions of one individual, their partner.

We manipulated whether the four participants functioned as a single social group or two separate groups (the two pairs of partners) by constructing artificial social groups. In the Group condition the experiment proceeded as outlined above. In the Pairs condition, after completing the personality questionnaire, participants were told they had been allocated to two separate groups (Apples and Oranges) based on their personality scores; the two groups wore distinctively coloured bibs and sat together during training, separated from the other group by a screen but still receiving identical training on the shared display and speakers. All four participants sat around a table together during interaction, as in the Group condition; individuals were partnered with their group-mate, i.e. the Apples sat opposite each other, as did the Oranges.

We observed linguistic divergence in both conditions: participants were more likely to use the same object names as their partner than their non-partners. However, divergence was significantly higher in the Pairs condition than the Group condition: the mere allocation to social groups resulted in the partnered pairs diverging in their label choices, each pair using different labels for the same objects, despite their full exposure to the labels used by the entire group. Post-experiment debrief suggested this was not a conscious strategy, and would be afunctional in the object-selection communication task anyway.

This data shows that linguistic divergence can occur purely as a consequence of the existence of social groups. This suggests that socially learnt linguistic systems will rapidly acquire the social functions central to the password hypothesis even without pre-existing functional pressure for social differentiation.

References

- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, 4, 258-267.
- Nettle, D. (1999). *Linguistic diversity*. Oxford: Oxford University Press.
- Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111, 317-328.
- Roberts, G. (2011). An experimental study of social selection and frequency of interaction in linguistic diversity. *Interaction Studies*, 11, 138-159.
- Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96-102.

A GENERAL AUDITORY BIAS FOR HANDLING SPEAKER VARIABILITY IN SPEECH? EVIDENCE IN HUMANS AND SONGBIRDS

BUDDHAMAS KRIENGWATANA

*Department of Psychology, University of Amsterdam,
Amsterdam, the Netherlands
bkrieng@alumni.uwo.ca*

PAOLA ESCUDERO¹, ANNE H KERKOVEN², CAREL TEN CATE²

¹*MARCS Institute, University of Western Sydney
Sydney, Australia*

²*Institute for Biology Leiden, Leiden University
Leiden, the Netherlands*

*Paola.Escudero@westernsydney.edu.au, annekerkhoven@hotmail.com,
c.j.ten.cate@biology.leidenuniv.nl*

1. Abstract

Different speakers produce the same speech sound differently, yet listeners are still able to reliably identify the speech sound. A compelling example of our ability to distinguish speech sounds despite enormous variability arising from speaker, gender, and age differences is in the case of vowels. Despite the large between-speaker variation within a vowel category and striking overlap between vowel categories, human adults, pre-linguistic infants, and even nonhuman animals are able to classify vowels of different speakers and genders. How is this achieved, and are they achieved in the same way by human adults, infants, and nonhuman animals?

Perceptual adjustments to accommodate for speaker differences in vowels may possibly be achieved pre-attentively via low level processing mechanisms. Combined with findings suggesting nonhuman animals also adjust for speaker differences, this raises an intriguing possibility that there is a tendency for the vertebrate auditory system to automatically accommodate for speaker differences in vowel production. If this is the case, then exposure to speaker-

variability in vowel production need not be necessary in order for listeners to compensate for speaker and gender differences.

The aim of this study was to compare the ability of humans and zebra finches to categorize vowels despite speaker variation in speech in order to test the hypothesis that accommodating speaker and gender differences in isolated vowels can be achieved without prior experience with speaker-related variability.

Using a behavioral Go/No-go task and identical stimuli, we compared Australian English adults' (naïve to Dutch) and zebra finches' (naïve to human speech) ability to categorize /ɪ/ and /ɛ/ vowels of a novel Dutch speaker after learning to discriminate those vowels from only one other speaker. The Go/No-go task requires subjects to make a response towards vowel stimuli assigned to one category (Go) and to inhibit responses toward vowel stimuli assigned to the other category (No-go). Experiments 1 and 2 presented vowels of two speakers interspersed or blocked, respectively. If experience with speaker variability in vowel production is necessary for successful normalization to occur, then we predicted that zebra finches and humans would not be able to discriminate the vowels of the second, new speaker.

Results demonstrate that categorization of vowels is possible without prior exposure to speaker-related variability in speech for zebra finches, and in non-native vowel categories for humans. This study is the first to provide evidence for what might be a species-shared auditory bias that may supersede speaker-related information during vowel categorization, although it is also possible that different perceptual mechanisms underlie performance in humans and songbirds. The role of experience with speaker-related variability may be to tune the auditory system to the most relevant acoustic parameters that define phonetic categories. Our results do not seem to be adequately explained by existing vowel normalization algorithms (e.g. formant ratios, exemplar-based models). Future investigations of alternative accounts of vowel normalization should incorporate the possibility of an auditory bias for disregarding between-speaker variability, and bear in mind that there are many similarities between humans and zebra finches in vocal production, characteristics of the acoustic vocal signal, auditory perception, and need for accurate perceptual categorization. Thus, it may not be so surprising that there are also parallels in perceptual mechanisms that allow both species to overcome the problem of separating variability associated with content of the signal from variability arising from the individual signaler.

CUMULATIVE VOCAL CULTURES IN ORANGUTANS AND THEIR ONTOGENETIC ORIGIN

ADRIANO R. LAMEIRA

JEREMY KENDAL

*Department of Anthropology, Durham University,
Durham, UK*

adriano.lameira@durham.ac.uk

jeremy.kendal@durham.ac.uk

MARCO GAMBA

*Zoology Department, University of Torino
Torino, Italy*

marco.gamba@unito.it

Recently, several lines of evidence indicate that orangutans (*Pongo* sp.) – the earliest diverging great ape lineage – are capable of expanding their species-specific vocal repertoire with new (voiced and voiceless) calls. These calls are shared and learned between individuals of the same cultural community. Orangutans represent, thus, a desirable model species for the study of language and speech precursors within the human lineage, since human spoken languages are fundamentally characterized by being learned. In the first section of this talk, based on the largest and most comprehensive call database ever assembled in orangutans (currently comprising 9 wild populations across Sumatra and Borneo), and perhaps among any great ape species, we show that orangutan vocal repertoires across populations show a nested structure – a signature of cultural build up that indicates that orangutan vocal cultural repertoires emerge and culturally evolve through a process of “sound upon sound”. The identification of these orangutan vocal cultures raises, however, questions about their ontogenetic origins. In the second section of this talk, we present a case of extreme vocal malleability in a wild Sumatran orangutan infant (approximately 5 years old), who exhibits a repertoire 2 to 4-fold the size of that of adults. A modest dataset of less than 150 recordings collected from this young individual increased the known orangutan vocal repertoire hitherto described by more than 20%. This flexibility verifies that assemblages of cultural calls may be indeed acquired by infant orangutans and subsequently passed on through generations. Altogether, our findings indicate that vocal

cultures in orangutans are real and may ontogenetically emerge through similar mechanisms as human vocal cultures. Like children, orangutan infants exhibit a latent degree of vocal malleability that expressively surpasses that of adults, and they experience a process of cultural trimming within their “linguistic” community in the process of development of the adult repertoire. Once present in our last great ape common ancestor, similar vocal skills would have allowed the rise and preservation of vocal cultures comparable at a basic level with modern human spoken languages.

1. Background

Language evolution is a scientific puzzle. As far as our understanding of language evolution goes, many puzzle pieces remain yet unrevealed. One of the most critical pieces that have proven particularly elusive is the identification of a vocal cultural system among our closest relatives, the great apes. Unlike humans, virtually no primate species shows the capacity to culturally transmit and preserve vocal cultures. A growing number of studies of great ape vocal behavior indicate, however, that orangutans may indeed be capable of learning new call types and expanding their vocal repertoire beyond the species-specific repertoire [1–4]. How these cultural call repertoires ontogenetically develop, and whether they exhibit the potential to evolve cumulatively as human cultures, remains to be investigated.

1.1. Methods

Nestedness analyses were conducted based on vocal repertoire assemblages across orangutan populations as part of the largest database ever assembled of orangutan calls, and results have been replicated resorting to computer simulations.

Infant data were collected opportunistically in the wild (Sikundur Research Station) and analysed via automated cluster analyses based on acoustic similarity as measured through pairwise comparisons using dynamic time warping.

1.2. References and Citations

- 1 Lameira, A.R. *et al.* (2013) Orangutan (*Pongo* spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *J. Acoust. Soc. Am.* 134, 1–11
- 2 Wich, S. *et al.* (2009) A case of spontaneous acquisition of a human sound by an orangutan. *Primates* 50, 56–64
- 3 Wich, S.A. *et al.* (2012) Call cultures in orang-utans? *PLoS One* 7, e36180
- 4 Hardus, M.E. *et al.* (2009) Tool use in wild orang-utans modifies sound production: a functionally deceptive innovation? *Proc. R. Soc. B Biol. Sci.* 276, 3689–3694

LEARNABILITY PRESSURES INFLUENCE THE ENCODING OF INFORMATION DENSITY IN THE LEXICON

MOLLY L. LEWIS AND MICHAEL C. FRANK

*Psychology Department, Stanford University
Stanford, CA, USA
mll@stanford.edu, mcfrank@stanford.edu*

A universal feature of language is that words vary in their length within a language, in terms of morphemes, syllables, and phonemes. There have been several accounts of this variability in the literature that appeal to the form of language itself, such as the frequency of a word, or its predictability in context. Information theory, however, suggests another factor that might influence length: the predictability, or complexity, of a word's meaning (Frank, A. & Jaeger, 2008). This theory predicts that if speakers try to maintain a constant rate of information across the speech stream, then more complex meanings should be longer.

Previous work has shown evidence for this *complexity bias*. Experimentally, participants tend to assign longer words to more complex meanings, relative to shorter words (Lewis, Sugarman, & Frank, 2014). And natural languages also show this bias (Lewis & Frank, under review). To estimate the bias in each language, ratings of conceptual complexity were collected for 499 English words and then translated into 79 additional languages using Google Translate. For each language, there was a correlation between word length and conceptual complexity (grand mean $r = .34$), and this bias held controlling for word frequency (grand mean $r = .22$) and other semantic variables, like concreteness. Despite the presence of this bias across all languages we examined, however, there was also a large degree of variability ($SD = .12$; Fig. 1).

In our work here, we explore one possible account of this variability: that the degree to which a language encodes conceptual complexity in the lexicon is related to the learnability pressure on the language. Learnability pressure has been argued as one factor influencing the morphological complexity of a language (Bentz & Winter, 2013; Lupyán & Dale, 2010; Nettle, 2012). Under this hypothesis, languages are thought to adapt to their particular social context, depending on the cohesiveness of the population acquiring the language. Languages that are acquired by a diverse population of speakers—many adult learners, for example—might be morphologically simpler, than those acquired only by children. Consistent with this prediction, languages that are spoken by more people tend to be

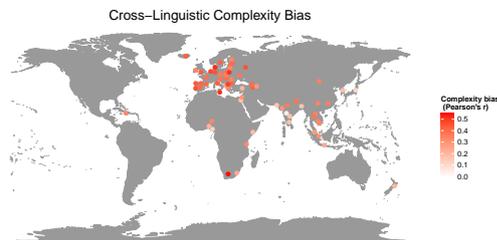


Figure 1 Magnitude of the complexity bias across 79 languages. Each point corresponds to a language.

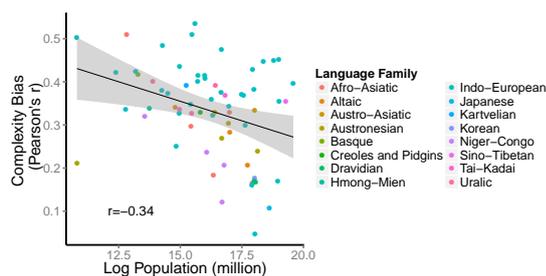


Figure 2 Relationship between complexity bias and speaker population. Each point corresponds to a language.

morphologically simpler.

We hypothesized that the same force might influence the degree to which languages encode information density in the lexicon. We calculated the correlation between a language's complexity bias and its population of speakers. Consistent with previous work, we found that languages with more speakers tend to have a smaller complexity bias ($r = -.34$; Fig. 2), and this result remained reliable even after controlling for language family. Thus, learnability pressure may force languages to rely on non-lexical strategies to maintain a uniform information density across speech.

References

- Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, 3(1), 1–27.
- Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.
- Lewis, M., & Frank, M. C. (under review). The length of words reflects their conceptual complexity.
- Lewis, M., Sugarman, E., & Frank, M. C. (2014). The structure of the lexicon reflects principles of communication. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, 5(1), e8559.
- Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1597), 1829–1836.

A DEVELOPMENTAL PERSPECTIVE ON LANGUAGE ORIGIN: CHILDREN ARE OLD HANDS AT GESTURE

CASEY J. LISTER

*School of Psychology, University of Western Australia,
35 Stirling Highway, Crawley, Western Australia 6009
casey.lister@research.uwa.edu.au*

TIARN BURTENSHAW, NICOLAS FAY, BRADLEY WALKER, JENEVA OHAN

*School of Psychology, University of Western Australia,
35 Stirling Highway, Crawley, Western Australia 6009
21282388@student.uwa.edu.au, nicolas.fay@gmail.com, bradley.walker@uwa.edu.au,
jenava.ohan@uwa.edu.au*

The capacity for language is a distinguishing feature of our species. A problem for those studying its origin is that our pre-linguistic ancestors, who used language in its earliest forms, no longer exist. Instead, we must draw conclusions based on studies of modern humans with fully fledged languages. This makes it difficult to assess the impact of culture and convention on the creation of novel sign systems. The current study addresses this issue through a referential communication task that examines how participants aged 6-12 years create novel sign systems using gestures or vocalisations (sounds that are not words). As children have less developed linguistic systems, and less exposure to conventionalised signs, they offer a new perspective on how people create novel sign systems when prevented from using their pre-existing language.

1. Method

Fifty-four children were recruited from three age groups (6-7, 8-9 & 10-12 years). Each child was presented with two lists of 18 concepts (6 Nouns, 6 Verbs and 6 Adjectives). They communicated the concepts in one list using gestures, and in the other, using vocalisations. The children were filmed and the signs they created were played back to adult participants (N = 36) who attempted to guess the meaning of each sign. A coder also gave each sign an iconicity rating based on the degree to which it resembled its referent. Signs were rated on a 7-point scale (0 = no iconicity, 6 = high iconicity). Fay, Arbib and Garrod (2013) and Fay, Lister, Ellison and Goldin-Meadow (2014) suggested that communication through gesture is more effective than through vocalisation because gesturing enables the creation of more highly iconic signs. By rating the iconicity of each sign, we were able to experimentally test this hypothesis using participants with a less established conventionalised language system.

2. Results & Discussion

As Figure 1 shows, Identification Accuracy (i.e. the proportion of signs that were correctly identified by the adult participants), and Iconicity were both greater in the gesture condition compared to the vocal condition. Adults were also better able to identify the signs produced by older children, compared to younger children. This suggests that with age, children become increasingly adept at producing understandable signs. The positive correlation between Identification Accuracy and Iconicity confirms that greater iconicity is associated with improved comprehension. Previously the effect of sign iconicity upon identification success had been purely speculative.

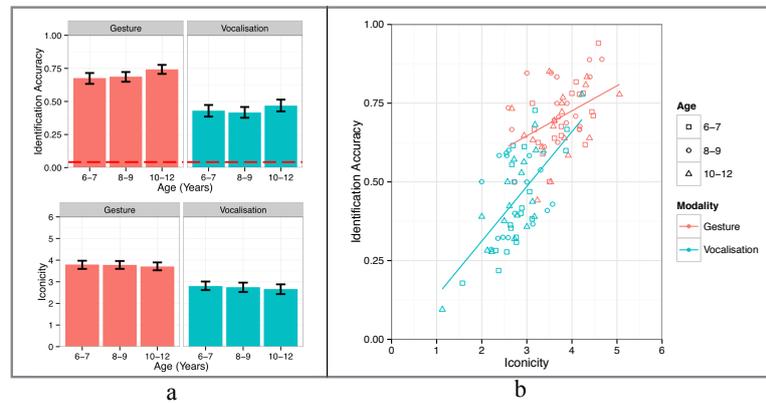


Figure 1. Comparison between Identification Accuracy (red line represents accuracy level predicted by chance) and Iconicity in each modality, across age groups (a) and correlation between Identification Accuracy and Iconicity (b).

This study supports the suggestion that gesturing enables the production of more iconic, better understood signs. Moreover, we have demonstrated that this relationship exists among younger participants who possess a smaller repertoire of conventionalised gestures and vocalisations, thereby reducing the impact of culture and convention upon the creation of a novel communication system.

References

- Fay, N., Lister, C. J., & Ellison, T. M. (2014). Creating a communication system from scratch: gesture beats vocalization hands down. *Frontiers in ...* <http://doi.org/10.3389/fpsyg.2014.00354/abstract>
- Fay, N., Arbib, M., & Garrod, S. (2013). How to Bootstrap a Human Communication System. *Cognitive Science*, 37(7), 1356–1367. <http://doi.org/10.1111/cogs.12048>

EMERGENCE OF SIGNAL STRUCTURE: EFFECTS OF DURATION CONSTRAINTS

HANNAH LITTLE, KEREM ERYILMAZ AND BART DE BOER

*Artificial Intelligence Laboratory, Vrije Universiteit Brussel
Brussels, Belgium*

hannah@ai.vub.ac.be, kerem@ai.vub.ac.be, bart@arti.vub.ac.be

Recent work has investigated the emergence of structure in speech using experiments which use artificial continuous signals. Some experiments have had no limit on the duration which signals can have (e.g. Verhoef et al., 2014), and others have had time limitations (e.g. Verhoef et al., 2015). However, the effect of length constraints on the structure in signals has not been experimentally investigated.

Physical, functional or cultural pressures will effect how long signals in the real world can be. Obviously, speech is constrained by breath. Social and functional pressures for transmitting information quickly, succinctly and with little effort will also create pressures for signals to be shorter (Piantadosi et al., 2011).

Signal duration will affect signal structure. Having shorter signals may limit redundancy and influence how quickly signal units are discretised and reused.

We carried out a signal creation experiment. Participants created continuous auditory signals using a Leap Motion. The pitch of signals could be manipulated by the position of a participant's hand in relation to the Leap Motion (see Little et al., 2015, for a summary and justification of the paradigm). Participants created signals for a set of meanings. No two meanings had any features (shape, colour or texture) which were shared. Participants took part in two conditions. In the unconstrained condition, signals had no limit on duration (signals had an average length of 3 seconds, $sd = 2.3$). In the constrained condition, signals, monitored using a progress bar, could only be 1 second, shorter than 80% of signals in the unconstrained condition. The experiment had 3 phases, with the meaning space expanding in each phase; 5, 10, 15 meanings in phase 1, 2, 3 respectively. Each phase consisted of a practice session, a signal creation task (participants created signals for each randomly selected meaning), and a signal recognition task (participants heard their own signals and chose between 4 possible meanings for each).

In the constrained condition, participants were worse at recognising their signals (mean = 64% correct), than in the unconstrained condition (mean = 86%). Success levels were not significantly affected by the growth of the meaning space. This discrepancy in success suggests that in the constrained condition, participants

may have had a harder time creating distinct signals. In the constrained condition, signals were much simpler, with a lot of participants relying on static pitch, rather than on patterns and pitch changes. We were able to measure the amount of movement within signals by calculating the variance of the signal trajectory coordinate values, and showed that the amount of movement in trajectories was significantly lower in the constrained condition than in the unconstrained condition (we compared a mixed linear model with a null model, $\chi^2(1) = 9, p < 0.001$).

We also found that in the unconstrained condition, there was a significant downward trend in the amount of systematicity in signals (measured by trajectory predictability given the rest of the signal repertoire) as the meaning space expanded. Signals for meanings introduced later were less predictable than those in earlier phases (we compared a mixed linear model with a null model, $\chi^2(1) = 4, p < 0.05$). This trend did not occur in the constrained condition, maybe suggesting that the limited signal duration stopped participants creating new strategies for new meanings, or constrained the use of redundant features in new signals, both of which would make signals less predictable.

Our results highlight why experimental studies need to consider the effects which time constraints will have on structure, systematicity and redundancy in artificial signals. Further, our time constraints impeded the production of distinct signals, generating a pressure for more efficient strategies for differentiating signals. One potential strategy, which accommodates the crowding of signal spaces, is the use of combinatorial structure. However, further experimental work needs to be done to see if this is the case.

Acknowledgements

Financial support from the ERC starting grant, ABACUS, project number 283435.

References

- Little, H., Eryılmaz, K., & De Boer, B. (2015). A new artificial sign-space proxy for investigating the emergence of structure and categories in speech. In *The proceedings of the 18th international congress of phonetic sciences*. University of Glasgow: Glasgow.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *PNAS*, *108*(9), 3526–3529.
- Verhoef, T., Kirby, S., & De Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, *43*, 57-68.
- Verhoef, T., Roberts, S. G., & Dingemanse, M. (2015). Emergence of systematic iconicity: Transmission, interaction and analogy. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. Jennings, & P. Maglio (Eds.), *The 37th annual meeting of the cognitive science society (cogsci 2015)* (pp. 2481–2487). Austin, TX: Cognitive Science Society.

THE EVOLUTION OF REDUNDANCY IN A GLOBAL LANGUAGE

GARY LUPYAN JUSTIN SULIK

Department of Psychology, University of Wisconsin-Madison
lupyan@wisc.edu jsulik@wisc.edu

Why are there different languages? Religious mythology aside, the usual story is that languages diverge when an initial group of speakers disperses, allowing their ways of speaking to begin to drift independently, instead of together (Sapir, 1921). But consider the explanatory inadequacy of such a *neutral drift* account if applied to a biological organism. Why do birds have different beaks? Because there is random drift in beaks shapes and once a population of birds disperses, their beaks drift independently, instead of together. When explaining animal morphology, we often appeal to adaptive fit: some beaks are better suited for some environments than others. Might similar logic apply to languages as well? Might linguistic diversity reflect, in part, the adaptation of languages to different environments in which they are learned and used?

A version of this idea—the linguistic niche hypothesis—was tested by Lupyán and Dale (2010). The authors reasoned that while all natural languages must be learnable by infants, some languages (e.g., English) are further constrained to be learnable by adults. Insofar as some grammatical paradigms (e.g., complex morphology) are more difficult for adults to learn, languages with many adult learners may become adapted to be more learnable by adults via simplification of their morphological paradigms. Examining relationships between morphological complexity and the socio-demographic niches of different languages confirmed the hypothesis: languages spoken by more people (those with more nonnative speakers) have simpler morphological paradigms.

But might these difference in morphology reflect a more fundamental property? A possibility proposed by Lupyán and Dale (2010) is that languages selected to be *only* learnable by children are selected to be morphologically complex because morphology (e.g., agreement systems, obligatory markings of tense, etc.) increases redundancy, redundancy which may pose challenges for

adult learners (Dale & Lupyan, 2012) but this redundancy may facilitate learning by children by providing additional cues for cohering the linguistic signal.

The current work tests the idea that languages primarily learned by infants have greater informational *redundancy* than languages with more nonnative speakers/adult learners. First, we took advantage of a published dataset for 11 Indo-European languages (Piantadosi, Tily, & Gibson, 2011) from which we could compute the average informativeness of each word based on an N-gram model. The results showed that languages spoken by more people / those with more nonnative speakers, had much higher informativeness per word (lower redundancy), $r=.84$, $p<.001$. These differences in redundancy are in line with theoretic predictions, are confounded by numerous differences between the languages. We undertook a stronger test of our hypothesis by comparing variants of two variants of English: American (AmEng) and British English (BrEng), which differ in the environment in which they are learned and used. AmEng is learned by more adult learners than BrEng (e.g., ~95% of BrEng are native speakers, but only 80%-85% of AmEng speakers are).

We trained N-gram language models on corpora of spoken AmEng (COCA) and spoken BrEng (BNC). We then tested the models both on withheld subsets of the training corpora, and on entirely new corpora created from scripts of American and British TV shows. The results show that for a large set of starting conditions, models of AmEng texts have overwhelmingly higher perplexity (lower redundancy) than BrEng texts and that models trained on AmEng are more generalizable, hinting at the greater compositionality of AmEng.

The work provides prima-facie evidence that linguistic divergence is not random. Knowing about the (social) environment in which a language is learned and used allows us to predict its current structure and possible future trajectory. These results point to specific aspects of language that may be under active selection through cultural evolution, helping us understand the circumstances that led to languages having their current structure.

References

- Dale, R. A., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems*, *15*(3), 1150017.
- Lupyan, G., & Dale, R. A. (2010). Language Structure Is Partly Determined by Social Structure. *PLoS ONE*, *5*(1), e8559.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *PNAS*, *108*(9), 3526–3529.
- Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. Dover Publications.

NONHUMAN ANIMALS' USE OF OSTENSIVE CUES IN AN OBJECT CHOICE TASK

HEIDI LYN¹, STEPHANIE JETT², MEGAN BROADWAY¹, AND MYSTERA SAMUELSON¹

*Department of Psychology, University of Southern Mississippi,
Long Beach, MS, USA
heidi.lyn@usm.edu*

*Department of Psychology, University of South Alabama
Mobile, AL
sjett@southalabama.edu*

One recent argument concerning the evolution of language centers on the ability of the last common ancestor of apes and humans to engage in Gricean communication (that is, communication in which the speaker has the clear intent to produce a response, but also with the intent that the hearer can recognize the communicative intent of the speaker (e.g. Moore, 2015; Scott-Phillips, 2015; Tomasello, 2008). The *standard* argument (see Moore, 2015) has been that true Gricean communication requires fourth-order meta-representation (the speaker intends for the hearer to understand that the speaker intends for the hearer to understand the message) and is therefore unique to humans. In this conception, animal communication is limited to strict associations (coded communication) and Gricean or ostensive communication was the key innovation in the evolution of language.

An *alternative* view suggests that the representation required for ostensive communication is much simpler (see Moore, 2015 for an explanation) – the speaker intends to communicate a message to the hearer and the speaker also intends for that message to represent something in the world (two first-order meta-representations). On the receiving end, then, the hearer must both understand that the speaker is intending to communicate (through overt or covert intentional cues) and must decipher the message (the signal). The intent to

communicate would itself be communicated through an ostensive act or cue – an act designed to draw attention to and facilitate the receipt of a communicative signal. Ostensive cues can include eye contact and shifting of joint attention, among other behaviors. These ostensive cues remain separate from the communicative signal, which could include gestures, verbalizations, etc. Under the standard view, nonhuman animals should rely entirely on associative learning to follow a signal such as pointing. In this scenario, nonhumans should perform at the same level when a communicator uses ostensive cues as when those cues are eliminated.

To evaluate this theory, we tested 3 bonobos (*Pan paniscus*) from the Ape Cognition and Conservation Initiative (ACCI) in Des Moines, IA on the object choice task (point following) both with and without ostensive cues (in this case, ostensive cues included gaze alteration between the gesture and the recipient). When ostensive cues were removed, the apes' performance fell from almost perfect to chance levels, indicating that ostensive cues are vital for the performance of bonobos in this task ($p < .01$, binomial tests).

A new study has been initiated with domestic dogs (*Canis familiaris*) at the Humane Society of South Mississippi in Gulfport, MS. A total of forty dogs will be tested on a variety of ostensive cues to determine which if any are most salient. To date, twenty-five dogs have begun testing, but only twelve passed the initial evaluation of following eye gaze. Of those twelve, eight could follow a distal point, but only two could follow a cross-body point (required for our study). In this stage, the ostensive cues tested were gaze alterations among the three points of the joint attention triad – gesture and recipient, gesture and referent, and recipient and referent. Neither dog could follow points with all ostensive cues removed, although one had no difficulty when alteration to any one point of the triad was eliminated (e.g. eye gaze only moved between gesture and recipient, but not to the referent).

These findings strongly indicate that nonhuman animals utilize overt ostensive cues to recognize gestural communication from humans, suggesting that the development of ostensive cues was not the key innovation that triggered the evolution of human language. Rather, the dog data reinforce the likely crucial social component of gesture comprehension in nonhumans, as the dogs in this study almost certainly had less human interaction and performed less well than pet dogs in other studies, again indicating more than strict associative learning.

References

- Moore, R. (2015). Meaning and ostension in great ape gestural communication. *Animal Cognition*, doi:10.1007/s10071-015-0905-x
- Scott-Phillips, T. C. (2015). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, 56(1), 56-80.
- Tomasello, M. (2008) *Origins of human communication*. MIT Press, Cambridge.

LANGUAGE ADAPTS TO SIGNAL DISRUPTION IN INTERACTION

VINICIUS MACUCH SILVA, SEÁN ROBERTS

*Language and Cognition Department, Max Planck Institute for Psycholinguistics
Nijmegen, Netherlands
vini.macuch@gmail.com, sean.roberts@mpi.nl*

Linguistic traits are often seen as reflecting cognitive biases and constraints (e.g. Christiansen & Chater, 2008). However, language must also adapt to properties of the channel through which communication between individuals occurs. Perhaps the most basic aspect of any communication channel is noise. Communicative signals can be blocked, degraded or distorted by other sources in the environment. This poses a fundamental problem for communication. On average, channel disruption accompanies problems in conversation every 3 minutes (27% of cases of other-initiated repair, Dingemanse et al., 2015). Linguistic signals must adapt to this harsh environment. While modern language structures are robust to noise (e.g. Piantadosi et al., 2011), we investigate how noise might have shaped the early emergence of structure in language.

The obvious adaptation to noise is redundancy. Signals which are maximally different from competitors are harder to render ambiguous by noise. Redundancy can be increased by adding differentiating segments to each signal (increasing the diversity of segments). However, this makes each signal more complex and harder to learn. Under this strategy, holistic languages may emerge. Another strategy is reduplication - repeating parts of the signal so that noise is less likely to disrupt all of the crucial information. This strategy does not increase the difficulty of learning the language - there is only one extra rule which applies to all signals. Therefore, under pressures for learnability, expressivity and redundancy, reduplicated signals are expected to emerge.

However, reduplication is not a pervasive feature of words (though it does occur in limited domains like plurals or iconic meanings). We suggest that this is due to the pressure for redundancy being lifted by conversational infrastructure for repair. Receivers can request that senders repeat signals only after a problem occurs. That is, robustness is achieved by repeating the signal across conversational turns (when needed) instead of within single utterances.

As a proof of concept, we ran two iterated learning chains with pairs of individuals in generations learning and using an artificial language (e.g. Kirby et

al., 2015). The meaning space was a structured collection of unfamiliar images (3 shapes x 2 textures x 2 outline types). The initial language for each chain was the same written, unstructured, fully expressive language. Signals produced in each generation formed the training language for the next generation. Within each generation, pairs played an interactive communication game. The director was given a target meaning to describe, and typed a word for the matcher, who guessed the target meaning from a set. With a 50% probability, a contiguous section of 3-5 characters in the typed word was replaced by 'noise' characters (#). In one chain, the matcher could initiate repair by requesting that the director type and send another signal. Parallel generations across chains were matched for the number of signals sent (if repair was initiated for a meaning, then it was presented twice in the parallel generation where repair was not possible) and noise (a signal for a given meaning which was affected by noise in one generation was affected by the same amount of noise in the parallel generation).

For the final set of signals produced in each generation we measured the signal redundancy (the zip compressibility of the signals), the character diversity (entropy of the characters of the signals) and systematic structure (z-score of the correlation between signal edit distance and meaning hamming distance). In the condition without repair, redundancy increased with each generation ($r=0.97$, $p=0.01$), and the character diversity decreased ($r=-0.99$, $p=0.001$) which is consistent with reduplication, as shown below (part of the initial and the final language):

Gen 0	Outline 1	Outline 2	Gen 5	Outline 1	Outline 2
Shape A	luna	lapi	Shape A	kakakakakakakkaka	kakakakakakakak
Shape B	monepilu	nenana	Shape B	lelelelelelelelelelele	lalalalalalalalalala
Shape C	lunenena	pinenalu	Shape C	mamamamamam	memememe

Linear regressions revealed that generations with repair had higher overall systematic structure (main effect of condition, $t = 2.5$, $p < 0.05$), increasing character diversity (interaction between condition and generation, $t = 3.9$, $p = 0.01$) and redundancy increased at a slower rate (interaction between condition and generation, $t = -2.5$, $p < 0.05$).

That is, the ability to repair counteracts the pressure from noise, and facilitates the emergence of compositional structure. Therefore, just as systems to repair damage to DNA replication are vital for the evolution of biological species (O'Brien, 2006), conversational repair may regulate replication of linguistic forms in the cultural evolution of language. Future studies should further investigate how evolving linguistic structure is shaped by interaction pressures, drawing on experimental methods and naturalistic studies of emerging languages, both spoken (e.g Botha, 2006; Roberge, 2008) and signed (e.g Senghas, Kita, & Ozyurek, 2004; Sandler et al., 2005).

References

- Botha, R. (2006). Pidgin languages as a putative window on language evolution. *Language & Communication*, 26(1), 1-14.
- Christiansen, M. H. and Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(5):489–508
- Dingemans, M., Roberts, S.G., Baranova, J. et al. (accepted) Universal Principles in the Repair of Communication Problems. PLOS ONE.
- Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.
- O'Brien, P. J. (2006). Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chemical reviews*, 106(2):720–752.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526-3529.
- Roberge, P. T. (2008). The creation of pidgins as a possible window on language evolution. *LOT Occasional Series*, 10, 101-138.
- Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7), 2661-2665.
- Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, 305(5691), 1779-1782.

BIOLOGICAL SYSTEMS OF INTEREST TO RESEARCHERS OF CULTURAL EVOLUTION

LUKE MCCROHON

DMM.com

Tokyo, Japan

luke.mccrohon@gmail.com

Cultural evolution has played an important role in the evolution of language (e.g. see Kirby & Hurford, 2002; Kirby, Cornish, & Smith, 2008). This cultural process is however far less well understood than its biological equivalent, which has led to the proposal of various analogies between biological and cultural evolution (Sereno, 1991). These analogies, however, have been rightfully criticised as misleading (e.g. Smith, 2012). Despite this, we argue much insight can still be gained from the study of biology, and in this paper survey several lesser-known biological systems that are informative for the study of cultural evolution.

The first class of systems we discuss are species which undergo an *alternation of generations* between two distinct reproductive forms as part of their life cycles. Examples of such species include those of the phylum *Cnidaria* (Collins, 2002), which includes jellyfish and corals, and the parasitic fungi of the genus *Gymnosporangium* best known for causing cedar-apple rust (Petersen, 1974). A parallel is drawn with the inherently two-stage replication of cultural information from brains to the environment and then back from the environment to brains.

Second, we consider *prions*, misfolded variants of the mammalian PrP protein which can cause transmissible neurological diseases. Li, Browning, Mahal, Oelschlegel, and Weissmann (2010) have shown that, without changes in the genetic encoding of the base protein, changes in the secondary (folding) structure of prions can be selected and evolve via a darwinian process. We argue prions are therefore interesting from a cultural evolution perspective due to their replication via direct copying of form, without the separation of genotype and phenotype found in other biological systems (which is also absent in cultural transmission).

Next we introduce *clonal transmissible cancers*; infectious cancers evolved from a species' own cells (Murchison, 2008; Metzger, Reinisch, Sherry, & Goff, 2015). Both the inter-cellular selection process leading to the emergence of these cancers, as well as their subsequent evolution as pathogens (and the host response), are suggested as a model for the often proposed evolution of maladaptive cultural variants. This host-parasite relationship is briefly contrasted with the sym-

biosis between ants and certain epiphytes (plants that grow on other plants) which are known to grow structures specifically to house ant colonies (Huxley, 1980). It is suggested that, in the general case, this is likely a better model for thinking about linguistic gene-culture interactions.

And finally, we discuss the species *Oxytricha trifallax* which is claimed to have the most complex genome architecture of any known eukaryote (Chen et al., 2014). Possessing two nuclei per cell, and undergoing large-scale genome remodeling during reproduction (including deletions, rearrangements and inversions), it is argued to provide a good test ground for theories concerning the nature and necessary properties of generalised darwinian replicators.

References

- Chen, X., Bracht, J. R., Goldman, A. D., Dolzhenko, E., Clay, D. M., Swart, E. C., Perlman, D. H., Doak, T. G., Stuart, A., Amemiya, C. T., Sebra, R. P., & Landweber, L. F. (2014). The Architecture of a Scrambled Genome Reveals Massive Levels of Genomic Rearrangement during Development. *Cell*, *158*(5), 1187–1198.
- Collins, A. (2002). Phylogeny of medusozoa and the evolution of cnidarian life cycles. *Journal of Evolutionary Biology*, *15*(3), 418-432.
- Huxley, C. (1980). Symbiosis between ants and epiphytes. *Biological Reviews*, *55*(3), 321–340.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS*, *105*(31), 10681-10686.
- 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 Verlag.
- Li, J., Browning, S., Mahal, S. P., Oelschlegel, A. M., & Weissmann, C. (2010). Darwinian evolution of prions in cell culture. *Science*, *327*(5967), 869-872.
- Metzger, M., Reinisch, C., Sherry, J., & Goff, S. (2015). Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. *Cell*, *161*(2), 255 - 263.
- Murchison, E. P. (2008). Clonally transmissible cancers in dogs and tasmanian devils. *Oncogene*, *27*(S2), S19–S30.
- Petersen, R. (1974). The rust fungus life cycle. *The Botanical Review*, *40*(4), 453-513.
- Sereno, M. (1991). Four analogies between biological and cultural/linguistic evolution. *Journal of Theoretical Biology*, *151*, 467-507.
- Smith, A. (2012). Linguistic replicators: a wild goose chase? In E. C. T.C. Scott-Phillips, M. Tamariz & J. Hurford (Eds.), *The evolution of language (evolang 9)* (p. 541-542). World Scientific.

PRELIMINARY RESULTS FROM A COMPUTATIONAL MULTI AGENT MODELLING APPROACH TO STUDY HUMPBACK WHALE SONG CULTURAL TRANSMISSION

MICHAEL MCLOUGHLIN^{1*}, LUCA LAMONI^{2*}, ELLEN GARLAND², SIMON
INGRAM¹, ALEXIS KIRKE¹, MICHAEL NOAD³, LUKE RENDELL², EDUARDO
MIRANDA¹

¹*Interdisciplinary Centre for Computer Music Research / School of Marine Science and
Engineering, Plymouth University, Plymouth, United Kingdom*

²*School of Biology, University of St. Andrews, St. Andrews, United Kingdom*

³*Cetacean Ecology and Acoustics Laboratory, School of Veterinary Science, The
University of Queensland, Gatton, Australia*

**Michael McLoughlin and Luca Lamoni contributed equally to this work*

Corresponding author: Michael.mcloughlin@plymouth.ac.uk

Humpback whale (*Megaptera novaeangliae*) songs are a striking example of cultural transmission in non-humans (Garland et al., 2011). During the migration and mating season of this species, males produce complex, stereotyped sound sequences defined as ‘songs’ (Payne & McVay, 1971). Within a population, males conform to a common yet slowly evolving song. Change can also occur more rapidly when a completely new song is adopted by the entire population in a relatively short time (termed ‘*revolution*’) (Noad, Cato, Bryden, Jenner, & Jenner, 2000). These phenomena can only occur if the whales are learning song from each other. While it is possible to record the shared song within a population and how this evolves in time, the individual mechanisms and learning strategies behind the cultural transmission of song remain unknown. Furthermore, it is not clear how populations maintain conformity in songs that change over variable timescales (evolution vs. revolution). This paper presents a spatially explicit multi-agent model designed to investigate humpback whale song learning and transmission. Models with an emphasis on cultural evolution have previously been used to describe the emergence of genetic diversity in whales, and these models have been adapted to demonstrate how cultural dynamics can have the same impact on genetic diversity in humans (Whitehead,

Richerson, & Boyd, 2002). In these studies however, the exact nature of the cultural evolution is deliberately left vague. Our model seeks to extend and explore that developed in (Kirke, Miranda, Rendell & Ingram, 2015) to specifically study humpback whales songs cultural transmission and may prove to be a valuable reference point for future studies for the early evolution of human language. In detail, the model simulates both the movement and acoustic behavior of humpback whales. The migratory movement of whales between feeding and breeding grounds is enabled using flocking algorithms and movement rules that also govern the interactions among agents. Agents in the model are also equipped with a first order Markov model to generate songs (list of symbols). The transition matrix, or 'grammar', can be updated by 'learning' from other singing agents; the influence of a song on a listener agents' grammar is determined by the distance between the listener and the singer (Kirke, Miranda, Rendell, & Ingram 2015). Each agent is initialized with a randomly generated grammar. This modelling architecture enables us to study how songs are transmitted within and between populations and to record the population convergence on one or multiple song grammars. Modelling results are compared qualitatively to known song evolution patterns and specifically validated against real song data recorded in the South Pacific during the last 11 years. The model was run with varying values of spatial parameters. Namely, the size of the feeding ground, the minimum distance between agents, the size of the acoustic active space, and the size of the breeding ground(s). In total, 56 runs were implemented to explore this parameter space. Four main scenarios emerged. Firstly, in 34% of the experiments the majority of agents converged on one or multiple song grammars, depending primarily on the formation of discrete, spatially segregated groups. This result echoes what is commonly observed in the wild, where spatially segregated populations generally sing different song grammars at any given time. In the second scenario (12%), the agents' convergence was more variable compared to scenario 1 due to a combination of widely spaced breeding grounds and weak attraction between agents. Thirdly, 20% of the runs showed the highest variability in final song grammar due to strong convergence on grammars characterized by lower transition matrix probabilities. Finally, 34% of the runs showed no sign of song learning, as grammars did not converge. Across scenarios song grammars tended to decrease in size/length along each run, resulting in short and simple songs. Future work will include equipping agents with different learning strategies, a more realistic representation of humpback whale song structure and the ability to innovate song.

Acknowledgements

The authors wish to thank The Leverhulme Trust for funding this project.

References

- Garland, E. C., Goldizen, A. W., Rekdahl, M. L., Constantine, R., Garrigue, C., Hauser, N. D., Poole, M. M., Robbins, J., Noad, M. J. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Curr Biol*, *21*(8), 687-691. doi: 10.1016/j.cub.2011.03.019
- Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., & Jenner, K. C. S. (2000). Cultural revolution in whale songs. *Nature*, *408*(6812), 537-537. doi: 10.1038/35046199
- Payne, R. S., & McVay, S. (1971). Songs of Humpback Whales. *Science*, *173*(3997), 585-597. doi: 10.1126/science.173.3997.585
- Kirke, A., Miranda, E., Rendell, L., Ingram, S. (2015). Towards Modelling Humpback Whale Song Evolution using Multi-agent Systems, *Proceedings of Transdisciplinary Approaches to Cognitive Innovation (Off the Lip 2015)*, 7-11 September, Plymouth (UK)
- Whitehead, H., Richerson, P. J., & Boyd, R. (2002). Cultural Selection and Genetic Diversity in Humans, *Selection*, *3*(1), 115-125.

HUMAN-LIKE BRAIN SPECIALIZATION IN BABOONS: AN INVO ANATOMICAL MRI STUDY OF LANGUAGE AREA HOMOLOGS IN 96 SUBJECTS

ADRIEN MEGUERDITCHIAN, DAMIEN MARIE, SCOTT A. LOVE,
KONSTANTINA MARGIOTOUDI, ALICE BERTELLO, ROMAIN LACOSTE
*Laboratoire de Psychologie Cognitive, Brain and Language Research Institute, Aix-
Marseille Univ./CNRS, Marseille & Station Primatologie CNRS Rousset France*

MURIEL ROTH, BRUNO NAZARIAN, JEAN-LUC ANTON, OLIVIER COULON
Institut des Neurosciences de la Timone, Aix-Marseille Univ./CNRS, Marseille France

Language is a unique system of communication in humans and involves complex hemispheric specialization of the brain (Vigneau et al., 2006, 2011). Brain regions such as the motor cortex, Broca's area and the Planum Temporale play key-roles within the language network. Given the phylogenetic proximity between humans and nonhuman primates, the investigation of the cortical organization in apes and monkeys within a comparative approach might enable detecting the potential precursors of hemispheric specialization for language processing. Most comparative studies have focused on great apes, particularly chimpanzees (Hopkins & Cantalupo, 2008). Similarly to humans, leftward asymmetries of the planum temporale (Gannon et al., 1998; Hopkins & Nir, 2010) and rightward asymmetries of the superior temporal sulcus (Leroy et al., 2015) have been documented in chimpanzees, but not in non-hominidae Old World monkey species. The aim of the present study is to investigate the neuroanatomical asymmetries of some of these key-cortical regions for language in a non-hominidae Old World monkey species. T1-weighted anatomical images were acquired *in vivo* in 96 olive baboons (*Papio anubis*) at the Centre IRMF (Institut de Neurosciences de la Timone) from anesthetized baboons housed in social groups at the Station de Primatologie CNRS. The depths of the central sulcus (CS) following the motor cortex and of the superior temporal sulcus (STS) have been quantified in both hemispheres in each subject using semi-automatic procedures from the free software BrainVisa. For the planum temporale (PT), the surface area was manually traced on a computer in both hemispheres (Analyze 11.0 software). We

found, for the first time in a non-hominidae species, human-like significant neuroanatomical asymmetries in favor of the left hemisphere for the PT surface and in favor of the right hemisphere for the STS depth. Interestingly, inter-hemispheric asymmetries of the CS depth were significantly driven by the contralateral direction of hand preference (*i.e.*, left- or right-hand), which were previously assessed in those individuals using a bimanual coordinated task. These collective findings suggest that the continuity of hemispheric specialization between apes and humans extends to baboons for key structures of language and handedness. These findings argue that prerequisites of hemispheric specialization for language and handedness might date back not to the common ancestor of hominidae at 14-17 million years ago but to the common ancestor of Catarrhini at 30-40 million years ago.

Acknowledgements. Research funded by ANR-12-PDOC-0014_01

References

- Gannon, P. J., Holloway, R. L., Broadfield, D. C., & Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. *Science*, *279*, 220-222.
- Hopkins, W. D., & Cantalupo, C. (2008). Theoretical speculations on the evolutionary origins of hemispheric specialization. *Current Directions in Psychological Science*, *17*, 233-237.
- Hopkins, W. D. & Nir, T. M. (2010). Planum temporale surface area and grey matter asymmetries in chimpanzees (*Pan troglodytes*): the effect of handedness and comparison with findings in humans. *Behavioural Brain Research*, *208*, 436-443.
- Leroy, F., Cai, Q., Bogart, S., Dubois, J., Coulon, O., Monzalvo, K., Fischer, C., Glasel, H., Van der Haegen, L. et al. (2015). A new human-specific brain landmark: The depth asymmetry of superior temporal sulcus. *Proceedings of the National Academy of Sciences USA*, *112*, 1208-1213.
- Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage*, *30*, 1414-1432.
- Vigneau, M., Beaucousin, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B. & Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *Neuroimage*, *54*, 577-593

THE EVOLUTION OF REPAIR: EVIDENCE FROM ONLINE CONVERSATIONS

According to Dingemanse et al (2013), a baby's first word is not "Mama" or "Papa" but the single-syllable "Huh?". Already in early infancy children use "Huh?" to divert the flow of the interaction and elicit a response from their caregiver. The development of such forms of coordinated joint action is a prerequisite for the ontogeny of language (Clark, 2009) and is equally important in adult language use. Studies of utterances such as "Huh?" have shown that they constitute a large family of interactive **repair** mechanisms that are used by interlocutors to deal with problems of intersubjectivity (Schegloff, 1992). Repair consists of two main components (1) Mechanisms for initiating repair, i.e. signaling to others that there is some "trouble" (2) Mechanisms for performing the repair, i.e. resolving the problem via elaboration or reformulation. Crucially, (1) and (2) can be performed by the same or by different people. For example, the utterance "*The next shape is the green one, oops I meant the red one*" is a self-initiated, self-repair, whereas in the exchange below B initiates repair with "huh?", but the correction is performed by A.

A: *hhe next shape is the green one*

B: *huh?*

A: *oops I meant the red one*

One important dimension along which repair mechanisms differ is their ability to locate and diagnose the problem. For example, "Huh?" and "What?" do not specify the nature of the problem, whereas "when?", "where?", "who" diagnose the problem as concerning a time, place, or person. Even more specific are partial repeats such as "Partially diagnose the what?".

Studies on the emergence of referring conventions have demonstrated the importance of repair: If participants are able to repair each other's referring expressions, this leads to quicker convergence on more systematized, abstract representations (Galantucci and Garrod, 2011). This is a recurrent finding which occurs across modalities (Healey et al., 2007).

To investigate repair in closer detail, this talk presents an analysis of interactions on the social media site reddit.com. Currently 7% of all US adults use reddit, and the archive, consisting of billions of messages from 2005-2015 is freely available from archive.org. Since each message specifies whether it was edited by the user, this allows automated identification and analysis of a large subset of repair mechanisms. By examining users' conversations throughout this 10 year period, we show how the community conventionalizes its own repair mechanisms via (1) repurposing of existing mechanisms, and (2) the development of novel mechanisms. Examples of both are given below:

1. Repurposing self-repair markers as other-repair

Garcia and Jacobs, (2014) showed that people append asterisks to their turns in order to perform self-repair, e.g. "*Let's grab a beere. *beer*". Over a period of 6 years the asterisk was gradually repurposed to perform other-repair, e.g.

A: Let's go for a drink
B: *drink

2. Developing novel forms of repair-initiation

During the same period, users developed the convention of using FTFY (Fixed That For You) to perform other-repair, a mechanism that became increasingly honed for targeting the preceding turn, e.g.

A: When is It you're free? Tomorrow let's go get a pizza.
B: "get some dim sum and a beer" FTFY. It's been a while.

By tracing the development of repair in the corpus, we argue that in addition to referring conventions, interlocutors also develop community-specific routines for identifying, signaling and correcting problems in the interaction.

References:

- Clark, E. V. (2009). First language acquisition. Cambridge University Press.
- Dingemanse, M., Torreira, F., & Enfield, N. J. (2013). Is "Huh?" a universal word? PLOS One. DOI: 10.1371/journal.pone.0078273
- Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. Frontiers in human neuroscience, 5, 11.
- Healey, P. G., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games. Cognitive Science, 31(2), 285-309.
- Jacobs, J. B., & Garcia, A. C. (2013). Repair in chat room interaction. Handbook of pragmatics of computer-mediated communication, 565-588.
- Schegloff, E. A. (1992). Repair after next turn. AJS, Vol. 97, No. 5 1295-1345.

ARBITRARY HIERARCHY: A PRECEDENT FOR LANGUAGE?

DOMINIC MITCHELL

*Department of Computer Science, University of Bath
Bath, United Kingdom
D.Mitchell@bath.ac.uk*

Arbitrary and Displaced Language

It is claimed that while certain of the design features of language are found in other animals' communication, only in that of humans are all features found simultaneously (Hockett, 1960). Here, focussing on one design feature (language's arbitrary nature), evidence from a computer model suggests that the two parts of Hockett's argument may be related. The model describes pressures which sustain the use of an arbitrary value in a society but which in doing so preclude the use of the value for displaced reference (another design feature). This suggests that in addition to optimizing pre-existent features, the evolution of human language may have involved resolving conflict between features which were mutually exclusive, a hallmark of transitional change (Maynard-Smith & Szathmary, 1997)

Hierarchy Formed by Historical Asymetry

In many species individuals are organized into a transitive hierarchy. This reduces the cost of aggression in cases of competition for limited resources when the lower-ranked individual in an encounter becomes less likely to escalate conflict with the higher-ranked individual. How such hierarchies are formed is difficult to explain because models have shown that the probability of a group forming a transitive hierarchy based on individuals' Resource Holding Power (Parker, 1974) is very low. An explanation in terms of historical asymetries has been proposed (Van Doorn, Hengeveld, & Weissing, 2003). The idea is that when deciding whether to contest or concede a resource individuals are influenced by the outcomes of previous encounters. Encounters experienced or observed by the individual are used to determine this winner-loser effect. A hierarchy based on winner-loser effects is said to be arbitrary because individuals do not fully exploit their RHP but accept their place in the hierarchy based on a limited number of escalated conflicts despite the fact that this record of their conflicts may not accurately represent their RHP. An explanation for this is that the cost paid in injury resulting from the aggression required to make further discovery may be greater than the benefit that results from improvement in rank.

Model

This explanation was examined by means of individual based computer simulation. In a two dimensional world individuals move randomly in search of limited resources, if two alight on the same resource the possibility for conflict is created. The outcome of this encounter is determined by individuals playing the Hawk/Dove game (Maynard Smith, 1982) in which an individual must either contest the resource and risk aggression or cede it. Two sources of information are available to players to help decide which move to play: the opponent's RHP rank which correlates with the likelihood of winning an escalated conflict should one arise; and the opponents social rank which is assumed to be created by winner-loser interactions. A parameter determines the extent of variance between RHP rank and social rank. Individuals are able to detect that they have been placed in a position ranked lower than their RHP merits. Individuals are selected for reproduction according to fitness determined by the benefit of a higher ranking and the cost of aggression. The evolutionary outcome when individuals can improve their arbitrarily assigned position, by leveraging RHP, was tested. Within a limited range determined by the cost of aggression relative to the value of resource, and the degree of variance between RHP rank and social rank, the strategy that respects the arbitrary hierarchy is stable against invasion by the RHP strategy, beyond this range individuals that ignore the social arrangement predominate in the population and a hierarchy determined on the basis of physical characteristics is formed.

Significance for Language Evolution

It is often asked in the context of language evolution why an individual should benefit from respecting an arbitrary symbolic order when tangible evidence presents a more reliable alternative (Knight, 2008). In the current case social coordination on the arbitrary determination raises the cost of challenge to it sufficiently to render the strategy disadvantageous. Despite the fact that interactions are dyadic and that no norm is inserted exogenously into the model, the society is able to sustain an arbitrary determination, one which it pays individuals even with full knowledge, to respect. It is suggested that the same dyadic interactions, were they to occur in isolation from the social context, would result in the replacement of the arbitrary determination by the logic of individual physical superiority. Here there is a paradox: the same inertia in hierarchy which underwrote the arbitrary agreement may preclude its re-presentation, this is because displacing its use implies disruption to the current instance of hierarchy. Considered in isolation from other design features, a precedent for arbitrary language can be found, more challenging is to explain how arbitrary value and displacement can coexist given that the current model succeeds in producing one only by precluding the other.

References

- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203, 88-96.
- Knight, C. (2008). 'honest fakes' and language origins. *Journal of Consciousness Studies*, 15(10), 236.
- Maynard Smith, J. (1982). *Evolution and the theory of games*. New York: Cambridge University Press.
- Maynard-Smith, J., & Szathmary, E. (1997). *The major transitions in evolution*. New York: Oxford University Press.
- Parker, G. A. (1974). Assessment strategy and the evolution of fighting behaviour. *Journal of theoretical Biology*, 47(1), 223-243.
- Van Doorn, G. S., Hengeveld, G. M., & Weissing, F. J. (2003). The evolution of social dominance ii: multi-player models. *Behaviour*, 140(10), 1333-1358.

MAKE NEW WITH OLD: HUMAN LANGUAGE IN PHYLOGENETICALLY ANCIENT BRAIN REGIONS

MARIE MONTANT, JOHANNES C. ZIEGLER

Laboratoire de Psychologie Cognitive, Aix-Marseille University and CNRS, Marseille, France

Marie.Montant@univ-amu.fr

BENNY B. BRIESEMEISTER, TILA BRINK

Freie Universität Berlin, Allgemeine und Neurokognitive Psychologie, Center for Applied Neuroscience, Berlin, Germany

BRUNO WICKER, AURELIE PONZ

Institut de Neurosciences de la Timone, Aix Marseille University and CNRS, Marseille, France

MIREILLE BONNARD

Institut de Neurosciences des Systèmes, Aix-Marseille University and INSERM, Marseille, France

ARTHUR JACOBS

Freie Universität Berlin, Allgemeine und Neurokognitive Psychologie, Center for Applied Neuroscience, and Dahlem Institute for Neuroimaging of Emotion (D.I.N.E.), Berlin, Germany

MARIO BRAUN

Centre for Cognitive Neuroscience (CCNS), University of Salzburg, Salzburg, Austria

A common view in cognitive science is that human language consists of computations and transformations of mainly abstract symbolic information that takes place in various (mostly left) neocortical regions. According to this view, when hearing or reading emotionally loaded words, like “vomit” or “love”, the emotional content of these words is accessed through the activation of the perisylvian language network in charge of processing symbolic meaning. Alternatively, because language is one of the most recent cognitive products of human evolution, it has been argued that the language network has evolved through the re-use of already existing cortical and subcortical brain regions (Anderson, 2010). Thus, reading or hearing emotionally loaded words should therefore activate phylogenetically ancient and heteromodal brain structures that are in charge of emotions. In a functional magnetic resonance imaging (fMRI) experiment, we show that the same brain region is activated whether people observe facial expressions of disgust or whether they read words that refer to core disgust. This particular region corresponds to a portion of the insula (i.e. the anterior part) that is also known to be involved in the perception of disgusting odors (Wicker et al., 2003). In a subsequent transcranial magnetic stimulation (TMS) experiment, we show that transient disruption of the anterior insula affects the processing of core disgust words in a reading task. Participants are much slower to recognize visually presented disgust words, compared to neutral words, when TMS is applied over the anterior insula rather than the vertex. Altogether, these results are compatible with theories of embodied emotion (Niedenthal, 2007) and neural re-use (Anderson, 2010), according to which phylogenetically ancient brain structures that process basic emotions in all mammals actively participate in high-level cognitive skills, such as language.

References

Anderson ML (2010) Neural reuse: a fundamental organizational principle of the brain. *Behav Brain Sci* 33(4):245-266; discussion 266-313.

Niedenthal PM (2007) Embodying emotion. *Science* 316(5827):1002-1005.

Wicker B, et al. (2003) Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron* 40(3):655-664.

THE EFFECT OF MODALITY ON SIGNAL SPACE IN NATURAL LANGUAGES

HOPE E. MORGAN

*Linguistics Dept., University of California San Diego
San Diego, USA
hmorgan@ucsd.edu*

Natural language in the visual-gestural modality provides an opportunity to discover which aspects of sub-lexical structure are common between the two language modalities^a—spoken language and signed language—and which aspects result from emergent properties grounded in the signaling and perceptual systems. The project reported here finds that there are fundamental differences in how words are constructed in each modality, offering evidence that contrastive features in both types of phonological system are emergent, not innate. What *is* shared are more general design features, such as those proposed by Hockett (1960): i.e., *productivity*, *arbitrariness*, *discreteness*, etc. However, the findings suggest that the principle of *duality of patterning* is in need of refinement if it is to apply across modalities.

A dataset of 1,868 signs in Kenyan Sign Language (KSL) were coded for 39 formational characteristics, such as *number of hands*, *handshapes*, *palm orientation*, *finger orientation*, *type of movement*, *manner of movement*, *1st major area*, *2nd major area*, *1st minor area*, etc. During coding and analysis, minimal pairs were gathered that differed by the narrowest possible degree—i.e., by only one formational feature.

Previous researchers have mentioned that sign languages have few minimal pairs (Sandler 1996: 202; Brentari 1998: 4; Kooij 2002: 160), but a comprehensive account of phonological contrasts has not been available until now. The results of the current study finds that there are around 370 true

^a Tactile Sign Language used by deaf-blind individuals could be an emerging third natural language modality (Edwards 2014)

minimal pairs in the dataset, with only 40 signs that contrast with more than one other sign.^b Conservatively calculating from the set of recombinable elements in the language indicated by minimal pairs, there are millions of possible signs,^c a very large combinatoric space for a primarily monosyllabic language; and much larger than monosyllables in spoken languages (Kirby & Yu 2007).

Why do sign languages have such a large phonological space? A likely explanation relates to the fact that signs are not comprised of strings of segments as are words in spoken languages; and therefore cannot use sequential, syntagmatic structure to construct words and create distinct form-meaning mappings. Instead, as the present study suggests, signed words depend on a multitude of simultaneous features for boundless *productivity* in word creation.

These features are also *discrete* and have *arbitrary* properties,^d but they do not fully conform to Hockett's view of *duality of patterning*, which appears to rest on syntagmatic contrast such that segments can be reordered (e.g., 'tack', 'cat', and 'act' [Hockett 1960: 92]). Thus, some refinement of Hockett is necessary to incorporate language without strings of segments.

References

- Brentari, D. (1998). *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press.
- Edwards, T. (2014). *Language Emergence in the Seattle DeafBlind Community*. PhD Thesis. University of California Berkeley.
- Hockett, C. F. (1960). The Origin of Speech. *Scientific American*, 203(3), 88–96.
- Kirby, J. P. and A. C. L. Yu. (2007). Lexical and phonotactic effects on wordlikeness judgments in Cantonese. *Proceedings of the 16th International Congress of Phonetics Science*: 1389–1392.
- Kooij, E. van der. (2002). *Phonological Categories in Sign Language of the Netherlands*. LOT.
- Sandler, W. (1996). Phonological features and feature classes: The case of movements in sign language. *Lingua* 98: 197–220.

^b I.e., there are not large “rhyming” clusters, such as in English: ‘saw’, ‘paw’, ‘law’, ‘thaw’, ‘raw’, etc. In sign language, an example of such a cluster is signs with the same movement and handshape, but different locations. The biggest cluster in KSL contains six signs.

^c Based on the following phonemes in KSL: 33 handshapes, 29 locations, 5 absolute orientations, 3 path shapes, 4 path directions, 6 manners of movement, 6 syllable types, and 10 other misc. features

^d They also have many motivated, non-arbitrary properties—a topic beyond the scope of this paper.

LINGUISTIC STRUCTURE EMERGES IN THE CULTURAL EVOLUTION OF ARTIFICIAL SIGN LANGUAGES

YASAMIN MOTAMEDI, MARIEKE SCHOUWSTRA, KENNY SMITH, SIMON
KIRBY

*Centre for Language Evolution, University of Edinburgh
Edinburgh, UK
s0813837@sms.ed.ac.uk*

The growing body of research into homesign and emerging sign languages offers insight into languages at their earliest stages of creation and development. The study of such languages allows us to monitor the types of structures that emerge and how they develop through the first generations of a language; for example, although evidence of lexical categories in Nicaraguan Sign Language and spatial grammar in Al-Sayyid Bedouin Sign Language appear in initial generations, these structures have been shown to take time to conventionalize and become systematized (Goldin-Meadow et al, 2014; Padden et al, 2010). Furthermore, recent laboratory experiments in which hearing participants are asked to communicate using gesture can be used to test the factors that shape languages, such as cross-linguistic word order preferences (Goldin-Meadow et al., 2008; Schouwstra and de Swart, 2014), while minimizing interference from participants' native languages. Because of this, results can be compared to data from natural emerging sign languages.

We present an iterated learning study (Kirby, Griffiths and Smith, 2014) that uses the silent gesture paradigm to investigate how the use and transmission of manual communication systems drives the emergence of systematic structure.

Pairs of participants take part in an artificial language learning experiment in which they are first trained on a set of gestures and then must communicate with a partner using only gesture. In the training stage, participants are shown videos of a previous participant gesturing a concept taken from a meaning space of 24 concepts. These concepts are presented orthographically and share either a functional association (person, location, object or action) or a semantic association (based on six professions) with other items in the meaning space. For example, "hairstylist" and "hair salon" share a semantic but not a functional association, and "hairstylist" and "police officer" share a functional but not a semantic association. The first participants in each transmission chain are trained on gestures from a seed set (generation 0), where a different

individual produces a gesture for each meaning. Participants in subsequent generations are trained on gestures produced in the previous generation, for a total of five generations. In the testing stage, pairs of participants take it in turns to be director (the gesturer) and matcher (the interpreter). The director is presented with a concept from the meaning space and must communicate that concept to their partner using only gesture (presented via video streaming between computers in two separate experiment booths). The matcher then attempts to match their partner's gesture to the correct item from the meaning space, presented as a grid of lexical items.

We use both a gesture coding system as well as direct video frame analysis to produce a set of measures capturing the presence of systematic structure in the sets of gestures our participants produce. Our data show three main results concerning the structures that emerge: 1. The entropy of gesture shapes used by participants reduces over time, suggesting that participants increasingly re-use and re-combine gestures from a smaller pool of gesture shapes; 2. The gestural systems become more efficient over time as the range of movement used by participants reduces; 3. Markers for functional categories in the meaning space emerge over generations in the evolution of the gestural systems, such as a roof

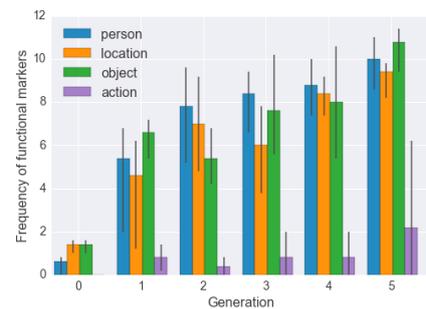


Figure 1. Mean frequency of functional markers at each generation. Coloured bars represent the four categories: person, location, object and action. Error bars represent bootstrapped 95% confidence intervals.

gesture used to signal the location category, or a point at the director's body to signal the person category (figure 1 shows the frequency of the markers used in each category across the five generations). These results suggest that, as the systems are used in communication and transmitted through generations, gestures develop from pantomimes to

conventionalized signs that demonstrate language-like segmentation through the marking of functional categories. Our results also indicate that the gestures produced by participants become more learnable as the systems are transmitted to naïve learners, and that participants in later generations become increasingly aligned with their communication partner. We suggest that the need for learnable and efficient communicative systems may drive the emergence of structure in the gestures our participants produce.

References

- Goldin-Meadow, S., Brentari, D., Coppola, M., Horton, L., & Senghas, A. (2014). Watching language grow in the manual modality: Nominals, predicates, and handshapes. *Cognition*, *136C*, 381–395.
- Goldin-Meadow, S., So, W. C., Ozyürek, A., & Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(27), 9163–8.
- Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28C*, 108–114.
- Padden, C., Meir, I., Aronoff, M., & Sandler, W. (2010). The grammar of space in two new sign languages. *Sign Languages: A Cambridge Language Survey*. Cambridge University Press, Cambridge, UK, 570–592.
- Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–6.

A SOCIAL DIMENSION OF LANGUAGE EVOLUTION

ALBERT F. H. NACCACHE

*Archaeology Department, Lebanese University, (ret.)
Beirut, Lebanon
anaccash@nidal.com*

1. Context and claim

The field of Language Evolution has not paid much attention to the social dimension of its object of study. Yet, even those who reject “language as communication” and look at language as strictly “a computational cognitive mechanism that has hierarchical syntactic structure at its core” (Bolhuis et al, 2014), cannot but agree that, ever since this mechanism started being implemented among the groups making up the genus *Homo*, the specific realizations of the basic syntactic elements have varied between these groups, and that the same must have held when syntactic rules emerged.

This is still the case today, and “the linguistic system in the traditional sense is a description of the linguistic structures used in a speech community” (Blythe & Croft, 2009, p. 48). Since it emerged, the linguistic computational cognitive mechanism has not only been embodied and encultured, but also socially embedded in “speech communities.”

In this exercise I will try to take advantage of the paleo-archaeological database, both empirical and analytical, to sketch the historical evolution of the size of “speech communities” within the *Homo* species. No claim is made here about a hypothesis or theory of language origin and evolution. All that is claimed is to highlight, within the tapestry of human evolution, one basic thread of the history of the social dimension of linguistic communication.

2. Size matters

What is attempted here is to track the historical changes of one aspect of the social dimension of language evolution, specifically, that of the size of the “speech communities” in which language has been practiced since it emerged. Although its implementation has varied throughout the history of language use, and its definition is still debated (Patrick, 2002), the speech community is a core concept in empirical linguistics, and size, the most basic of its markers, provides a realistic target for investigation over the long-time span. This is because the size of speech communities throughout the history of *Homo* can be estimated by inferences drawn from the presently available empirical databases provided by archaeology, using a set of explicit assumptions.

The first database we used is that of the estimates of the size of the global *Homo* population since its emergence ~2M years ago, estimates which are

available in the literature (Biraben, 2003; Hawks, 2008; Weiss, 1984, etc.). These estimates cover the whole historical stretch of our exploration, from the emergence of language till today, and provide us, for every period, with the sum of the populations of all the primary (first-language speakers) speech communities that existed at that time. This is uncontroversially true for all the period preceding the Neolithic, when it is fair to assume that only primary speech communities existed, even if some traders were bi or multilinguals.

The second database is the “distribution of world languages by number of first-language speakers” (Lewis, Simons & Fennig, 2015), which gives us today’s estimate and is also used as a base from which to project a Late Upper-Paleolithic estimate. The present-day distribution has a long tail consisting of a small number of languages with large numbers of speakers, languages that developed since the Neolithic under the influence of known historical and socio-cultural processes. Once these languages are subtracted from the distribution, the remaining core part and bulk of today’s distribution provides us with a good template from which to project a similarly shaped distribution that would have held 15,000 years ago, when we can confidently assume that the distribution was normal, and when the total world population was ~6M people and the largest speech communities numbered less than 5,000 people (Ong, 1977).

The third database consists in estimates of group size (Grove, 2010) and of range and/or density of occupation during the Paleolithic era. Given some conservative assumptions, these data can be used to estimate the size distribution of speech communities at the few points in the Paleolithic for which there is evidence of a change in group size and/or range or density. For instance, if we know the maximum group size and the population density at a given period, the maximum size for speech communities at that time can be inferred at the cost of some explicit assumptions. It is then straightforward to correlate this maximum size to the total population at the same period to derive the normal distribution of speech communities at that time. Available group size estimates allowed us to do this exercise at three dates in the Paleolithic. Even if the estimates are subject to a greater margin of error the further back in time we go, they are enough to plot the evolution of the average size of speech communities, a basic social dimension of language, over the history of *Homo*, from its origin till the “historical” period, for which data are much more easily derived.

The main aim of the present exercise was to produce this plot sketching the evolution of one characteristic of the human linguistic behavior throughout *Homo*’s history. As an indication of the heuristic potential of the resulting curve, just consider its main inflection points. Clearly, the inflection points must mark stages in the social development of the human linguistic communication system. But the stages are also suggestive of steps in the elaboration of the syntactic structures at the core of linguistic behavior. In my concluding remarks I will, both briefly, characterize these stages in social development and offer a potential scenario of the accretion of the syntactic structures.

References

- Biraben, J.-N. (2003). L'évolution du nombre des hommes. *Population et Sociétés*, **394**, pp. 1-4.
- Blythe, R. A. & W. A. Croft (2009). The Speech Community in Evolutionary Language Dynamics. *Language Learning*, **59**, Suppl. 1, pp. 47–63.
- Bolhuis, J., I. Tattersall, N. Chomsky & R. C. Berwick (2014). How could language have evolved? *PLOS Biology*, **12**, 8, e1001934
- Grove, M. (2010). The Archaeology of Group Size. pp. 391-411 in Dunbar, R., Gamble, C. & J. Gowlett (eds). *Social Brain, Distributed Mind*. Oxford. Oxford University Press.
- Hawks, J. (2008). From Genes to Numbers: Effective Population Sizes in Human Evolution. pp. 9-30 in Bocquet-Appel, J.-P. (ed.) *Recent Advances in Palaeodemography. Data, Techniques, Patterns*. Springer.
- Lewis, M. P., G. F. Simons & C. D. Fennig (eds.) (2015). *Ethnologue: Languages of the World, Eighteenth edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Ong, W. J. (1977). *Interfaces of the Word: studies in the evolution of consciousness and culture*. Cornell University Press.
- Patrick, P. L. (2002). The Speech Community. pp. 573-598 in J. K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *Handbook of language variation and change*. Oxford: Blackwell.
- Weiss, K. M. (1984). On the Number of Members of the Genus *Homo* Who Have Ever Lived, and Some Evolutionary Implications. *Human Biology*, **56**, 4, pp. 637-649.

SHARED BASIS FOR LANGUAGE AND MATHEMATICS REVEALED BY CROSS-DOMAIN SYNTACTIC PRIMING

TOMOYA NAKAI^{1,2}, KAZUO OKANOYA^{1,3,4}

1 Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan, 2 Japan Society for the Promotion of Science, Research Fellow, 3 Center for Evolutionary Cognitive Science, The University of Tokyo 4 Riken Brain Science Institute
cokanoya@mail.ecc.u-tokyo.ac.jp (K.O.)

It has been proposed that the capacity of recursive computation plays a central role in language evolution, and that such capacity provides a basis not only for syntax in language, but also for mathematics (Hauser et al., 2002). It is possible that an evolutionarily older linguistic function is “recycled” in novel cultural inventions such as mathematics (Dehaene & Cohen, 2007). Recently, Scheepers et al. (2011) found that the syntactic structures of prime mathematical expressions affected the processing of following target linguistic stimuli. In their subsequent study, Scheepers and Sturt (2014) revealed that such cross-domain syntactic priming effect was bidirectional.

Although these priming studies had a critical importance, their questionnaire-based experimental settings had limitations in evaluating on-line syntactic processing. In the present study, we created a new task that focused on the on-line syntactic priming between language and mathematics. We recruited 34 college students

(all native Japanese speakers, aged 18–26). Participants were asked to perform a calculation task and a semantic decision task for consecutively presented mathematical expressions and sentences (in Japanese), respectively. For both domains, we created stimuli with left-branching (“ $4*3+8$ ” and “*kuroi neko-ga hashiru*” [*a black cat runs*]) and right-branching structures (“ $8+3*4$ ” and “*neko-ga hayaku hashiru*” [*a cat runs fast*]). Linguistic stimuli were created by inserting either an adjective or adverb into the phrase composed of a NP and an intransitive verb. This experimental setting allowed us to examine the implicit structural priming effect between two domains. To consider the influence of participants’ sensitivity to structural information in mathematics, we recruited students in both scientific and non-scientific departments.

By using two-way repeated measures analysis of variance (ANOVA), we found a significant main effect of congruency (structural priming effect) only for students in scientific departments ($P < 0.05$). We found no significant main effect of modality (language to math/math to language) or interactions. Structurally congruent stimuli induced lower error rates compared to incongruent stimuli, both from language to mathematics, and from mathematics to language. Our results support the idea that language and mathematics have a shared basis in their syntactic structure, with individual variability related to the environment. We also suggest a putative application of current task settings in neuroimaging experiments.

Acknowledgements

This work is supported by a Grant-in-Aid for JSPS Fellows (No. 26-9945) from the Japan Society for the Promotion of Science.

References

- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384–398.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579.
- Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural priming across cognitive domains from simple arithmetic to relative-clause attachment. *Psychological Science*, 22(10), 1319–1326.
- Scheepers, C., & Sturt, P. (2014). Bidirectional syntactic priming across cognitive domains: From arithmetic to language and back. *The Quarterly Journal of Experimental Psychology*, 67(8), 1643–1654.

MEASURING CONVENTIONALIZATION IN THE MANUAL MODALITY

SAVITHRY NAMBOODIRIPAD^a, DANIEL LENZEN^b, RYAN LEPIC^{a,c}, TESSA VERHOEF^c

^a*Department of Linguistics*, ^b*Department of Cognitive Science*, ^c*Center for Research in Language, University of California, San Diego, San Diego, USA*
snambood@ucsd.edu

Gestures produced by users of spoken languages differ from signs produced by users of sign languages in that gestures are more typically *ad hoc* and idiosyncratic, while signs are more typically conventionalized and shared within a language community. To study how gestures may change over time as a result of the process of conventionalization, we designed a social coordination game to elicit repeated silent gestures from hearing nonsigners, and used Microsoft Kinect to unobtrusively track the movement of their bodies as they gestured (following Lenzen, 2015). Our approach follows both a tradition of lab experiments designed to study social coordination and transmission in the emergence of linguistic structure (Schouwstra et al., 2014) and insights from sign language research on language emergence. Newly emerging sign languages are being discovered, and even established sign languages are relatively young; it is therefore possible to observe linguistic conventionalization as it happens naturally (Senghas et al., 2005). Working with silent gesture, we were able to simulate and quantify effects of conventionalization that have been described for sign languages (Frishberg, 1975), including changes in efficiency of communication and size of articulatory space, in the laboratory.

Participants took turns either giving clues about (the Communicator) or guessing (the Guesser) items from a set of English nouns. Items were presented on a screen visible only to the Communicator, and once the Communicator confirmed that the Guesser had guessed correctly, the Guesser pressed a button to advance to the next item. Trial length was recorded as the time (ms) between button presses. Participants switched roles halfway through each of four rounds.

Each item appeared once per round. In Round 1, the Communicator could use gesture and speech to ensure that both participants were familiar with the entire set of items going into Rounds 2-4, which were gesture-only. The Communicator's movements were recorded as sequences of locations in XYZ-space using the Kinect. 10 pairs of undergraduates received course credit for participating in the study. Participant pairs had never met before, and no participant reported knowledge of a sign language.

We examined Rounds 2-4, in which the Communicator was gesturing without speech about a set of items known to both participants. We observed rapid alignment between participant pairs across the rounds; as participants became familiar with the items in the game, they correctly guessed the items at faster rates. Trial lengths (s) started longer in Round 2 (M=11.54), and became shorter in Round 3 (M=5.65) and Round 4 (M=4.39). A linear mixed-effects model showed that ROUND significantly affected TRIAL LENGTH ($\chi^2=87.09$, $p < 0.0001$), reducing trial length by about 3.57s (S.E. +/-0.37) each round.

Two additional analyses concerned volume of gesture space and distance traveled by the hands. The Kinect measurements showed that gesture spaces started larger (m³) in Round 2 (M=0.15) and became smaller in Round 3 (M=0.11) and Round 4 (M=0.10). A linear mixed-effects model showed that ROUND significantly affected GESTURE SPACE ($\chi^2=51.01$, $p < 0.0001$), reducing the volume of the gesture space by about 0.03m³ (S.E. +/- 0.004) each round. The total distance traveled by the hands also started longer (m) in Round 2 (M=11.58) and became shorter in Round 3 (M=6.22) and Round 4 (M=5.14). A linear mixed-effects model showed that ROUND significantly affected HAND TRAVEL DISTANCE ($\chi^2=75.85$, $p < 0.0001$), reducing the distance that the hands traveled by about 3.22m (S.E. +/- 0.36) each round. (Figures associated with trial length, volume of gesture space, and distance traveled by the hand are included in the supplementary materials.)

We chose an experimental setup known to result in rapid conventionalization (Scott-Phillips & Kirby, 2010), and with Kinect we were able to measure changes in gesture that are also hallmarks of conventionalization in sign language. This approach opens the door for more direct future comparisons between *ad hoc* gestures produced in the lab and natural sign languages in the world. By operationalizing concepts like reduction and articulatory space, which, out of necessity have been typically discussed in vague terms, we anticipate that this approach will also be beneficial for future studies of (sign) language emergence.

Acknowledgments

We wish to thank Carol Padden, Jim Hollan, Jeremy Karnowski, and Nick Gibson. This work was supported in part by the UC San Diego Chancellor's Interdisciplinary Collaboratories Fellowship to DL, RL, and SN, and the NWO Rubicon Fellowship to TV.

References

- Frishberg, N. (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language*, 51(3), 696-719.
- Lenzen, D. (2015). Measuring Time Gestures with the Microsoft Kinect. In Noelle, D.C. et al. (Eds.), *Proc. of the 37th Annual Meeting of the Cognitive Science Society* (pp. 1285-1289). Austin, TX: Cognitive Science Society.
- Senghas, R.J., Senghas, A. & Pyers, J.E. (2005). The emergence of Nicaraguan Sign Language: Questions of development, acquisition, and evolution. In Parker, S.T. et al. (Eds.), *Biology and Knowledge Revisited: From Neurogenesis to Psychogenesis* (pp. 287-306). New York: Psychology Press.
- Schouwstra, M., Abramova, K., Mota Medi, Y., Smith, K. & Kirby, S. (2014) From silent gesture to artificial sign languages. *Workshop on Signals, Speech and Signs*, Vienna, 35-39
- Scott-Phillips, T., Kirby, S. (2010). Language Evolution in the Laboratory. *Trends in Cognitive Sciences*, 14(9), 411 - 417

THE ARBITRARINESS OF THE SIGN REVISITED: THE ROLE OF PHONOLOGICAL SIMILARITY

ANONYMOUS AUTHORS 1, 3, 4

*University Department, University Name,
City, Country
email@university*

ANONYMOUS AUTHORS 1, 3, 4

*University Department, University Name,
City, Country
email@university*

Recent research has suggested that the structure of the lexicon bears the hallmarks of an adaptation to support language learning (Monaghan et al., 2014). It has been suggested that systematically structured languages (i.e. where some feature of meanings is related to a feature of words) might aid in bootstrapping language acquisition- thus, explorations of how different types of systematic structure affect learnability might answer important questions and generate further testable predictions about the origins of language. In 2011, Monaghan, Christiansen, & Fitneva reported the results of a series of experiments and computational models of language learning that were designed to test the effect of systematicity on learning. In their study they used a feed-forward neural network model and an artificial language learning paradigm with human participants to explore the differences in learnability between languages where the relationships between forms and meanings were either systematic or arbitrary (i.e. where no feature of meaning is reliably associated with any feature of words).

In Monaghan et al.'s study, systematic associations between words and meanings are based on there being phonological similarities *within* a group of words (e.g. the fricative phonemes /f/ and /z/ being associated with similar meanings), and also phonological dissimilarity *between* groups (e.g. the plosive phonemes /g/ and /k/ being associated with a second group of meanings) Here, we extend the findings of Monaghan et al. (2011) using a new experimental methodology and a

number of computational simulations. In addition to systematic associations between words and meanings that are based on phonological similarity, we explore the learnability of systematic languages that are phonologically dispersed.

We replicated the model described by Monaghan et al. (2011), instantiating a version using a 2x2 design with systematicity (arbitrary vs. systematic) as one factor and phonology (clustered vs. dispersed) as a second factor. In the clustered condition of the simulation (which directly replicates Monaghan et al.), labels with similar phonemes (e.g. *f* and *ʒ*) were used to create one set of labels, with a set of dissimilar phonemes (e.g. *g* and *k*) used in a second set of labels. In the newly added dispersed conditions the coupling of phonemes based on their featural similarity was broken (pairing, for example, *f* and *g*).

Where Monaghan et al. (2011) contrasted fricative and plosive phonemes, our experiment used a set of phonemes that differed in plosivity (plosive vs. continuant consonants) as in Nielsen & Rendall, 2012. Additionally, our experiment moved from an alternative forced choice task to a signal detection protocol: after training, participants were presented with trials where they were shown a single image with a single label and tasked with responding whether the pairing was one that they had been trained on before. As with the model, the experiment was a 2 (systematic vs. arbitrary) x 2 (phonological vs. dispersed) design.

Our results suggest that human language learners learn systematic languages better than arbitrary ones, regardless of their degree of phonological dispersion. This stands in contrast to the results of the model, which overestimates the importance of phonological dispersion for learning- confusing similar phonemes at higher rates than do human participants. These results suggest that the types of systematic structures we might expect to see in real languages might not always be neatly phonologically clustered, but that systematic structure in its most general form is adaptive for the process of language learning.

Monaghan, P., Christiansen, M.H., & Fitneva, S.A. (2011). The Arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, 140, 325-347.

Monaghan, P., Shillcock, R.C., Christiansen, M.H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B*, 369.

Nielsen, A. & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, 4, 115-125.

DOMESTICATION AND EVOLUTION OF SIGNAL COMPLEXITY IN FINCHES

KAZUO OKANOYA

*Department of Cognitive and Behavioral Sciences, The University of Tokyo
Tokyo, Japan
cokanoya@mail.ecc.u-tokyo.ac.jp*

Among vocalizations birds make, a class of sounds that are consisted of more than two types of sound patterns arranged in a certain temporal sequence is called as a 'birdsong', not only because of the organization of sound patterns, but also because our musical aesthetics intuitively allow such an analogy. Scientific investigations of birdsong to date suggest that certain properties of birdsong extend beyond the musical to the developmental analogies.

Bengalese finches are domesticated strains of wild white-rumped munias imported from China to Japan 250 years ago. Bengalese finch songs are composed of multiple chunks and each chunk is a combination of 2-4 song notes. Furthermore, chunks are arranged in a finite-state probabilistic automaton. We studied how and why Bengalese finches sing such complex songs. We found the following facts. 1) The ancestral strain sing simpler songs. 2) There is high learning specificity in white-rumped munias but not in Bengalese finches. 3) Bengalese finches have larger song control nuclei and higher level of glutamate receptor gene expressions than white-rumped munias. 4) Both Bengalese finch and white-rumped munia females prefer complex songs as measured by the nest string assay and males with complex songs are physically fitter than the males with simpler songs. These results promoted sexual selection scenario of song complexity in Bengalese finches (Okanoya, 2004).

We further examined factors related with domestication. We examined songs of white-rumped munias in subpopulations of Taiwan (Kagawa, et al., 2012). Where there is a sympatric species to white-rumped munias, songs were simpler. This leads to a hypothesis that in the wild songs needed to be simple to secure species identification, but under domestication this constrains was set free. Not only that, analyses of isolated songs and cross-fostering results suggest that there are different degrees of learnability between white-rumped munias and Bengalese finches (Takahasi, et al, 2010; Kagawa et al, 2014).

Furthermore, recent suggestion of neural crest hypothesis that might account for the “domestication syndrome” fits well with the properties of Bengalese finches (Wilkins, et al., 2014). For example, Bengalese finches are sooner to recover from tonic immobility test than white-rumped munias (Suzuki, et al., 2013). Feces corticosterone level is lower in Bengalese finches (Suzuki, et al., 2014). Biting force is stronger in white-rumped munias than Bengalese finches, and time required to come back to the food cup after a foreign object was placed was quicker in Bengalese finches (Suzuki et al., in prep; Ikebuchi et al., in prep). All of these result suggest that the difference between Bengalese finches and white-rumped munias in socio-emotional factors might be related with the limited diffusion of the neural crest cells, since these properties are controlled by cells derived from the neural crest cells (Wilkins, et al., 2014).

Thus, evolution of song complexity involves not only factors related with strengthen of sexual selection and relaxation of species identification, but also socio-emotional factors due to domestication. These results on Bengalese finches must be useful in discussing possible biological origin of human speech in terms of proximate and ultimate factors. (Work supported by Kakenhi #15K14581)

References

- Kagawa, H., Yamada, H., Lin, R. S., Mizuta, T., Hasegawa, T., & Okanoya, K. (2012). Ecological correlates of song complexity in white-rumped munias: the implication of relaxation of selection as a cause for signal variation in birdsong. *Interaction Studies*, 13(2), 263-284.
- Kagawa, H., Suzuki, K., Takahasi, M., & Okanoya, K. (2014). Domestication changes innate constraints for birdsong learning. *Behavioural processes*, 106, 91-97.
- Okanoya, K. (2004). Song syntax in Bengalese finches: proximate and ultimate analyses. *Advances in the Study of Behavior*, 34, 297-346.
- Suzuki, K., Matsunaga, E., Yamada, H., Kobayashi, T., & Okanoya, K. (2014). Complex song development and stress hormone levels in the Bengalese finch. *Avian Biology Research*, 7(1), 10-17.
- Suzuki, K., Ikebuchi, M., & Okanoya, K. (2013). The impact of domestication on fearfulness: a comparison of tonic immobility reactions in wild and domesticated finches. *Behavioural processes*, 100, 58-63.
- Takahasi, M., & Okanoya, K. (2010). Song learning in wild and domesticated strains of white - rumped munia, *Lonchura striata*, compared by cross - fostering procedures: domestication increases song variability by decreasing strain - specific bias. *Ethology*, 116(5), 396-405.
- Wilkins, A. S., Wrangham, R. W., & Fitch, W. T. (2014). The “domestication syndrome” in mammals: A unified explanation based on neural crest cell behavior and genetics. *Genetics*, 197(3), 795-808.

PARROT “PHONOLOGICAL REGRESSION”: EXPANDING OUR UNDERSTANDING OF THE EVOLUTION OF VOCAL LEARNING

IRENE M. PEPPERBERG

*Department of Psychology, Harvard University,
Cambridge, MA USA
impepper@wjh.harvard.edu*

KATIA ZILBER-IZHAR; SCOTT SMITH

*Department of Psychology, Harvard University
katia.izhar@gmail.com, scottsmith@g.harvard.edu*

Development of vocal learning was a critical aspect in the evolution of spoken language, mainly because such learning allows for cultural transmission of the complex communication patterns that are the hallmark of human language. Notably, few nonhumans engage in such behavior. Oscine songbirds are the most commonly studied of these exceptional nonhumans, and research demonstrates striking avian-human parallels with respect to the ontogeny and neurological bases of vocal communication (e.g., Jarvis et al. 2005). Parrots also engage in vocal learning, but, unlike most songbirds, are adept at vocal mimicry (Chakraborty et al., 2015)—the capacity to reproduce, exactly, sounds such as those of human speech. Given that Grey parrots (*Psittacus erithacus*), at least, have shown some ability for referential use of such speech and advanced cognitive capacities (e.g., Pepperberg, 1999; Pepperberg & Carey, 2012), they could provide a particularly good model for studying how vocal communication may have evolved (e.g., Pepperberg, 2013).

One aspect of human language, with likely evolutionary importance, that of pre-speech babbling, has been studied extensively in both children and songbirds (e.g., Doupe & Kuhl, 1999) and somewhat in parrots. In parrots, however, researchers either investigated development of conspecific vocalizations in the wild (Berg et al., 2011) or the vocalizations of an adult bird, already fluent in human speech, that was learning novel labels (Pepperberg, Brese, & Harris, 1991). The development of human speech in a juvenile parrot, however, had not been tracked.

A recent study of such tracking found an interesting aspect of vocal learning, that of “phonological regression”, also seen in children, even if rarely reported (e.g., Bleile & Tomblin, 1991): Here, young children sometimes nearly perfectly produce words at a very early stage, but these correct first productions are then followed by less faithful renditions, only to be returned later to relative accuracy. Fledgling

songbirds may similarly occasionally countersing with adults using a fully adult rendition, then return to subsong before fully developing their vocalizations (e.g., Baptista, 1983). The present study examined the trajectory of vocal development of a young Grey parrot (Athena) as she learned referential English. By tracking Athena's acquisition of vowel-like sounds over the course of fifteen months, using audio recordings and acoustic software programs, her vocal development was analyzed over time, from her first squeaks to her more distinct pronunciations, and her progress compared with human children and other parrots in the lab. Not one, but multiple U-shaped curves characterized her acquisition of isolated labels, from what initially seemed to be almost exact renditions of an English label, to much less clear versions, and on to more faithful copies. The results indicate that, like human children, parrots can experience the phenomenon of phonological regression, a finding which provides additional evidence for avian-human vocal learning parallels.

References

- Baptista, L.F. (1983). Song learning. In A.H. Brush & G.A. Clark Jr (Eds), *Perspectives in ornithology* (pp 500-506). Cambridge: Cambridge University Press.
- Berg, K.S., Delgado, S., Cortopassi, K.A., Bessinger, S.R., & Bradbury, J.W. Vertical transmission of learned signature calls in a wild parrot. *Proceedings of the Royal Society B*: doi:10.1098/rspb.2011.0932.
- Bleile, K., & Tomblin, K. (1991). Regressions in the phonological development of two children. *Journal of Psychological Research*, 20(6), 483-99.
- Chakraborty, M., Walløe, S., Nedergaard, S., Fridel, E.E., Dabelsteen, T., et. al., (2015). Core and shell song systems unique to the parrot brain. *PLoS ONE*, DOI:10.1371/journal.pone.0118496
- Jarvis, J.D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W. et al. (2005). Avian brains and a new understanding of vertebrate evolution. *Nature Reviews Neuroscience*, 6, 151-159.
- Doupe, A.J., & Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, 22, 567-631.
- Pepperberg, I.M. (1999). *The Alex studies*. Cambridge, MA: Harvard University Press.
- Pepperberg, I.M. (2013). Evolution of vocal communication: An avian model. In J. Bolhuis & M. Everaert (Eds), *Birdsong, speech and language* (Ch. 26) Cambridge, MA: MIT Press
- Pepperberg, I.M., Brese, K.J., & Harris, B.J. (1991). Solitary sound play during acquisition of English vocalizations by an African Grey parrot (*Psittacus erithacus*): Possible parallels with children's monologue speech. *Applied Psycholinguistics*, 12:151-177.
- Pepperberg, I.M. & Carey, S. (2012). Grey Parrot number acquisition: the inference of cardinal value from ordinal position on the numeral list, *Cognition* 125:219-232.

EARLY LEARNED WORDS ARE MORE ICONIC

LYNN K. PERRY

*Department of Psychology, University of Miami,
Coral Gables, FL, USA
lkperry@miami.edu*

MARCUS PERLMAN, GARY LUPYAN

*Department of Psychology, University of Wisconsin-Madison,
Madison, WI, USA
mperلمان@wisc.edu, lupyman@wisc.edu*

BODO WINTER

*Department of Cognitive Science, University of California Merced,
Merced, CA, USA
bodo@bodowinter.com*

DOMINIC W. MASSARO

*Department of Psychology, University of California Santa Cruz,
Santa Cruz, CA, USA
massaro@us.edu*

Laboratory studies show that iconicity, the correspondence between form and meaning, can help young children to learn and generalize the meanings of new words (Imai, Kita, Nagumo, & Okada, 2008). These findings suggest that iconicity may serve an important function in language acquisition by jumpstarting early word learning (Imai & Kita, 2014). This mechanism has important implications for language evolution: Because languages evolve to be learnable by children (Christiansen & Chater, 2008), if iconicity is an aid to word-learning, early-learned words should be more iconic.

Recently, Perry et al. (2015) asked speakers of English and Spanish to rate the iconicity of ~600 words in their respective language (see also a study with British Sign Language: Vinson, Cormier, Denmark, Schembri, & Vigliocco, 2008). They found that iconicity predicted age of acquisition (AoA), even after controlling for a number of related factors. In the current paper, we ask *why* such

a relationship might exist and what consequences it has for our understanding of language evolution.

We asked English-speaking participants to rate the iconicity of 1,952 English words. 705 participants rated 25-26 words each. These iconicity ratings predict AoA: words rated as more iconic are acquired earlier even when we control for word frequency and systematicity (from Monaghan, Shillcock, Christiansen, & Kirby, 2014). Systematicity measures language-internal regularities between words' forms and meanings, which is in theory, independent of iconicity. In support of this distinction, we found that iconicity is not correlated with systematicity, and iconicity explains more variance in AoA than systematicity.

Next, we compared children's production frequencies (from CHILDES) to adult's production frequencies (from SUBTLEX), yielding a measure of how much a word is used by children compared to adults (e.g., "spoon" is used more by children; "before" more by adults). More iconic words were more likely to be used by children (above and beyond systematicity). In a related finding, words increasing in frequency over development were relatively *less* iconic, while those decreasing in frequency were relatively *more* iconic, showing that more mature vocabularies become increasingly dominated by arbitrary words.

Our findings suggest iconicity is not distributed equally across the lexicon but specifically characterizes early-learned words. Given the experimental evidence that iconicity aids word learning, this is exactly the distribution we would expect in a lexicon that is adapted to fit the cognitive systems of its users.

References

- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(05), 489–509.
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130298.
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65.
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is English? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130299.
- Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, *10*(9), e0137147.
- Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods*, *40*(4), 1079–1087.

COOPERATIVE COMMUNICATION: WHAT DO NONHUMAN ANIMALS HAVE TO TELL?

SIMONE PIKA

*Max Planck Institute for Ornithology, Humboldt Research Group,
Seewiesen, Germany
spika@orn.mpg.de*

Human language is thought to be a fundamentally cooperative enterprise, involving fast-paced and extended social interactions (Grice, 1957; Sperber & Wilson, 1986). Although it is still highly debated how human language originated, it has been suggested that it evolved as part of a larger adaptation of humans' species-unique forms of cooperation (Levinson, 1995; Tomasello, 2008). Earliest cooperative interactions can be observed around the age of 12 months, when human infants start to engage in turn-taking routines with their caretakers involving distinct gestures such as *showing*, *offering*, *giving* (for example, food, objects), and *pointing* to coordinate attention towards a social partner and an object of mutual interest. Intriguingly, Levinson and Holler (Levinson & Holler, 2014) thus suggested that the apparent gulf between animal and human communication may be bridged by looking for precursor adaptations to human language in turn-taking interactions and gestural signalling. Although, our closest living relatives, bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*) communicate via highly sophisticated gestural exchanges (e.g. Fröhlich, Wittig, & Pika, in press; Pika, Liebal, & Tomasello, 2005) and show cooperative abilities under experimental (Hare, Melis, Woods, Hastings, & Wrangham, 2007; Melis, Hare, & Tomasello, 2006; Pika & Zuberbühler, 2008) and/or wild conditions (Boesch & Boesch-Achermann, 2000; Mitani, 2009), studies into cooperative communication skills are relatively rare. For instance, by drawing on a conversation analysis framework Rossano (2013) showed that two mother infant dyads of captive bonobos used gesture sequences that strongly resemble the structure of turn-taking sequences of social action in human conversation. They utilized cooperative adjacency-pair structures and communicated at communication tempi similar to the timing of ordinary human conversation (Stivers et al., 2009).

The aim of the present paper is twofold: First, I aim to revisit the claim that communicative interactions of nonhuman animals lack the cooperative nature of human communication. Second, I try to encourage a critical evaluation of methods commonly used in the field of gesture research to draw inferences about similarities between human and nonhuman animal signaling to enable a higher sensitivity to the social characteristics and/or ecology of a given species (see for recent developments in other areas of cognitive ethology, Hare, 2001). To do so, I will (a) provide an overview of the state of the art, (b) present newest

data on collaborative communication in nonhuman animals, and (c) develop a framework, which could be used to predict patterns of collaborative communication in nonhuman animal species and to facilitate more systematic investigation.

References

- Boesch, C., & Boesch-Achermann, H. (2000). *The Chimpanzees of the Tai Forest: Behavioural Ecology and Evolution*. Oxford: Oxford University Press.
- Fröhlich, M., Wittig, R. M., & Pika, S. (in press). 'Should I stay or should I go?' – Initiation of joint travel in mother-infant dyads of two chimpanzee communities in the wild. *Animal Cognition*.
- Grice, H. P. (1957). Meaning. *Philosophical Review*, 66(3), 377-388.
- Hare, B. (2001). Can competitive paradigms increase the validity of experiments on primate social cognition? *Animal Cognition*, 4, 269-280.
- Hare, B., Melis, A. P., Woods, V., Hastings, S., & Wrangham, R. (2007). Tolerance Allows Bonobos to Outperform Chimpanzees on a Cooperative Task. *Current Biology*, 17(7), 619-623.
- Levinson, S. C. (1995). Interactional biases in human thinking *Social intelligence and interaction* (pp. 221-260): Cambridge University Press.
- Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1651), 20130302.
- Melis, A. P., Hare, B., & Tomasello, M. (2006). Chimpanzees recruit the best collaborators. *Science*, 311, 1297-1301.
- Mitani, J. C. (2009). Cooperation and competition in chimpanzees: Current understanding and future challenges. *Evolutionary Anthropology*, 18(5), 215-227.
- 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.
- Pika, S., & Zuberbühler, K. (2008). Social games between bonobos and humans: Evidence for shared intentionality? *American Journal of Primatology*, 70(3), 207-210.
- Rossano, F. (2013). Sequence organization and timing of bonobo mother-infant interactions. *Interaction Studies*, 14(2), 160-189. doi: 10.1075/is.14.2.02ros
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Cambridge, Massachusetts: Harvard University Press.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., & Heinemann, T. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 106(26), 10587-10592.

Tomasello, M. (2008). *Origins of Human Communications*. Cambridge, Massachusetts: MIT Press.

CONSTRUCTION GRAMMAR FOR APES

MICHAEL PLEYER

*Department of English, Universität Heidelberg, Kettengasse 12, Heidelberg, D-69117,
Germany, michael.pleyer@hggs.uni-heidelberg.de*

STEFAN HARTMANN

*German Department, University of Mainz, Jakob-Welder-Weg 18, Mainz, D-55099,
Germany, stefan.hartmann@uni-mainz.de*

Constructionist approaches describe language as a structured network of form-meaning pairings. These pairings vary in their degree of schematicity and prototypicality, ranging from lexical items to highly abstract syntactic patterns. Language acquisition is seen as based on general social and cognitive skills. Starting out from concrete, item-based constructions, children use these skills to extract and gradually abstract constructions from instances of actual language use (Tomasello 2006). Constructions are stored in a fine-grained taxonomic network, the so-called constructicon.

Constructionist approaches have been increasingly applied to language evolution research (e.g. Hurford 2012). In line with this growing research movement, we propose that constructionist approaches can prove useful in elucidating similarities and differences between human language and non-human primate communication systems. Specifically, we will discuss the question whether the nature of great ape gesture systems can be captured in terms of an inventory of (proto-)constructions – a proto-constructicon – and whether such a network is based on cognitive capacities homologous to the cognitive infrastructure underlying the acquisition, usage, and processing of constructions in humans.

Regarding the gesture systems of chimpanzees, Roberts et al. (2012: 586-587) note that they “have a multifaceted and complex repertoire of manual gestures, organised around prototypes, within which there is considerable variation.” Schematization and prototypicality can therefore be seen as important foundational features both of great ape gesture systems and of the human constructicon. In a usage-based, constructionist approach, linguistic knowledge

is seen to consist in abstractions from exemplar representations of experience with concrete usage events in context that form radial prototype networks (cf. e.g. Croft 2001). Importantly, Roberts et al. (2012: 587) note that there are gestures that are “intermediate between the prototypical forms” and that are not structurally discrete but instead graded. Similarly, usage-based accounts of language acquisition assume that knowledge of linguistic constructions in young children is characterised by fuzzy boundaries and graded representations (e.g. Abbot-Smith, Lieven & Tomasello 2008).

Another important point of comparison concerns the role of pragmatics in human and non-human primate communication. In studies of the gesture systems of great apes it was found that they flexibly use multiple different gestures in the same context for the same goal. They also use single gestures in different contexts with different goals (Liebal et al. 2014: 155). As Genty & Zuberbühler (2015) note, “several gestures appear to have several outcomes, suggesting that meaning resides more in the pragmatic context than in the morphological form of the signal,” although there are also some iconic and deictic gestures. Human linguistic constructions, in contrast, possess more specific conceptual content. Still, the meaning side of human linguistic constructions is characterised by the properties of prototypicality and schematicity. As such the meaning of human linguistic constructions is underdetermined without pragmatic context and is only properly instantiated in actual language use in particular situations (cf. also Scott-Phillips 2015).

Despite the marked differences between human constructions and great ape gestures, we propose that the striking similarities both can be analysed in a shared theoretical format. As a data-driven and bottom-up approach to signaling systems, Construction Grammar seems like an appropriate heuristic tool for that purpose.

References

- Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). Graded representations in the acquisition of English and German transitive constructions. *Cognitive Development, 23*, 48-66.
- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Hurford, J.R. (2012). *The Origins of Grammar: Language in the Light of Evolution II*. Oxford: Oxford University Press.

- Genty, E. & Zuberbühler, K. (2015). Iconic gesturing in bonobos. *Communicative & Integrative Biology*, 8(1), e992742
- Liebal, K., Waller, B.M., Burrows, A.M. & Slocombe, K.E. (2014): Primate Communication: A Multimodal Approach. Cambridge: CUP.
- Roberts, A.I. , Vick, S, Roberts, S.G.B., Buchanan-Smith, H.M. & Zuberbühler, K. (2012). A structure-based repertoire of manual gestures in wild chimpanzees: statistical analyses of a graded communication system. *Evolution and Human Behaviour*, 33, 578-589.
- Scott-Phillips, T. (2015). *Speaking Our Minds: Why Human Communication is Different, and How Language Evolved to Make It Special*. London: Pelgrave.
- Tomasello, M. (2006). Construction grammar for kids. *Constructions*, SV1-11/2006.

THE CULTURAL EVOLUTION OF STRUCTURE IN MUSIC AND LANGUAGE

ANDREA RAVIGNANI

*Artificial Intelligence Lab, Vrije Universiteit Brussel,
Brussels, Belgium*

*Sensory and Cognitive Ecology Group, Universitaet Rostock,
Rostock, Germany
andrea.ravignani@gmail.com*

TANIA DELGADO, SIMON KIRBY

*School of Philosophy, Psychology & Language Sciences, University of Edinburgh,
Edinburgh, UK
tdelgado09@gmail.com, smkirby@gmail.com*

Introduction. Humans are well-versed at perceiving and learning sequences. If confronted with language or other culturally-transmitted systems in the lab, humans introduce and amplify structural regularities making the systems easier to learn (Kirby, Griffiths, & Smith, 2014; Verhoef, Kirby, & de Boer, 2014). In particular, *systematicity* frequently evolves in these experiments. Individual items are easier to reproduce by virtue of patterns of similarity across the entire *set* of items that are learned (Cornish, Smith & Kirby 2013). Is this process of system formation language-, task- or domain-specific (Iversen, Patel, Nicodemus, & Emmorey, 2015; Patel, 2007)? Can cultural transmission explain universals in musical structure as it explains linguistic universals (Jackendoff, 2009; Trehub, 2015)?

Methods. We tested 48 participants in a non-linguistic iterated learning task, featuring transmission of information across generations of learners. Participants were asked to imitate sets of drumming sequences as accurately as possible using an electronic drum kit. Following Cornish et al (2013), the task involved immediate imitation of each sequence in the set, rather than exposure to the full set of items prior to production. While the first generations of participants had to reproduce drumming patterns with random inter-beat intervals (i.e. time between

onsets of adjacent drum hits), later generations were asked to copy the imperfectly recalled patterns of the previous generation of participants.

Results. First, similarly to other experiments using language-like stimuli, systematic structure emerged over time. Over experimental generations and within transmission chains, drumming sequences became: (i) more rhythmic; (ii) easier to learn, shown by measuring imitation fidelity; (iii) more systematic: i.e. the structure of each sequence in a recalled set provides expectations about the structure of other sequences. Second, the evolution of structure in this duration-based task strikingly resembled independent results in a visual sequencing task (Cornish et al., 2013), suggesting that either one domain-general or two analogous domain-specific mechanisms underpin similar pressures for systematic structure across domains. Third, our experimental transmission chains recreated rhythmic features which are statistical universals of world music (Savage, Brown, Sakai, & Currie, 2015). In fact, over generations, drumming sequences became: (i) more isochronous, (ii) composed of few (categorically distributed) alternating inter-beat intervals, related by small integer ratios, and (iii) more structured, containing repeating motivic patterns.

Conclusions. The emergence of systematic structure via cultural transmission: (a) does not require semantics or learning language-like behaviours; (b) operates similarly across domains and modalities of human cognition; (c) explains characteristics of musical rhythms appearing as statistical universals around the world (Savage et al., 2015; London, 2012). Future research should replicate this experiment in participants of different age groups, cultures and literacy in order to disentangle group-specific from human-universal cognitive biases (Ravignani, 2015; Trehub, 2015).

Acknowledgements

Andrea Ravignani was supported by FWO grant V439315N and KWA Universitaet Wien grant 000245 (to Andrea Ravignani), and European Research Council grants 283435 ABACUS (to Bart de Boer) and 230604 SOMACCA (to W. Tecumseh Fitch). We thank Piera Filippi, Bill Thompson, Bart de Boer, Hannah Little, Sabine van der Ham, Kenny Smith, and all members of the LEC Edinburgh for comments and advice.

References

- Cornish, H., Smith, K., & Kirby, S. (2013). Systems from sequences: An iterated learning account of the emergence of systematic structure in a non-linguistic task. *Proc. 35th Conference of the Cognitive Science Soc* 340-345.
- Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition*, 134, 232-244.
- Jackendoff, R. (2009). Parallels and Nonparallels between Language and Music. *Music Perception*, 26(3), 195-204.
- Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current opinion in neurobiology*, 28, 108-114.
- London, J. (2012) Three Things Linguists Need to Know About Rhythm and Time in Music. *Empirical Musicology Review*, 7(1-2), 5-11.
- Patel, A. D. (2007) *Music, language, and the brain*. Oxford university press.
- Ravignani, A. (2015). Evolving perceptual biases for antisynchrony: a form of temporal coordination beyond synchrony. *Frontiers in neuroscience*, 9.
- Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, 112(29), 8987-8992.
- Tamariz, M., Brown, J. E., & Murray, K. M. (2010). The role of practice and literacy in the evolution of linguistic structure. In *Proceedings of the 8th International Conference on the Evolution of Language* (pp. 313-320).
- Trehub, S. E. (2015). Cross-cultural convergence of musical features. *Proceedings of the National Academy of Sciences*, 112(29), 8809-8810.
- Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43, 57-68.

STRATEGIES IN GESTURE AND SIGN FOR DEMOTING AN AGENT: EFFECTS OF LANGUAGE COMMUNITY AND INPUT

LILIA RISSMAN, LAURA HORTON, MOLLY FLAHERTY, DIANE BRENTARI,
SUSAN GOLDIN-MEADOW

*Department of Psychology, University of Chicago
Chicago, USA*

*lrissman@uchicago.edu, laurahorton@uchicago.edu, mflaherty@uchicago.edu,
dbrentari@uchicago.edu, sgm@uchicago.edu*

ANN SENGHAS

*Department of Psychology, Barnard College of Columbia University
New York, USA*

asenghas@barnard.edu

MARIE COPPOLA

*Department of Psychological Sciences, University of Connecticut
Storrs, USA*

marie.coppola@uconn.edu

Languages use a variety of devices to indicate that an agent is present in an event but not particularly salient (see Siewierska, 2013 for a typology of passive voice and other agent demotion devices). Here we investigate agent-demotion in an emerging sign language in Nicaragua. Nicaraguan Sign Language (NSL) began when Homesigners (deaf individuals who use homemade gestures to communicate with hearing individuals) were brought together for the first time in the late 1970s (Cohort 1 signers). Cohort 2 signers entered the community after 1984, and Cohort 3 signers joined after 1994; these later cohorts learned their sign language from the previous generations. We asked Homesigners, Cohort 1 signers, and Cohort 2-3 signers to describe vignettes that varied in how salient the agent was. All groups used verbal morphology to distinguish agentive vs. non-agentive scenes. However, only signers who learned their language from older peers (i.e., Cohorts 2-3) used verbal morphology for agent demotion. This finding suggests that linguistic devices for demoting agents evolve more slowly than devices for distinguishing agents from non-agents, and that the former may emerge only when language is transmitted to a subsequent generation of learners.

In many sign languages around the world, agency is encoded via handshape morphemes in classifier predicates (Benedicto & Brentari, 2004). For example, *handling* handshapes (e.g., the hand represents how the pen is held as it is lifted off a table) are used in predicates that describe agentive events. By contrast, *object* handshapes (e.g. an extended index finger represents the pen itself as it rolls off a table) are used in predicates that describe agent-less events. In our study, we asked whether signers of an emerging language use handshape to distinguish among scenes with a more or less salient agent and, if so, at what point in the process of language emergence this distinction is first made.

We tested 4 adult homesigners living in Nicaragua, 8 Cohort 1 signers and 10 Cohort 2-3 signers. We collapsed data from Cohorts 2 and 3, both of whom received a linguistic system (NSL) as input, and compared their productions to Cohort 1 signers, who did not receive linguistic input but did have a linguistic community, and to Homesigners, who had neither linguistic input nor a linguistic community. As a test of their agent-demotion language, participants viewed video clips from a Body condition (a person manipulates an object, e.g., woman pushes over a book) vs. a Hand condition (a hand manipulates an object without the body/face shown, e.g., hand pushes over a book). We believed this Body/Hand

distinction would elicit agent-demotion language, as pilot studies showed that in the Hand condition, English speakers were 40% more likely to use passive, e.g. *the book was pushed over*. As a control, participants also described No-Agent scenes (the object moves by itself, e.g. book falls over).

Each predicate in a signer's descriptions was glossed and coded for handshape type (handling vs. object). We categorized each trial by response strategy: handling predicates only, object predicates only, or both handling and object predicates. Figure 1 shows these results as a function of signing group and condition:

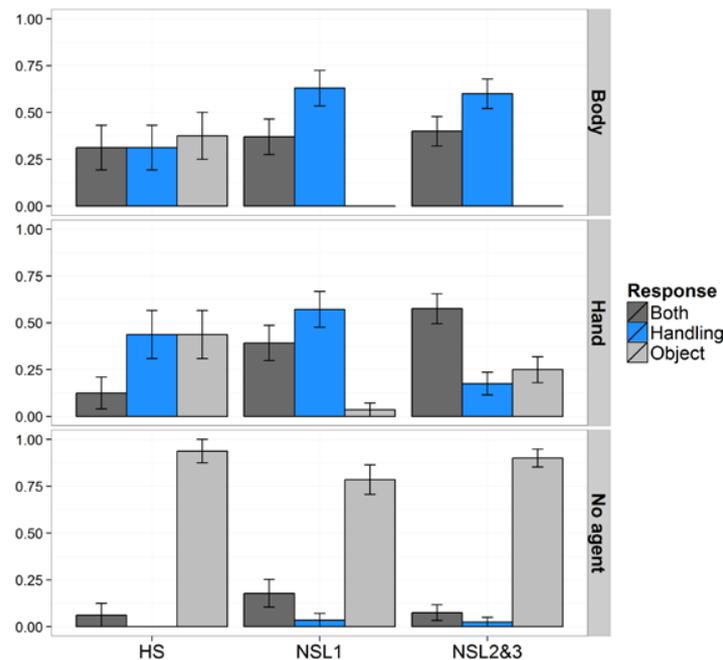


Figure 1. Distribution of handshape responses by Signing group & Condition.

All 3 groups were more likely to use handling handshape (blue bars) for agentive scenes and object handshapes (light gray bars) for non-agentive scenes, suggesting that marking agency requires neither a linguistic community nor linguistic input for its emergence. Note, however, that Homesigners did use some object handshapes alone for Body events, whereas Cohorts 1 and 2-3 did not, suggesting that agency is marked more categorically in signers who have a linguistic community. Moreover, only Cohorts 2-3 made a distinction between weakly agentive (Hand) and strongly agentive (Body) events. They produced utterances with two predicates for Hand events, one with a handling handshape and one with an object handshape, which may indicate the emergence of a serial verb strategy for demoting agents. In contrast, Cohort 2 signers favored single predicates with a handling handshape for Body events. The fact that only Cohorts 2-3 distinguish Body from Hand events suggests that linguistic input may be essential to the emergence of agent-demotion devices. In particular, for Cohort 1 signers the handling/object distinction may serve primarily to mark the fundamental agent/no-agent contrast, whereas Cohort 2 signers can use the distribution of handling and object handshapes to also mark the more subtle demoted-agent contrast.

References

- Benedicto, E., & Brentari, D. (2004). Where did all the arguments go?: Argument-changing properties of classifiers in ASL. *Natural Language & Linguistic Theory*, 22(4), 743-810.
- Senghas, A., & Coppola, M. (2001). Children Creating Language: How Nicaraguan Sign Language Acquired a Spatial Grammar. *Psychological Science*, 12(4), 323-328.
- Siewierska, A. (2013). Passive Constructions. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*.

SOCIAL BIASES VERSUS EFFICIENT COMMUNICATION: AN ITERATED LEARNING STUDY

GARETH ROBERTS¹, MARIYA FEDZECHKINA²

¹*Department of Linguistics*, ²*Department of Psychology*
University of Pennsylvania, Philadelphia, PA
gareth.roberts@ling.upenn.edu, mfedze@sas.upenn.edu

A crucial question in the evolution of language concerns the “problem of linkage” (Kirby, 1999, 20): How do the constraints acting on individual language users give rise to observed patterns of linguistic diversity? For instance, recent work has suggested that some properties of language benefit efficient information transmission (e.g., Piantadosi et al., 2011), and Fedzechkina et al. (2012) showed experimentally that efficiency can increase through restructuring by language learners. A number of other experiments investigating the problem of linkage have employed the iterated-learning model (ILM), which simulates cultural transmission by using the output of one learner as the training data for the next (Kirby et al., 2014). A typical simplification is that learners are exposed to data undifferentiated by source, very often from only one individual. Real-world transmission, by contrast, involves multiple models distinguished by such variables as contact frequency and social status, which are known to influence the spread of linguistic variants (Labov, 2001). An interesting case concerns situations where social pressures apparently run counter to efficiency. For instance, modern plural second-person pronouns in English (*yous*, *y’all*, *yinz*) reduce ambiguity, but are often avoided on social grounds (leaving ambiguity, or requiring less efficient workarounds). Social factors may also explain the retention of distinctions that do little communicative work, such as *who/whom*.

In this study, we focus on two questions left open by prior work: a) Is the increase in efficiency observed by Fedzechkina et al. amplified by iterated learning? b) How is this process influenced by the presence of social pressures that run counter to communicative efficiency? To investigate this, we conducted an iterated learning experiment on Amazon Mechanical Turk, in which participants were exposed equally to two dialects of an “alien language”. Both dialects exhibited strict SOV word order, but one dialect redundantly marked case, while the other did not. Participants in the *Bias* condition were encouraged to view the aliens speaking the redundant dialect as potential trading partners. In the *No bias* condition all aliens were potential trading partners. There were five chains in each condition, with two

participants in each generation, whose output in the test phase was presented as the redundant dialect to the next generation. Case marking behaved significantly differently in the two conditions (Figure 1). All first-generation participants were exposed to 50% case-marked sentences. In the No bias condition, this proportion declined fast and disappeared completely within four generations for all chains. In the Bias condition, it also declined, but more slowly, disappearing eventually in only three chains. Our results suggest that the effect of learners' biases towards efficiency is amplified by transmission and can thus account for (some) observed cross-linguistic typological patterns. However, while languages in both conditions became more efficient over generations, the biases towards efficiency were modulated by social factors—redundant case-marking persisted longer when associated with a preferred social group.

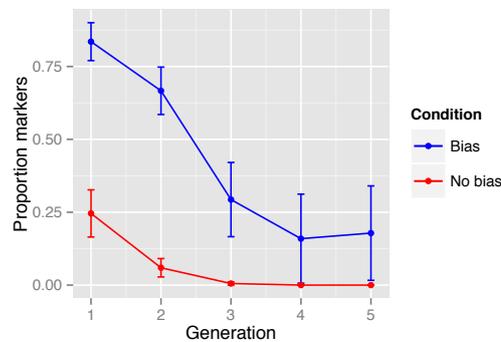


Figure 1. Mean proportion of sentences produced with case markers per generation.

References

- Fedzechkina, M., Jaeger, T. F., & Newport, E. (2012). Language learners restructure their input to facilitate efficient communication. *109*(44), 17897–17902.
- Kirby, S. (1999). *Function, Selection and Innateness: The Emergence of Language Universals*. Oxford: Oxford University Press.
- Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114.
- Labov, W. (2001). *Principles of Linguistic Change. Volume 2: Social Factors*. Malden, MA: Blackwell.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.

HOMO SAPIENS IS HOMO LOQUENS: THE POWER OF VOCAL LEARNING

JOANA ROSSELLÓ

*General Linguistics Department, Universitat de Barcelona,
Barcelona, Spain
joana.rossello@ub.edu*

Homo sapiens is a species of vocal learners, apparently the only extant one among great apes and terrestrial mammals. This is a fact about human biology. Humans, however, differ hugely from songbirds, the other sophisticated vocal learners known thus far (Arriaga & Jarvis 2013). Only the human ones convey meaning, which to a great extent comes from words. Words, because they are invented, are not the outcome of natural selection (Bickerton 2014); birdsong motives instead conform to it. This notwithstanding, words are built on top of a vocal learning (VL) system which in a way naturalizes them. Words end up residing in and circulating among brains, which must have affected (and affect) the functioning of the latter. This crucial role as a vehicle for words in speech is one of the different aspects which have been grossly underestimated with regard to the contribution of VL to human language and cognition.

By rooting language in a VL system, one gets rid of false dilemmas such as whether what took the lead in the process was communication or cognition, as Tomasello and Chomsky respectively contend. With VL as point of departure the issue dissolves: VL entails *per se* a codependence between social behavior and individual capacity. A second dilemma that fades away is the order of appearance of the computational system (syntax) or phonology (words) in evolution. Yet, if a VL system has a computational system (CS) of its own, the same that combines the words that run on it, the issue ceases to make sense. It has been claimed (Bolhuis, 2014) that the CS underlying birdsong is devoid of the asymmetrical hierarchy present in natural language. This problem is not unsurmountable. As a matter of fact, *merge* alone, the single operation of the CS, yields no asymmetrical structure (*eating apples* is meant to be an instance of

eating and not of apples). A suggestion is that the asymmetry derives from word-related conceptual properties (Boeckx, 2013).

Another eliminative gain is that primate *calls* as such could not have played any role in language evolution, against Miyagawa's integration hypothesis (see Nóbrega & Miyagawa, 2015). Apart from cognitive and behavioral traits – intentionality displayed in the gestures of apes in particular –, there is nothing in great apes' vocalizations that could have been recruited for language: primate *calls* constitute a hurdle in the way to language because of being subcortically controlled and occasionally endowed with a referential power of a mind-independent sort in contrast with that of words, indirect and mind dependent.

The primacy of speech over sign as the default modality must not be elusive anymore (Goldin-Meadow, 2008). Speech is the default option because only speech, by being *auditory-vocal*, fits a VL system, a modality of sensorimotor integration independently available in nature. A gestural protolanguage, however, could not have existed since there is no sensorimotor visuo-gestural other than the signed modality itself (and other human occasional mimicry practices), which emphasizes the crucial role of the sensorimotor integration in language. The essential overlapping of neural correlates for both modalities is also consistent with the idea that sign is modeled after speech.

Highly suggestive of the sufficiency of this bottom-up approach to language evolution is the fact that the brain circuitry in charge of birdsong finds not only a correspondence in the neural pathways that are involved in learned vocal communication in humans but is also essentially coincident with the subcortical network (basal ganglia and thalamus) that is increasingly seen as involved in language (beyond speech) and high cognition as well (Lieberman 2013). *Homo Sapiens* seems to amount to *Homo Loquens*.

References

- Arriaga, G. & Jarvis, E. D. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain & Language*, 124: 96–116.
- Bickerton, D. (2014). *A non-selectionist origin for the human mind*. Ms.
- Boeckx, C. (2013). Merge: Bilingual Considerations. *Eng. Lin.*, 30(2), 1–21.
- Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. (2014). How could language have evolved? *PLoS Biology*. Biol 12(8):
- Goldin-Meadow, S. (2008) Gesture, Speech and Language. *Evolang* 7, 427-428.
- Lieberman, P. (2013). *The Unpredictable Species. What Makes Humans Unique*.
- Nóbrega, V. A., & Miyagawa, S. (2015). The precedence of syntax in the rapid emergence of human language in evolution as defined by the integration hypothesis, *Frontiers in Psychology* 6.

THE CULTURAL EVOLUTION OF COMPLEXITY IN LINGUISTIC STRUCTURE

CARMEN SALDANA, SIMON KIRBY AND KENNY SMITH

*Centre for Language Evolution, School of Philosophy, Psychology and Language
Sciences, The University of Edinburgh, Edinburgh, UK
c.c.saldana@sms.ed.ac.uk, simon.kirby@ed.ac.uk, kenny.smith@ed.ac.uk*

Languages are culturally transmitted through a repeated cycle of learning and communicative interaction, a process known as iterated learning. Previous work has shown how different features of linguistic structure evolve from the trade-off between different competing pressures acting on language learning and communication such as compressibility and expressivity (Kirby, Cornish, & Smith, 2008; Perfors, Tenenbaum, & Regier, 2011; Lupyan & Dale, 2015; Regier, Kemp, & Kay, 2015; Kirby, Tamariz, Cornish, & Smith, 2015). In Kirby et al. (2015), compositional miniature artificial languages evolve as a result of their transmission across “generations”. Where both compressibility and expressivity pressures are in play, signals in later generations are composed of atomic units, each mapping to a specific dimension of the meaning to be conveyed. However, the complexity of the languages which evolve in these experiments is necessarily limited by the objects (meanings) people were learning labels for. In particular, the sets of objects to be labelled do not require a language which exhibits hierarchical constituency and syntactic categories. In this paper, we increase the complexity of the meanings to be conveyed by including motion events that comprise shape, number, motion and aspect. The events are composed by a focal object which performs the action and optionally, an anchor object which remains static. By increasing the complexity of the meaning space, we expect the same mechanisms involved in the evolution of simple compositionality to lead to richer syntactic structure more closely resembling that found in real languages.

We ran an Iterated Artificial Language Learning study and manipulated the expressivity pressure. We designed a monadic condition (N=32) with an artificial pressure for expressivity, and a dyadic condition (N=80) with communication as a natural pressure for expressivity. Following Kirby et al. (2015) we use the transmission chain paradigm. Participants were trained on a set of meaning-signal mappings, and then tested on their ability to recall that language. The first participants in a chain were trained on a non-compositional randomly generated language. Subsequent participants were trained on the language produced by the

previous participants. The test phase of the monadic condition involved typing descriptions for motion event scenes using the language learned previously; participants were not allowed to reuse the same description for different meanings, introducing an artificial pressure for expressivity. The test phase in the dyadic condition required participants to communicate with their partner in the language that they previously learned; members of a dyad alternated between describing meanings for their partner, and interpreting descriptions provided by their partner.

In accordance with previous results, we found a significant increase in learning success and structure in both conditions along the evolution of compositional structure. Moreover, constituency was hinted at by the emergence of morphologically complex N-like and V-like syntactic lexical categories. These categories were used to form hierarchically compositional sentential structures with meaningful word order.

Despite the qualitative similarity of the results in the two conditions, we found that condition significantly affected the evolution of structure: languages in the dyadic condition became structured more rapidly and their level of structure was consistently higher. The levels of complexity in the emergent compositional systems were significantly different between conditions: the systems in the monadic condition showed higher system complexity on average and less transparent morphosyntactic structures (i.e. they exhibit functional elements such as category markers, and non-adjacent dependencies, not found in the dyadic condition).

Compositionality operating at the levels of morphology and syntax evolved through the trade off between compressibility and expressivity. Nevertheless, the difference in complexity found between the two conditions points to the need for further investigation into the nature of the pressure for expressivity in these experiments. In the dyadic condition, the need to maintain communication may lead to a conservative approach. If participants find a solution that works, they stick with it. In the monadic condition, the pressure for expressivity is quite different. The need to avoid reuse of the same description for different meanings leads to an anti-conservative approach, with participants actively generating novel signals. Future work should investigate whether an analog of this tendency to innovate is at play in real languages, and consequently whether a pressure for novelty needs to take its place alongside compressibility and expressivity in the evolution of complex linguistic structure.

References

- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS*, *105*(31), 10681–10686.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.
- Lupyan, G., & Dale, R. (2015). The role of adaptation in understanding linguistic diversity. In R. D. Busser & R. J. LaPolla (Eds.), *Language structure and environment: Social, cultural, and natural factors.cognitive linguistic studies in cultural contexts*, 6. John Benjamins.
- Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*(3), 306–338.
- Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence*. John Wiley & Sons.

Skepticism towards skepticism towards computer simulation in evolutionary linguistics

Carlos Santana

Department of Philosophy and Institute for Research in Cognitive Science,
University of Pennsylvania

The evolution of human language has been a perennially controversial topic of research, with recent criticisms of research into language evolution focusing on a lack of empirical evidence behind the theories. The high-profile authors of Hauser et al. (2014), for instance, argue that "[b]ased on the current state of evidence...the most fundamental questions about the origins and evolution of our linguistic capacity remain as mysterious as ever." This pessimism runs counter to the optimism expressed by many scientists who, while acknowledging the scarcity of evidence, think we have the tools to make progress on this difficult subject matter. In particular, many argue that computer simulations allow us to "compensate for the lacking empirical evidence by utilizing methods from computer science and artificial life" (Lekvam et al. 2014; see also Cangelosi and Parisi 2002).

Skepticism about the ability of computer simulation to add substantively to our knowledge of language evolution is widespread, however. Skeptics point out that the fact that a simulation results in a realistic-seeming outcome does not entail that that simulation recapitulated the actual historical processes yielding that outcome (e.g. Templeton 2007). They argue that it's easy to establish contradictory claims on the basis of simulation (e.g. Roberts 2010), and that simulations are often tailored to support the modeler's prior theory (Martins et al. 2014). More generally, skeptics worry that simulations rely on empirical assumptions unsupported by evidence (Hauser et al. 2014), leading some to conclude that quantitative simulation should play only a reduced role in evolutionary linguistics (Martins et al. 2014).

Responses to this sort of skepticism about simulating the evolution of language have either been vague (e.g. Cangelosi and Parisi 2002) or limited in scope (e.g. the appeal to robustness analysis by Irvine et al. 2013). While acknowledging the skeptic's valuable call for the need to integrate computer modeling with empirical sources of evidence, I develop a more filled out defense of the evidential value of computer simulations to evolutionary linguistics.

I argue that although simulations do not provide evidence on their own, they can extend the inferential reach of weak empirical evidence. Data from comparative biology, archeology, historical linguistics, psycholinguistics, and so on often has weak and unclear applicability to our understanding of the origins and evolution of language. Simulation, I suggest, can aid in confirming or disconfirming hypotheses about language evolution because they draw out the inferential import of this otherwise hard to digest empirical data. This means that simulations play an important evidential role without actually constituting a source of evidence in of themselves.

Drawing on analogous uses of simulations in other areas of biology, I illustrate multiple ways in which they can play this role. First, computer models can connect the dots between observed facts and particular hypotheses. As a number of authors have observed, one reason quantitative models are valuable is that they precisify informal theories. Simulations can thus be used to test theories against evidence whose import for a theory is otherwise unclear. Second, simulations can situate contemporary evidence in otherwise hard-to-test evolutionary settings. Much of the most exciting work in evolutionary linguistics involves experiments with human and animal subjects in tasks involving, for instance, signal learning or cultural transmission. There are legitimate worries, however, about what we can extrapolate from this lab data. Here simulations can help by taking mechanisms observed in the lab, and situating them in more realistic evolutionary settings, which are drawn from the best current theories in quantitative evolutionary biology. Third, as Irvine et al. (2013) have argued, simulations can be used as part of a robustness analysis to draw out plausible actual mechanisms of language evolution.

Given these ways in which simulation can extend the inferential reach of empirical evidence, I conclude that the common forms of skepticism about simulating the evolution of language slightly miss the mark. The issue is not that simulations can't play an important evidential role, but instead that to play that role they must relate to empirical evidence in the right sort of ways. Our energies are thus better directed at improving the relationship between modeling and experimental and observational work, rather than at decrying the use of computer simulation at all.

References

- Cangelosi A, & Parisi D (2002) *Simulating the evolution of language* (Vol. 1). London: Springer.
- Hauser MD, Yang C, Berwick RC, Tattersall I, Ryan MJ, Watumull J, Chomsky N and Lewontin RC (2014) The mystery of language evolution. *Front. Psychol.* **5**:401.
- Irvine L, Roberts SG, & Kirby S (2013). A robustness approach to theory building: A case study of language evolution. In the *35th Annual Meeting of the Cognitive Science Society* (CogSci 2013) (pp. 2614-2619).
- Lekvam T, Gambäck B, and Bungam L (2014) Agent-based modeling of language evolution *Proceedings of 5th workshop on Cognitive Aspects of Computational Language Learning* 49.
- Martins M, Raju A, & Ravignani A. (2014). Evaluating the role of quantitative modelling in language evolution. *The past, present and future of language evolution research*, 84-93.
- Roberts G (2010) An experimental study of social selection and frequency of interaction in linguistic diversity. *Interaction Studies*, 11(1), 138-159.
- Templeton AR (2007) Genetics and recent human evolution. *Evolution* 61(7) 1507-1519.

FROM NATURAL ORDER TO CONVENTION IN SILENT GESTURE

MARIEKE SCHOUWSTRA, KENNY SMITH, SIMON KIRBY

*Centre for Language Evolution, School of Philosophy, Psychology & Language Sciences,
University of Edinburgh, Edinburgh, United Kingdom
Marieke.Schouwstra@ed.ac.uk, kenny@ling.ed.ac.uk, simon@ling.ed.ac.uk*

Silent gesture, an experimental paradigm in which adult hearing participants describe events using only their hands, has been valuable for investigating the origins of word order. Goldin-Meadow et al. (2008) found a language-independent preference for SOV for extensional transitive events (e.g., boy-ball-throw), but participants prefer SVO for intensional events (e.g., boy-search-ball; Schouwstra & de Swart, 2014).

The SVO/SOV pattern for intensional/extensional events arises independently of participants' native language, and, we will claim, represents naturalness, reflecting cognitive preferences to put Agents first (Jackendoff, 2002) and more abstract/relational information last. However, existing languages tend not to condition word order on event type and are instead more regular. Understanding this transition from naturalness to conventionalised regularity is a major goal of language evolution research. We present a new approach to this challenge using a novel experimental paradigm in which silent gesture is both used for communication (Christensen et al, 2016) and culturally transmitted through artificial generations of lab participants (Smith et al, in prep). This allows us to investigate how individual humans improvise solutions to communicative challenges, how pairs of individuals create conventions through interaction, and how these conventions are transmitted over time through learning.

In experiments 1a and 1b, 48 participants were assigned into dyads. Stimuli were 64 line drawings of intensional and extensional events. Participants alternated between the role of actor and interpreter, in six rounds of 32 trials. As actor they described an image using only their hands, and as interpreter they selected (from an array of 8) the image they thought was intended by the actor. Experiment 1a showed intensional and extensional events equally often; in experiment 1b, extensional events were more frequent (75%) than intensional events (25%).

The word orders showed signs of conventionalisation: over the rounds, word order became less conditioned on meaning. 7 of 12 dyads in experiment 1a con-



Figure 1 An example of an extensional event, 'artist drops hammer' (left) and an intensional event, 'nun dreams of violin' (right).

verged on a single word order:^a all of them SVO (the dominant order of English, the native language of all participants). In experiment 1b, 3 dyads converged on SVO word order, and 3 other dyads on SOV. The dyads who failed to converge used a mix of SOV and SVO orders.

Experiments 2a and 2b were carried out with 16 groups of 8 participants (8 groups per condition) in a gradual turnover design. In round one, participants 1 and 2 were communicators, and 3 was an observer. Each consecutive round, one of the communicators left, the observer became a communicator, and a new participant became observer. Experiment 2a used equal amounts of extensional and intensional stimuli, and experiment 2b had more extensional (75%) than intensional (25%) events. In Experiment 2a, 5 out of 8 groups converged on SVO word order, while in Experiment 2b, 3 groups converged on SVO and 2 groups converged on SOV. This shows that even in a turnover setup, where participants interact for a maximum of two rounds instead of six, convergence of word order is possible, and dependent on the frequency of event types.

Our experiments show that in silent gesture communication and transmission, semantically conditioned word order tends to disappear in favour of regular word order. The frequency of event types determines how regularisation progresses. This suggests that where pressures for naturalness and regularity are in conflict, languages start natural, but naturalness will give way to regularity as signalling becomes conventionalised through repeated usage.

Acknowledgements

Marieke Schouwstra was funded by the British Academy.

^aFor each dyad we counted the proportion of the most frequent word order in the last 2 rounds, and those in which this was $\geq .8$ were counted as converging.

References

- Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, *146*, 67 - 80.
- Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *PNAS*, *105*(27), 9163–9168.
- Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1–17.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.
- Schouwstra, M., & Swart, H. de. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436.
- Smith, K., Abramova, K., Cartmill, E., & Kirby, S. (in prep). The experimental study of sign language emergence.

CHILDREN'S PRODUCTION OF DETERMINERS AS A TEST CASE FOR INNATE SYNTACTIC CATEGORIES

CATRIONA SILVEY

*Department of Psychology, University of Chicago
Chicago, USA
casilvey@uchicago.edu*

CHRISTOS CHRISTODOULOPOULOS

*Department of Computer Science, University of Illinois at Urbana-Champaign
Urbana, USA
christod@illinois.edu*

A central debate in language evolution is whether humans have a specific innate capacity for language, or whether domain-general learning abilities can explain the acquisition of linguistic structures. One testing ground for these hypotheses has been children's early use of English determiners, specifically the definite and indefinite articles 'the' and 'a'. The argument goes as follows: if children have an innate syntactic determiner category, they should interchangeably use 'the' and 'a' with all nouns as soon as they begin producing them with a determiner. However, if children initially learn determiner-noun combinations as islands and only gradually abstract a syntactic category, they should initially use particular nouns with only one determiner (Valian, Solt, & Stewart, 2009). These two possibilities can be quantified as 'overlap': the number of nouns children produce with both 'a' and 'the', divided by the number of nouns children produce with either. If overlap is 0, children use each noun only with one of the two determiners, suggesting island-based learning. If overlap is 1, children use each noun with both determiners, suggesting a productive syntactic category. Results from this paradigm have been mixed. Some researchers find that children's overlap is low, suggesting that an abstract category of determiner is gradually constructed rather than being present from the start (Pine, Freudenthal, Krajewski, & Gobet, 2013). Others counter that children's overlap is not significantly different from their parents', suggesting an innate syntactic category (Valian et al., 2009).

Yang (2013) addresses an important problem with using overlap as a measure of productivity. As Valian et al. (2009) observe, the fewer times a noun appears, the more likely it will appear with only one determiner. Therefore, low overlap may simply be the consequence of many nouns appearing only few times.

Yang therefore uses the frequencies of noun types and determiners to predict expected overlap if determiners and nouns freely combine within these frequency constraints. His model accurately predicts empirical overlap values in early child language. Yang interprets this result as showing that from the start, children have an abstract determiner category. This finding has since been cited as evidence for innate syntactic categories (Bolhuis, Tattersall, Chomsky, & Berwick, 2014).

We replicate Yang's model on the six children from the CHILDES corpus analysed in Yang (2013). We show that while the model holds on average across nouns, it poorly predicts the behaviour of individual nouns. As a result, it systematically underestimates the overlap that would occur if nouns and determiners freely combined within Zipfian constraints. Keeping constant the overall frequencies of nouns and determiners, we shuffle each child's productions so that determiners and nouns combine at random. For these shuffled data, overlap measures exceed those predicted by Yang's model. The model, then, predicts the children's data not because they resemble the product of a freely combinatorial grammar, but because determiners and nouns do **not** freely combine: many mid- to high-frequency nouns appear with only one determiner. While Yang acknowledges these 'use asymmetries', he characterises them as 'unlikely to be linguistic'. We argue, however, that a) these asymmetries significantly constrain both children's and adults' data, and b) they are linguistic, specifically the product of lexical semantics interacting with the discourse functions of 'a' and 'the'. Since the target of acquisition is therefore not a freely combinatorial system, but one conditioned on semantics and discourse factors, children's productions are more accurately represented as a gradual acquisition of these factors, rather than as either islands or grammatical combinations isolated from discourse. More broadly, studies using naturalistic corpora to test hypotheses about language acquisition and evolution should be wary of either taking constrained usage patterns as evidence of lack of grammar, or abstracting away from them in aid of revealing underlying rules, since these constraints are a non-arbitrary part of the function of the language.

References

- Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How could language have evolved? *PLOS Biology*, *12*(8), e1001934.
- Pine, J. M., Freudenthal, D., Krajewski, G., & Gobet, F. (2013). Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner. *Cognition*, *127*(3), 345–360.
- Valian, V., Solt, S., & Stewart, J. (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language*, *36*(4), 743–778.
- Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6324–7.

VOCAL LEARNING IN FUNCTIONALLY REFERENTIAL CHIMPANZEE FOOD CALLS

KATIE E. SLOCOMBE

*Department of Psychology, University of York,
York, UK
Katie.slocombe@york.ac.uk*

STUART K. WATSON, ANNE M. SCHEL, CLAUDIA WILKE, EMMA WALLACE,
LEVEDA CHENG, VICTORIA WEST

*Department of Psychology, University of York,
York, UK
sw218@st-andrews.ac.uk, annemschel@gmail.com, cw562@york.ac.uk,
ekw510@york.ac.uk levedacheng@hotmail.com, v.west@newcastle.ac.uk*

SIMON W. TOWNSEND

*Department of Psychology, University of Warwick,
Coventry, UK
Simon.W.Townsend@warwick.ac.uk*

One standout feature of human language is our ability to reference external objects and events with socially learnt symbols, or words. Exploring the phylogenetic origins of this capacity is therefore crucial to a comprehensive understanding of the evolution of language. While non-human primates can produce vocalizations that function as if they refer to external objects in the environment, the psychological mechanisms underlying call production in terms of a caller's motivation and a caller's ability to alter the structure of these calls is likely different to humans (Wheeler and Fischer, 2012). Indeed it is generally argued that the acoustic structure of context specific calls elicited by salient external stimuli (e.g. food, predator) is directly determined by arousal states induced by the external stimuli (Wheeler and Fischer, 2012). This apparent lack of flexible control over the structure of functionally referential vocalizations represents a key discontinuity with language. We tested the degree of flexibility in the acoustic structure of functionally referential chimpanzee food calls (Slocombe and Zuberbuhler, 2005) and whether the structure of these calls could be influenced by vocal learning processes.

We examined the food preferences and acoustic structure of food calls of two groups of adult chimpanzees, prior to and for 3 years after their integration into a single group at Edinburgh Zoo, UK. Prior to social integration in 2010 the resident Edinburgh (ED) chimpanzees (N = 6) and the immigrant Beekse Bergen (BB) chimpanzees (N = 7) had significantly different preferences for apples and produced acoustically distinct calls whose structure, in line with previous research, matched their preferences for this food (Slocombe and Zuberbühler 2006). Apples were regularly fed to both groups for at least 3 years before integration, so were not a novel food for either group. General arousal levels may have been elevated in 2010 as both groups adjusted to a new social environment and BB chimpanzees habituated to a new enclosure. However, in 2011, one year after integration and habituation to the new social and physical environment, the call structures and preferences for apples of the two groups remained stable, indicating changes in general arousal were not affecting call structures. Social network analysis (SNA) revealed two distinct subgroups in 2011, with individuals still preferring to associate with members of their original group and maybe lacking the motivation to converge their calls. In 2013, SNA showed the subgroups had dissolved and strong inter-group relations had developed. Although the ED calls stayed stable in their structure 2010-13, in 2013 BB calls changed significantly to converge with the lower frequency ED calls. Importantly this call convergence occurred independently of preferences for apples, which stayed stable over all years for both groups. This shows a decoupling of the affective response induced by the external stimulus (apples) and the structure of the call produced. We argue that these data represent the first evidence of non-human animals actively modifying and socially learning the structure of a meaningful functionally referential vocalization from conspecifics. Our findings indicate that functionally referential call structure is not solely determined by arousal processes in our closest living relative. Although this modest degree of acoustic change within an existing call type is not analogous to the impressive vocal learning shown by humans, this flexibility may be an important evolutionary *precursor* to socially learnt referential words that are so central to human communication.

References

- Slocombe, K.E., & Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour*, *72*, 989–999.
- Slocombe, K.E., & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology*, *15*, 1779–1784.
- Wheeler, B.C., & Fischer, J. (2012). Functionally referential signals: a promising paradigm whose time has passed. *Evolutionary Anthropology*, *21*, 195–205.

CHIMPANZEES PROCESS STRUCTURAL ISOMORPHISMS ACROSS SENSORY MODALITIES

RUTH SONNWEBER

*Department of Cognitive Biology, University of Vienna
Vienna, Austria
ruth-sophie.sonnweber@univie.ac.at*

ANDREA RAVIGNANI

*Artificial Intelligence Lab, Vrije Universiteit Brussel,
Brussels, Belgium*

*Department of Cognitive Biology, University of Vienna
Vienna, Austria
andrea.ravignani@gmail.com*

Humans and other animals are constantly exposed to environmental stimuli, from which they extract sensory regularities (Fitch, 2014; ten Cate and Okanoya, 2012). Moreover, they often relate and integrate one-dimensional quantities across sensory modalities (Ludwig et al., 2011), for instance relating conspecific faces to voices (Seyfarth and Cheney, 2009). If basic patterns like repetitions and identities are perceived in different sensory modalities (Ravignani et al., 2013; Ravignani et al., 2015; Sonnweber et al., 2014), it could be advantageous to detect cross-modal isomorphisms, i.e. modality-independent representations of structural features, which could be used in visual, tactile, and auditory processing. Humans can transfer structural regularities learnt in one modality, e.g. visual sequences, to another modality, e.g. unfamiliar sound sequences (Altmann et al., 1995). To date, this ability to map structural regularities across domains has not been demonstrated in other animals. Here we show that two chimpanzees trained to choose symmetric sequences of geometric shapes spontaneously detected a visual-auditory isomorphism. Although chimpanzees were never trained to associate sounds to images, their response latencies in choosing symmetric visual sequences was shorter when presented

with (structurally isomorphic) symmetric, rather than foil sound triplets. Thus, previously unheard sound sequences influenced the choice of visual sequences solely based on structural similarities. This provides the first evidence of structure learning across modalities in a non-human animal. Our findings suggest that human language is not a prerequisite to map abstract structures between modalities. Cross-modal abilities might instead have constituted a precursor to human linguistic abilities (Cuskley and Kirby, 2013), involving evolutionary old neural mechanisms (Ghazanfar and Takahashi, 2014).

Acknowledgements

Research was supported by ERC grants 230604 SOMACCA (to W. Tecumseh Fitch) and 283435 ABACUS (to Bart de Boer). We thank C. Cus, E. O'Sullivan, S. Prislei, and G. Schiestl for invaluable comments and R. Hofer, N. Kavcik, and J. Oh for support.

References

- Altmann, G., Dienes, Z., and Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 899.
- Cuskley, C., and Kirby, S. (2013). *Synaesthesia, cross-modality and language evolution*. In Oxford Handbook of Synaesthesia, J. Simner and H. E.M., eds. (Oxford University Press), pp. pp. 869-907.
- Ludwig, V.U., Adachi, I., and Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *PNAS*. 108, 20661-20665.
- Fitch, W. T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Physics of life reviews*, 11(3), 329-364.
- Ravignani, A., Sonnweber, R.-S., Stobbe, N., and Fitch, W.T. (2013). Action at a distance: dependency sensitivity in a New World primate. *Biol. Lett.* 9, 20130852.
- Ravignani, A., Westphal-Fitch, G., Aust, U., Schlumpp, M. M., & Fitch, W. T. (2015). More than one way to see it: Individual heuristics in avian visual computation. *Cognition*, 143, 13-24.
- Seyfarth, R.M., and Cheney, D.L. (2009). Seeing who we hear and hearing who we see. *PNAS*. 106, 669-670.
- Sonnweber, R., Ravignani, A., and Fitch, W.T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Anim. Cogn.*, 18(3), 733-745.
- ten Cate, C., & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B*, 367(1598), 1984-1994.

Ghazanfar, A.A., and Takahashi, D.Y. (2014). The evolution of speech: vision, rhythm, cooperation. *Trends. Cogn. Sci.* 18, 543-553.

**RULE LEARNING IN BIRDS:
ZEBRA FINCHES GENERALIZE BY ITEM POSITION,
BUDGERIGARS BY STRUCTURE**

MICHELLE J. SPIERINGS

*Institute of Biology Leiden (IBL) and Leiden Institute for Brain and Cognition (LIBC),
Leiden University,
Leiden, Netherlands
m.j.spierings.2@biology.leidenuniv.nl*

CAREL TEN CATE

*Institute of Biology Leiden (IBL) and Leiden Institute for Brain and Cognition (LIBC),
Leiden University,
Leiden, Netherlands
c.j.ten.cate@biology.leidenuniv.nl*

The ability to abstract a rule that defines the structure of strings of sounds is a core mechanism underlying the language faculty, but might not be specific to language learning or even to humans. Up until now, it is unclear whether and to what extent non-human animals possess the ability to abstract a rule defining the relationship among arbitrary auditory items in a string and to generalize this rule to strings of acoustically novel items (ten Cate & Okanoya, 2012; ten Cate, 2014). In this study we tested both a songbird (zebra finch) as well as a parrot species (budgerigar) on these rule learning abilities. Subjects were trained in a go/no-go design to discriminate between two sets of sound strings that corresponded to either an *XYX* or an *XXY* structure. After this discrimination was acquired, each subject received a number of test strings (mixed with the training strings) that followed the same structural rules, but consisted of either new combinations of known elements or of novel elements belonging to other categories. If the animals paid attention to sound-specific features of the training strings, their responses would differ between the test strings with known sounds and the strings with novel sounds. If, however, the birds learned the structure of the strings, independent of the sounds, they should respond similarly to all test strings. Both species learned to discriminate

between the two stimulus sets during training. However, their responses to the test strings were strikingly different. Zebra finches categorized test stimuli with known elements by the positions that these elements occupied in the training strings. A subsequent experiment with artificially created sound elements showed that this was independent of whether the strings consisted of conspecific or unknown sounds. In contrast, the budgerigars categorized both novel combinations of familiar elements as well as strings consisting of novel elements by their underlying structure. They thus abstracted the relationship among items in the *XYX* and *XXY* structures, indicating a level of abstraction comparable to analogical reasoning, a cognitive ability long thought to be unique for humans and thus far only known from great apes and crows (Thompson & Oden 2000; Smirnova et al; 2015). Our study is the first clear indication that abstract rule learning in auditory strings is not specific to language or to humans.

References

- Smirnova, A., Zorina, Z., Obozova, B., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current Biology*, 25, 1-5.
- ten Cate, C. (2014). On the phonetic and syntactic processing abilities of birds: from songs to speech and artificial grammars. *Curr. Opin. Neurobiol.* 28, 157-164.
- ten Cate, C., & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Phil. Trans. R. Soc. B.*, 367(1598), 1984-1994.
- Thompson, R.K.R., & Oden, D.L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cogn. Sci.*, 24, 363-396.

INFORMATION DYNAMICS OF LEARNED SIGNALLING GAMES

MATTHEW SPIKE, SIMON KIRBY, KENNY SMITH

*Language Evolution and Computation Research Unit, Linguistics and English Language,
University of Edinburgh, Edinburgh, UK
matthew.spike@ed.ac.uk*

Signalling games involving agent learners exist in various guises, from the game-theoretic Roth-Erev learners of Skyrms (2010), to the Naming Game (Steels, 1997), and agents employing varieties of observational learning (e.g. Oliphant & Batali, 1996; Smith, 2002). The agent-based nature of this work means that the resulting dynamics have an inherently unpredictable character: individual simulations may or may not be representative of average behaviour, if such a thing exists at all. Typically, the best way of overcoming this problem is by running large numbers of simulations and observing the aggregate behaviour. This contrasts with other frameworks — for example, classical or evolutionary game theory. In these cases, there is some *macro-level* property of the model which drives the overall dynamic of the game. For example, fitness of individual agents in evolutionary models is evaluated using the global average communicative success. Because of this, it is possible to calculate the mean-field dynamic for any known mixture of strategies in the population, revealing any attractors or stable points. In the case of agent-based models, because overall dynamics are completely determined by individual pairwise interactions — at the *micro-level* (Mühlenbernd, 2013) — the likely result of any interaction is not a direct consequence of the global communicative success of a population, which as a result cannot serve to describe the overall dynamics. Hence, identifying attractors and stable points poses a much harder problem. In order to resolve this problem, we introduce a new information-theoretic measure of optimality which *can* describe the overall dynamics of signalling populations of learning agents.

Typically, information theory (Shannon, 1948) has proven difficult to apply to problems involving meaningful communication as it has no way of describing semantic or referential content. Although there have been attempts to address this (e.g. Corominas-Murtra, Fortuny, & Solé, 2014), these still include a problematic macro-level term such as described above. However, we are able to avoid this under the assumption that agent signalling production and reception behaviours are derived from a single shared set of signal meaning associations. In this case, we can use the signal production behaviour of individual agents to describe their *in-*

dividual optimality in terms of the conditional entropy of meanings given signals, $H(M|S)$, where low entropy represents low ambiguity. Employing this measure, we show that the overall entropy of a system has two components determined by the average *individual entropy* and average *alignment entropy*: individual entropy measures the optimality of a single agent's own signalling system, while alignment entropy is the extra uncertainty due to the divergence of any agent from the population mean. We draw on results such as (Xue, 2006) which show that any population of agents which *imitate* each other with positive probability will inevitably drive the alignment entropy to zero.

This allows us to dissect the overall dynamics of any signalling game involving associative agents, which we do by analysing the pairwise interaction defined by its model of learning. In particular, we can describe any population as a point in an entropy *state-space*. Certain points within this space represent final stable states of the population in terms of their optimality. As such, we are able to show that the way 'imitative' learning by itself causes populations to move around the state-space resembles a type of genetic drift. Moreover, we identify the features which must exist to ensure populations develop optimal signalling: firstly, the *imitative* property described above; secondly, the learning model must on average *reduce conditional entropy* in any pairwise interaction. Finally, there must be a way to prevent learning slowdown: i.e. agents must retain *plasticity*. Using these three factors as a diagnostic, we are able to determine the dynamics of any population model involving associative signalling agents without recourse to numerical simulation, including whether or not it will develop optimal signalling. This applies to not just modelling work, but any theory of the emergence of novel lexicons.

References

- Corominas-Murtra, B., Fortuny, J., & Solé, R. V. (2014). Towards a mathematical theory of meaningful communication. *Scientific reports*, 4, 4587.
- Mühlenbernd, R. (2013). *Signals and the Structure of Societies*. Phd thesis, University of Tübingen.
- Oliphant, M., & Batali, J. (1996). Learning and the Emergence of Coordinated Communication. *Center for research on language newsletter*, 11(1), 1–46.
- Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(July, October), 379–423, 623–656.
- Skyrms, B. (2010). *Signals: Evolution, Learning & Information*. Oxford: Oxford University Press.
- Smith, K. (2002). The cultural evolution of communication in a population. *Connection Science*, 14(1), 65–84.
- Steels, L. (1997). Self-organising vocabularies. In *Proceedings of the 5th international workshop on artificial life: Synthesis and simulation of living systems*. ALIFE-96.
- Xue, J. (2006). A Polya Urn Model of Conformity.

MINIMAL PRESSURES LEADING TO DUALITY OF PATTERNING

MATTHEW SPIKE, KENNY SMITH, SIMON KIRBY

*Language Evolution and Computation Research Unit, Linguistics and English Language,
University of Edinburgh, Edinburgh, UK
matthew.spike@ed.ac.uk*

Hockett (1959) identified *duality of patterning* as a fundamental design feature of human language. Ladd (2012) re-analyses ‘duality’ as two levels of systematicity, one fully embedded within the other: a meaningless, *combinatorial* level, providing the building blocks for a meaningful, *compositional* level. Explanations for the emergence of combinatoriality include ease of production and reception (e.g. Roberts, Lewandowski, & Galantucci, 2015), robustness (Ay, Flack, & Krakauer, 2007), cultural adaptation to physical constraints (Zuidema & Boer, 2009), and learnability (Verhoef, Kirby, & Boer, 2014). Likewise, explanations for compositionality range from communicative function (e.g. De Beule, 2008) to cultural selection for expressivity and learnability (e.g. Kirby, Tamariz, Cornish, & Smith, 2015). Tria, Galantucci, and Loreto (2012) are the first to outline an integrated model of the emergence of duality, showing that distinct mechanisms of ‘noise/recognition’ and ‘blending/repair’ lead to the emergence of combinatorial and compositional structure respectively.

Our proposal is that both combinatoriality and compositionality are functional responses to maintain expressivity and learnability against noise. This depends on the level of analysis at which noise applies: signal-directed noise leads to combinatoriality, and noise which affects signal/meaning associations drives compositionality. Our approach contrasts with that of Tria et al.: we also investigate the emergence of the two levels of patterning, but we aim to show that they are driven by identical — not distinct — pressures. We employ an exemplar-based computational model of cultural learning subjected to twin pressures of expressivity and learnability. In common with Tria et al., utterances are modelled as strings drawn from a potentially infinite set of characters, subject to noise during transmission/storage. *Learnability* is modelled in terms of *compressibility*, a consequence of noisy pressures causing sub-strings across all exemplars to become more similar. *Expressivity* is an opposing force causing competition between similar sub-strings with a shared meaning. Besides these processes, agents are modelled simply as a set of exemplars associating full strings and complex meanings, with an exemplar memory of fixed size. We show that both combina-

torial *and* compositional structures act to maintain learnable, expressive systems against noise: similarly to Tria et al., we find that combinatoriality is modulated by signal-directed noise, which can be situated in both perception and cognition. Furthermore, compositionality also requires pressures for learnability and expressivity but, as with Kirby et al., whether systems become compositional or holistic depends on the presence of noise in the shape of an information bottleneck, which can be located in both transmission and memory. Given these results, we propose that combinatoriality emerges when noise puts the signal space under pressure to maintain learnability and expressivity: compositionality occurs when noise puts the signal/meaning association space under similar pressures. This helps dispel apparent conflicts between physical, perceptual and cognitive accounts of combinatoriality on the one hand, and acquisition vs. interaction-based accounts of compositionality on the other. However, this does not guarantee that duality of patterning will arise in any socially learnt communication system. Neither does either level of patterning predict the other: we suggest that duality is a response to noise at two levels of analysis.

References

- Ay, N., Flack, J., & Krakauer, D. C. (2007). Robustness and complexity co-constructed in multimodal signalling networks. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1479), 441–7.
- De Beule, J. (2008). the Emergence of Compositionality, Hierarchy and Recursion in Peer-To-Peer Interactions. *The Evolution of Language - Proceedings of the 7th International Conference (EVOLANG7)*, 75–82.
- Hockett, C. F. (1959). Animal ‘languages’ and human language. *Human Biology*, 31(1), 32–39.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102.
- Ladd, D. R. (2012). What is duality of patterning, anyway? *Language and Cognition*, 4(04), 261–273.
- Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, 141, 52–66.
- Tria, F., Galantucci, B., & Loreto, V. (2012). Naming a structured world: a cultural route to duality of patterning. *PloS one*, 7(6), e37744.
- Verhoef, T., Kirby, S., & Boer, B. de. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43, 57–68.
- Zuidema, W., & Boer, B. de. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, 37(2), 125–144.

**METALINGUISTIC AWARENESS OF TRENDS AS A DRIVING FORCE
IN LINGUISTIC EVOLUTION: AN EMPIRICAL STUDY**

KEVIN STADLER, ELYSE JAMIESON, KENNY SMITH, SIMON KIRBY

*Centre for Language Evolution
School of Philosophy, Psychology and Language Sciences
The University of Edinburgh, Edinburgh, Scotland
kevin.stadler@ed.ac.uk*

Like other culturally replicated traits, human languages – whether spoken or signed – evolve continually. But, unlike many other cultural traits, linguistic conventions are *arbitrary*: the exact form of a morpheme or syntactic word order convention matter much less than the fact that there *is* a convention that is understood by all. Consequently, once a convention is established there is a pressure to maintain it, and not to replace it with another form – so what selective forces cause languages to keep evolving?

Sociolinguistic research of the past decades has shown that language changes spread across social groups in an orderly fashion, a process which is typically explained by the notion of *prestige* – a metalinguistic property of linguistic variants that determines whether a speech community will seek to adopt the new variant or not. A positively evaluated variant might be taken up thanks to its *overt prestige* value, while the spread of negatively evaluated variants (those perceived as ‘wrong’ or otherwise imbued with negative associations) is said to be due to *covert prestige*. Crucially, the establishment of social prestige is itself a puzzle to be solved: the choice of which linguistic form becomes ‘prestigious’ is as arbitrary as the choice of using one form over another. The prestige value of a variant needs to be negotiated and spread across the speech community in the first place, a process which requires just as much explanation as the diffusion of the linguistic form that it is supposed to explain.

To put the notion of metalinguistic prestige on a more solid footing, Labov (2001, ch. 14) suggested that the steady advancement of changes across generations might be driven by adolescents’ awareness of the directionality of linguistic changes, combined with a pressure to discriminate themselves from older speakers. While experiments have shown that the latter pressure can indeed drive linguistic divergence of an artificial language when social group membership is marked explicitly (Matthews, Roberts, & Caldwell, 2012), empirical evidence that humans are able to exploit information on the directionality of ongoing changes

is still missing.

In this work we report results from a first quantitative investigation of the human capacity for tracking language evolution in progress. Using a questionnaire methodology we collected data on speakers' implicit and explicit awareness of three ongoing syntactic changes to verb positioning in the local variety of Scots spoken in Shetland, an island group to the North of Great Britain. 77 participants were asked to report their perceived usage levels of different age and speaker groups for the three changing variables as well as a stable, non-changing control. Our results show that individuals can reliably identify which of the competing linguistic variants are older and which are newer. The data also indicates that individual perceptions of apparent time differences (when younger speakers are leading a change, with the usage levels of older speakers 'lagging behind') can be used reliably to determine the directionality of the changes in progress.

The efficacy of 'trend-amplifying' selection mechanisms such as the one suggested by Labov has already been demonstrated theoretically by means of computational modelling (Stadler, Blythe, Smith, & Kirby, 2016). In particular, Mitchener (2011) showed that a model of language change that is based on perceived usage differences between age groups can successfully produce directional selection of arbitrary variants. Our quantitative results indicate that the information required by such mechanisms is indeed readily available to humans, and might consequently be used to coordinate changes across a community. Our results are in support of the idea that language evolution, rather than just being the result of drift and the incidental accumulation of errors in transmission and acquisition, is actively maintained and driven by individuals. Evidence to this end suggests that the origin of human language as we know it rests not only on an increased linguistic capability, but on metalinguistic capacities that are sensitive to variation and able to exploit usage patterns to actively guide linguistic divergence and change.

References

- Labov, W. (2001). *Principles of linguistic change. Social factors* (Vol. 2). Malden, Massachusetts: Blackwell Publishers Inc.
- Matthews, C., Roberts, G., & Caldwell, C. A. (2012). Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory. *Evolution and Human Behavior*, 33(6), 759–770.
- Mitchener, W. G. (2011). A mathematical model of prediction-driven instability: how social structure can drive language change. *Journal of Logic, Language and Information*, 20, 385–396.
- Stadler, K., Blythe, R. A., Smith, K., & Kirby, S. (2016). Momentum in language change: a model of self-actuating s-shaped curves. *Language Dynamics and Change*, 6(2).

THE GRAMMAR OF THE BODY AND THE EMERGENCE OF COMPLEXITY IN SIGN LANGUAGES

ROSE STAMP

*Sign Language Research Lab, University of Haifa,
Haifa, Israel
rose_stamp@hotmail.com*

WENDY SANDLER

*Sign Language Research Lab, University of Haifa,
Haifa, Israel
wsandler@research.haifa.ac.il*

In all human languages, spoken and signed, complex expressions are compositional: their meanings are determined by the meanings of their constituents and the rules for combining them (e.g., Krifka 2001; Jackendoff 2011; Pfau et al 2012; Smith & Kirby 2012). In sign languages only, however, while the hands convey words, individual actions of face, head, and torso can manifest different linguistic functions, often simultaneously, creating visual compositional, complex configurations. Squinted eyes in Israeli Sign Language (ISL) signal the interlocutor to retrieve shared information (Dachkovsky & Sandler 2009); brow raise signals yes/no questions in ISL as in ASL (Liddell 1980); and the combination of squinted eyes and brow raise in ISL signals a yes/no question about shared information (Nespor & Sandler 1999). Similarly, different head and body postures can represent different participants (e.g., Lillo-Martin 1995; Metzger 1995), concepts (van der Kooji & Crasborn 2006), or places in a discourse. Figure 1 schematizes how movements of different articulators contribute to the overall meaning of an utterance in contemporary ISL, to create a Grammar of the Body (Sandler to appear).

However, corporeal and linguistic complexity do not emerge all at once. An earlier, preliminary study of a newly emerging sign language in a Bedouin community with a high incidence of deafness, Al Sayyid Bedouin Sign Language, shows instead that the different articulators are recruited gradually across generations to convey increasingly complex linguistic functions (Sandler 2012). **This, suggests that the recruitment of the body in sign languages provides a visual map of the emergence of complexity in a new language.** In our current project, we adopt this initial finding as a strategy to systematically trace the diachronic development of linguistic complexity across three generations (including the first generation) in another sign language that originated only 80 years ago: Israeli Sign Language (ISL). We coded and analyzed two-minute narratives from 15 signers, five in each of three age groups, focusing on the form and function of head and torso actions. In this way we are able to identify increasing systematicity and complexity of linguistic structure as the language gets older, with the body as our guide.

We find that the signers use the head and torso differently and with increasing complexity across generations. Specifically, (1) while older signers use their articulators more than younger signers, it is younger signers who exploit a more variegated head and torso movement pattern by activating the side-to-side axis in addition to the forward and back movement favored by older signers. (2) An analysis of the language functions conveyed reveals that younger signers exploit the additional axis exclusively for marking specific linguistic functions, including parentheticals, questions and coordination. (3) Older signers tend to move their head and torso together as a unit whereas younger signers are able to activate their head and torso independently more than older signers, assigning separate functions to each articulator simultaneously. Finally, (4) younger signers sign much faster than older signers, signalling an increase in efficiency in their language.

The expanded use of the spatial axes in young signers is compatible with the finding that younger signers locate different referents and concepts using the additional side-to-side axis (Padden et al. 2010; Meir 2012). Our study also provides a whole-body context for the finding that the use of eye and head signals on relative clauses becomes significantly more systematic and linguistic in younger ISL signers (Dachkovsky 2014). The study confirms that articulator use and linguistic complexity increase in tandem across generations, and that the bodily organization of articulators in sign languages is a key to the organization of emergent linguistic structure.

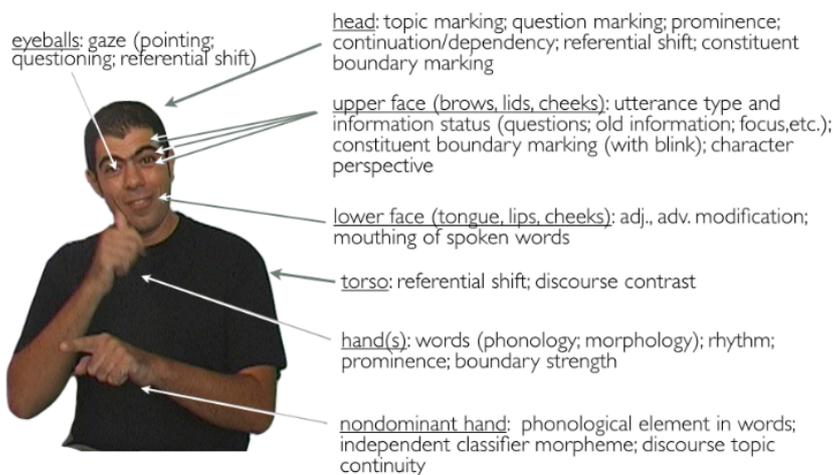


Figure 1: Body articulators recruited for linguistic functions (following Sandler 2012).

References

- Dachkovsky, S. (2014). *Grammaticalisation of facial intonation: the case of squint in ISL relative clauses*. Talk presented at the 30th Annual Meeting of the Israel Association of Theoretical Linguistics. Beer-Sheva, October 20-21.
- Dachkovsky, S. & Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Language and Speech*, 52(2/3), 287-314.
- Jackendoff, R. (2011). What is the human language faculty?: Two views. *Language*, 87(3), 586-624.
- Krifka, M. (2001). Compositionality. In Wilson, R.A. and Keil F. (Eds.), *The MIT Encyclopedia of Cognitive Sciences*. 152-3.
- Liddell, S. K. (1980). *American Sign Language Syntax*. The Hague: Mouton.
- Lillo-Martin, D. (1995). The point of view predicate in American Sign Language. In K. Emmorey and J. Reilly (Eds.), *Language, Gesture, and Space*, (pp. 155-170). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Meir, I. (2012). The evolution of verb classes and verb agreement in signed languages. *Theoretical Linguistics*, 38(1-2), 145 – 152.
- Metzger, M., (1995). Constructed dialogue and constructed action in American Sign Language. In C. Lucas (Ed.), *Sociolinguistics in Deaf Communities* (pp. 255-271). Washington, DC: Gallaudet University Press.
- Nespor, M. & Sandler, W. (1999). Prosody in Israeli Sign Language. *Language and Speech*, 42(2/3), 143-176.
- Padden, C., Meir, I., Aronoff, M., & Sandler, W. (2010). The grammar of space in two new sign languages. In D. Brentari (Ed.), *Sign Languages: A Cambridge Survey* (pp. 573-595). Cambridge: Cambridge University Press.
- Pfau, R., Steinbach, M., and Woll, B. (2012). *Sign Language: An International Handbook*. Berlin: De Gruyter.
- Sandler, W. (2012). Dedicated gestures and the emergence of sign language. *Gesture*, 12:3, 265-307.
- Sandler, W. (to appear). What comes first in language emergence? In Enfield, N. (Ed.), *Dependencies in Language: On the Causal Ontology of Linguistic Systems*. Language Science Press, Studies in Diversity Linguistics Series.
- van der Kooij, E. & Crasborn, O. (2006). Explaining prosodic body leans in Sign Language of the Netherlands: Pragmatics required. *Journal of Pragmatics*, 38, 1598-1614.
- Smith, K. & Kirby, S. (2012). Compositionality and linguistic evolution. Wernig, M, Hinzen, W, and Machery, E. (Eds.), *The Oxford Handbook of Compositionality*, Oxford: Oxford University Press. 493-509.

FAILURES OF PERSPECTIVE TAKING IN AN OPEN-ENDED SIGNALING TASK

JUSTIN SULIK & GARY LUPYAN

*Department of Psychology, University of Wisconsin-Madison
Madison, WI USA
jsulik@wisc.edu lupyan@wisc.edu*

Imagine needing to communicate some meaning in the absence of a shared, conventional signal. A modern human might do this in a number of ways. Lacking the word *snake*, for example, one might vocally imitate its hiss, or gesturally imitate its slithering, or its biting strike. One could draw a simple stick-figure, or, if speaker and listener share conventional signals other than *snake*, one could say something like ‘legless reptile’. Each of these choices foregrounds — makes salient — a different feature of the snake: its hiss, its movement, its bite, its anatomy, its taxonomy.

Some of these choices about salience will be better than others in conveying the meaning to the interpreter. If the slithering gesture were to make most people guess *fish*, the signal would be a poor choice for communicating about snakes. How does a signaler select a signal in the absence of convention? In particular, what information drives the inference about which out of several potentially salient features is most likely to lead to successful communication? Understanding how signaling occurs in the absence of convention is crucial for understanding the origins of convention (Cubitt & Sugden, 2003), and ultimately for understanding the evolution of language with its reliance on conventional signals.

According to several influential theories (Lewis, 1969; Sperber & Wilson, 1995), people are able to take their interlocutor’s point of view into account when deciding how to signal, inferring either what would be salient from an interlocutor’s perspective, or what information would be relevant to them. We experimentally tested the assumption that people are able to use information about salience or relevance from another’s point of view using a word-guessing game.

In the game, a signaler is given an item, such as *bank*. He/she has to think of a one-word signal to help a guesser guess the item. A very good signal in this case is *teller* because most people guess *bank* given *teller* (Nelson, McEvoy, & Schreiber, 1998). On the other hand, *money* is a poor choice because very few people guess *bank* given *money* (Nelson et al., 1998). The challenge of choosing a good signal is twofold. First, communication of this sort is inherently asymmetric

in the sense that the signaler is **given** *bank*, while the guesser must **infer** *bank*. In addition, salience is often asymmetric, in the sense that *money* is likely to occur to a signaler given the item *bank*, but *bank* isn't likely to occur to the guesser given the signal *money* (Nelson et al., 1998). The question, then, is whether the signaler is able to override the comparatively high salience of *money* from his/her point of view to choose a signal that is more informative from the guesser's point of view, such as *teller*.

Our results show that signalers are more likely to use information about salience from their own perspective than the guesser's perspective in an unconstrained task like the one just described. This leads to low communicative success. For example, 40% of signalers chose *money* to signal *bank* — the most common choice — while 0% chose *teller* (we use this example as an illustration; the experiments used many other items). In an unconstrained task like this, signalers are quite poor in using information from the guesser's perspective, and guessers are even worse at inferring salience from the signaler's perspective.

In a second study we show that communicators do have access to information about salience from the opposite perspective, but they do not access this information outside of tightly constrained contexts. For example, when given a list of 5 potential signals including *money* and *teller*, and asked to pick which would be most likely to help someone guess *bank*, 42% now chose *teller* and just 25% *money*. In a third study, we show that contextual information can promote perspective taking if the context is clearly shared between signaler and guesser. Participants were given a list of items, all of which made *money* salient. They were then asked to help the guesser pick *bank*. This partially inhibited their choosing *money* as a signal for *bank* if told that the guesser also had the list.

In sum, the results show that in a novel signaling task, people are sometimes able to take another's perspective on salience, but this is difficult and achieved only under specific conditions. We give examples of such conditions and conclude that these impose severe limitations on any theory that relies on inference in perspective taking to explain the emergence of successful linguistic conventions.

References

- Cubitt, R. P., & Sugden, R. (2003). Common knowledge, salience and convention: A reconstruction of David Lewis' game theory. *Economics and Philosophy*, 19, 175-210.
- Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. <http://w3.usf.edu/FreeAssociation/>. (Last checked 14 January 2016)
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Malden, MA: Blackwell Publishing.

WHAT IS UNIQUE ABOUT THE EVOLUTION OF LANGUAGE COMPARED TO OTHER CULTURAL DOMAINS? AN EXPERIMENTAL STUDY OF LANGUAGE, TECHNOLOGY AND ART

MÓNICA TAMARIZ, JON W. CARR

*Language Evolution and Computation, PPLS, The University of Edinburgh
Edinburgh, UK*

monicatamariz@gmail.com, j.w.carr@ed.ac.uk

A comparative approach examining the differences and similarities between the evolution of language and cultural evolution in other human cultural domains (e.g. technology, art, social and political institutions) would help us understand how our cognitive biases interact with different human needs and affordances to produce the astonishing cultural diversity observed in our species. While between-species comparative studies of cultural evolution have deservedly received much attention (e.g. Horner, Whiten, Flynn & de Waal 2006), between-domain comparative studies within humans hardly exist. Instead, studies of cultural evolution in various human cultural domains have proceeded in parallel. Here we present an experimental approach to comparative cross-domain cultural evolution. Using a new paradigm, we find that evolutionary variables such as fidelity of transmission and selection are differentially affected by the *functions* inherent to three cultural domains: language, technology and art.

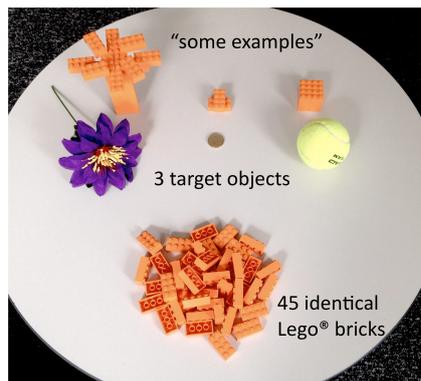


Figure 1.

We recruited 96 participants who were randomly assigned to the Art, Language, or Technology condition. Participants were presented with 45 identical 2x4 loose orange Lego bricks, a set of target objects, and an example Lego construction standing next to each object (see Fig. 1). Participants were then given five minutes to produce a new Lego construction for each object. In the Language condition, participants had to create signals communicate

the target objects to a partner participant in a naming game (a task similar to Garrod, Fay, Lee, Oberlander & MacLeod, 2007). In the Technology condition, they had to create three towers to raise the objects as high as possible off the table (a task similar to that in Caldwell & Millen, 2008). In the Art condition, they had to create three pieces of art inspired by the target objects. She was not given any instructions as to what to do with the examples. In each condition, we ran four chains of six generations. The examples shown to Generation i in a given chain were the constructions built by Generation $(i-1)$ in that chain; the initial generation received no examples.

The Lego constructions produced were coded for Height and Complexity (among others). Mutation was quantified by taking the average of three rater judgements of the dissimilarity between constructions of the same object in consecutive generations of the same chain. These values were submitted to repeated-measures ANOVAs, which returned significant effects of cultural domain on Complexity (linguistic signals were less complex than artworks or towers); Mutation (mutation was smallest in language and greatest in art); and Height (towers were taller than the linguistic or artistic constructions). These results clearly reflect different function-related evolutionary responses across domains: the artistic function encourages mutation, communication encourages fidelity of transmission and simpler signals, and the technological function favours innovation on previous ideas.

We believe this new cross-domain comparative approach complements existing lines of work by exploring how linguistic function specifically influences language evolution.

References

- Caldwell, C.A., & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, *29*, 165–171.
- Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*, 961–987.
- Horner, V., Whiten, A., Flynn, E., & de Waal, F.B.M. (2006). Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Sciences of the USA*, *103*, 13878–13883.

EVIDENCE OF DESCENT WITH MODIFICATION AND SELECTION IN ITERATED LEARNING EXPERIMENTS

MÓNICA TAMARIZ

*Language Evolution and Computation, PPLS, The University of Edinburgh, UK.
monica@ling.ed.ac.uk.*

JOLEANA SHURLEY

Dept. of Linguistics, UC Santa Barbara, USA. jshurley_12@live.com.

Iterated learning experiments claim to model the evolution of languages. We report a study that tests to what extent the languages in these experiments actually follow evolutionary dynamics. Specifically, we look for the signature of (1) descent with modification, leading to diversity, and (2) selection, or adaptation to environmental factors, as languages change over generations.

We ran a modified design of the classic study by Kirby et al. 2008 (experiment 2). In that experiment, individuals organized in diffusion chains were trained on 50% of a language (excluding homonyms) and then had to produce labels for 100% of the meanings. The languages showed a gradual decrease in learning error and an increase in systematic structure over generations. The relevant novelty in our design was the tree-like structure of the chains (Fig. 1), designed to allow us to obtain a set of related languages with a known phylogeny. Like Kirby et al. (2008), we also used a 3x3x3 meaning space, but each of our meanings was a scene where a *subject* gives a *direct object* to an *indirect object*.

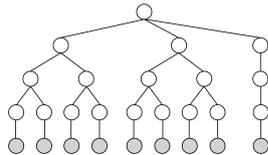


Figure 1. The tree-like design of our iterated learning experiment. Starting with a single random language (top circle), we ran diffusion chains to obtain a family of languages. Each circle represents a language, and each line, a participant.

1. Phylogenetic analysis finds evidence of descent with modification

In order to check whether the changes in the languages are consistent with descent with modification, we constructed phylogenetic trees based on the pairwise edit-distances between languages. Phylogenetic tree building works on the assumption that change is the result of descent with modification, so if our constructed trees are similar to the veridical phylogenetic tree, the hypothesis that the languages have changed through these mechanisms is supported. We

reconstructed two trees: tree A, based on the distances between final-generation language only (the shaded circles in Fig. 1), and tree B, based on distances between all languages. Monte Carlo analyses found that the constructed trees were significantly more similar to the veridical tree than expected by chance ($z\text{-score}_A=2.50$; $p_A<0.01$; $z\text{-score}_B=8.60$; $p_B<0.001$), strongly indicating that change in our language family is, indeed, the outcome of descent with modification.

2. Mantel tests and spectral analysis find evidence of selection

We also found three quantitative signatures of selection in the languages. First, Mantel tests on the languages revealed a consistent, directional increase in systematic structure over generations, indicative of adaptation of the languages to being easy to learn and expressive (Kirby et al. 2008). Second, partial Mantel tests looking separately at the initial and final bigrams of the signals quantitatively confirmed the emergence of division of labour in some of the languages. In one case, for instance, word beginnings expressed the subject, and word endings, the indirect object. This hints at a process of adaptation of the structure of signals to the structure of meanings. Third, in order to explore further the selective pressures of meaning on signals we used Monte Carlo tests and spectral analysis (Tamariz 2011). This analysis tests whether the frequencies of elements of the signals evolve randomly over generations, following neutral evolutionary (or drift) dynamics, or come to resemble the frequencies of meanings, consistent with selection. The spectra of the ngrams that formed the signals of early languages were indistinguishable from spectra of randomly produced languages; however, the spectra of the languages produced at later generations showed significant drops in low frequencies and peaks of high frequencies that mirrored the (nonrandom) frequency structure of the meanings compared to randomly produced languages. These are significant departures from the predictions of the drift model, which are strongly indicative of the adaptation of the language structure under selective pressure originating in the structure of the meanings.

These analyses together show that iterated learning experiments are appropriate models of evolution by descent with modification and selection.

References

- Kirby S, Cornish H & Smith K. 2008. Cumulative cultural evolution in the lab: An experimental approach to the origins of structure in human language. *PNAS*, 105 (31): 10681–10686.
- Tamariz M. 2011. Linguistic structure evolves to match meaning structure. In Carlson L, Hölscher C, Shipley T (eds), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

INTERPRETING SILENT GESTURE

BILL THOMPSON

*Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, Belgium,
bill@ai.vub.ac.be*

MARIEKE SCHOUWSTRA

*Language Evolution and Computation Research Unit, The University of Edinburgh,
Edinburgh, Scotland, Marieke.Schouwstra@ed.ac.uk*

HENRIETTE DE SWART

*Utrecht Institute of Linguistics & Departement Moderne talen, Universiteit Utrecht,
Utrecht, Netherlands, h.deswart@uu.nl*

Silent gesture, a lab methodology in which adult hearing participants describe simple events using only their hands, has proven to be a valuable window into the origins of word order (the ordering of Subject, Object and Verb) in language. Recent experiments using this methodology have uncovered word order biases in silent gesture *production*. People prefer SOV order when they describe extensional transitive events (*boy-pale-swing*, (Goldin-Meadow, So, Özyürek, & Mylander, 2008)), but prefer SVO order for events with different semantic properties, such as intensional events: *man-think of-ball*, *witch-build-house* (Schouwstra & de Swart, 2014). However, a core underexplored question (but see, e.g., Hall et al., 2015) is whether these biases also feature in silent gesture *interpretation*: when people interpret silent gesture strings, will they behave similarly to when they produce them? In this presentation, we describe experimental and computational analyses of silent gesture interpretation: first, in a silent gesture experiment we examine interpretation of the semantically conditioned word order variation found in the production experiments of Schouwstra and de Swart (2014); second, we develop a Bayesian computational model of silent gesture interpretation, and fit the model to our experimental data.

In our experiment, we recorded silent gestures of ambiguous actions (e.g., a gesture that could mean *build* as well as *climb*). For each ambiguous action, we composed ambiguous gesture strings describing transitive events: one in SVO and one in SOV order (e.g., *witch-climb/build-house* and *witch-house-climb/build*). We predicted that participants would be more likely to interpret SVO ordered videos as intensional events than as extensional events, and vice versa. Forty

one Dutch and forty Turkish participants watched twelve videos (6 in SVO and 6 in SOV order; we used two different versions to make sure that all participants saw each video in only one order), and were asked, after each video, to choose an interpretation from two line drawings of the two target events (one intensional and one extensional). We found that for SOV sequences, an extensional interpretation was given significantly more often ($M=.711$, $SE=.019$) than for SVO ($M=.569$, $SE=.020$), and the native language of the participants did not influence this.

Participants used word order as a key to the semantic distinction between intensional and extensional events: their word order preferences are semantically conditioned, as in silent gesture production. However, a key aspect of our findings is that this effect is weaker in interpretation than in production. What does this tell us about word order biases? We develop a computational model of the experimental task, based on the idea that silent gesture interpretation can be understood as inductive inference under uncertainty. Modelling interpretation as Bayesian inference allows us to explicitly specify the influence of word order preferences – or *prior beliefs* – on judgements about the unseen intentions of another gesturer. Our model assumes participants account rationally for the uncertainty that surrounds the gesturer’s usage of word order in the absence of labelled examples – consistent with the idea that the learner is entertaining emerging linguistic rules. By fitting the model to participants’ interpretations, we infer a quantitative estimate of their underlying inductive biases, which suggests skewed and asymmetric – but crucially defeasible – event class conditional ordering preferences. The biased model (M1), but not an unbiased alternative (M0), predicts the experimental data well. Furthermore, despite their weakened effect on interpretation, the biases we infer align well with results from production experiments, suggesting silent gesture interpretation is underpinned by computational principles that balance word order biases with communicative uncertainty. Only by understanding the interplay of biases in production and interpretation, and the computational principles that govern these processes during learning and communication, can we begin to understand the emergence of word order regularities in human language.

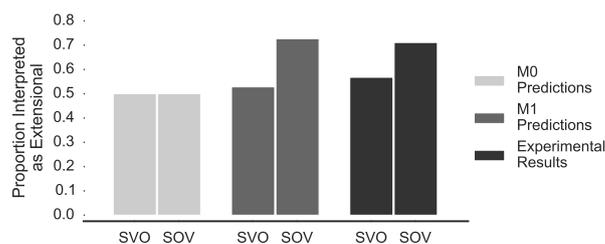


Figure 1. Experimental results (right), alongside predictions of the unbiased (M0, left) and biased (M1, middle) models. Bars show proportion of gestures interpreted as Extensional events.

References

- Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *PNAS*, *105*(27), 9163–9168.
- Hall, M. L., Ahn, Y. D., Mayberry, R. I., & Ferreira, V. S. (2015). Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime. *Journal of memory and language*, *81*, 16–33.
- Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436.

ARBITRARINESS OF ICONICITY: THE SOURCES (AND FORCES) OF (DIS)SIMILARITIES IN ICONIC REPRESENTATIONS

OKSANA TKACHMAN, CARLA L. HUDSON KAM
*Department of Linguistics, The University of British Columbia,
Vancouver, Canada
o.tkachman@alumni.ubc.ca, Carla.HudsonKam@ubc.ca*

Unrelated sign languages have more overlap in their form and structures than unrelated spoken languages, and this overlap has often been attributed to properties of the visual-manual modality that enable or even encourage iconic forms (see Perniss, Thompson, & Vigliocco, 2010, for discussion). Clearly, iconicity plays an important role in the development and evolution of signed languages. However, iconicity is a much more complex phenomenon than seems to be generally assumed. In particular, there is no single ‘iconicity’, there are many (Tolar, Ledeborg, Gokhale, & Tomasello, 2008). Signs can be based on culturally-specific (i.e., learned) relationships: for instance, EAT utilizes a grasping gesture in many Western sign languages and a V-handshape for chopsticks in many East Asian sign languages. Signs can also differ in which features of a referent are iconically represented. For example, a cat is referred to by whiskers in American Sign Language, by licking paws in Al-Sayyid Sign Language and by petting in Swedish Sign Language. However, note that even in the differences there are similarities: the signs for eat represent the action involved in prototypical eating events in the culture, including the tool(s) used, whereas the signs for cat more frequently represent some feature of the animal itself.

Our study investigates factors that might lead to favoring some features of referents over others in iconic representations. We investigate this by having hearing, sign-naïve adult participants invent gestured names for easily recognizable objects. The items participants were asked to create signs for differed along a number of dimensions that we hypothesize might influence the nature of the iconic representation, as shown in Figure 1. For instance, some of the items were man-made while others were part of the natural world, as it has been claimed that man-made objects are represented with handling (grasping) handshapes (Padden et al., 2013). We also investigated the effect of movement and size, for both man-made and natural categories. We anticipated that these

categories would have impact on the choice of representational features; for example, the size and shape of natural objects would be encoded in the gestures, and the man-made objects would be represented by the prototypical interaction of humans with those objects.

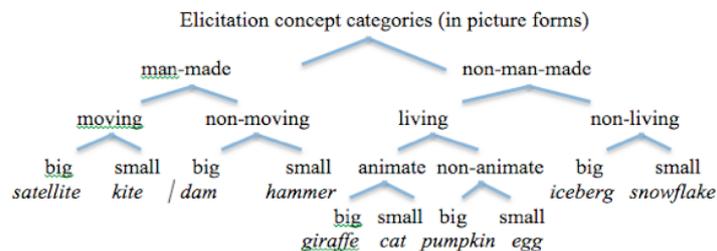


Figure 1. Elicitation item types with examples.

50 native speakers of English with no knowledge of sign languages, ages 18-72, participated in the study. They saw 110 pictures of familiar objects and were asked to 'name' them with their hands. Responses were videotaped. Each response is currently being coded for the type of iconic information encoded, specifically, whether the invented sign encodes referent shape, characteristic movement, or human handling of the object.

This study helps us better understand the roots of iconic representations and the forces that might shape the specific information encoded in iconic signs.

Acknowledgements

This research was funded by a Standard Research Grant from the Social Sciences and Humanities Research Council of Canada to C. L. Hudson Kam.

References

- Padden, C. A., Meir, I., Hwang, S. O., Lopic, R., Seegers, S., & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture, 13*, 287-308.
- Perniss, P., Özyürek, A., & Morgan, G. (2015). The influence of the visual modality on language structure and conventionalization: insights from sign language and gesture. *Topics in Cognitive Science, 7* (1), 2-11.
- Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Language Sciences, 1*, 227.
- Tolar, T.D., Lederberg, A.R., Gokhale, S., & Tomasello, M. (2008). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education, 13*, 225-240.

Experimental Evidence for Phonemic-like Contrasts in a Nonhuman Vocal System

SIMON W. TOWNSEND

*Department of Psychology, University of Warwick
Coventry, United Kingdom
simon.w.townsend@warwick.ac.uk*

ANDREW F. RUSSELL

*Department of Biosciences, University of Exeter, United Kingdom
a.russell@exeter.ac.uk*

SABRINA ENGESSER

*Department of Evolutionary Biology and Environmental Studies, University of Zurich
Zurich, Switzerland
sabrina.engesser@ieu.uzh.ch*

The capacity to generate new meaning through rearranging combinations of meaningless sounds, so-called phoneme structuring, is a fundamental component of language and is central to its productive nature (Hurford 2011). Despite its importance, surprisingly little is known about how unique this capacity is to humans or indeed the evolutionary steps that characterised its emergence. Animal vocalizations have been shown to often comprise combinations of meaningless acoustic elements, for example humpback whales (*Megaptera novaeangliae*), gibbons (*Hylobates spec.*), and a number of passerine birds, are capable of constructing elaborate and meaningful “song” vocalisations from a variety of meaningless call elements. A hierarchical structure is ultimately achieved through an assembling of the single units in a potentially rule-governed way, in some cases reaching a considerable level of complexity (Payne and McVay 1971, Clarke et al. 2006, Berwick et al. 2012). However, evidence that rearranging such combinations generates functionally distinct meaning is

lacking (Berwick et al. 2011). Here we provide evidence for this basic ability in calls of the chestnut-crowned babbler (*Pomatostomus ruficeps*), a highly social bird of the Australian arid zone. Using acoustic analyses, natural observations and a series of controlled playback experiments, we demonstrate that this species uses the same acoustically distinct elements (*A* and *B*) in different arrangements (*AB* or *BAB*) to create two functionally distinct vocalizations: the flight call (used during movement, *AB*) and the prompt call (used when provisioning nestlings, *BAB*). Specifically, the addition or omission of a contextually meaningless acoustic element at a single position generates a phoneme-like contrast that is sufficient to distinguish the meaning between the two calls. Our results indicate that the capacity to rearrange meaningless sounds in order to create new signals occurs outside of humans. We discuss the implications of our data for understanding the evolutionary progression of phoneme structuring and suggest that basic phonemic contrasts represent a rudimentary form of phoneme structure, and a potential early step towards the generative phonemic system of human language.

References

- Berwick, R.C., Okanoya, K., Beckers, G. I. J. L., & Bolhuis, J. J. (2011). Songs to Syntax: the Linguistics of Birdsong. *Trends in Cognitive Sciences*, 15(3):113–121.
- Berwick R.C., Beckers, G.J.L., Okanoya, K. & Bolhuis, J.J. (2012). A Bird's Eye View of Human Language Evolution. *Frontiers in Evolutionary Neuroscience* 4: 5. pmid:22518103
- Clarke, E., Reichard, U. H. & Zuberbühler, K. (2006). The Syntax and Meaning of Wild Gibbon Songs. *PLoS ONE* 1(1): e73.
- Hurford, J.R. (2011). *The Origins of Grammar: Language in the Light of Evolution*. Oxford University Press.
- Payne, R.S. & McVay, S. (1971). Songs of Humpback Whales. *Science*, 173(3997):585–597.

MODELING THE EMERGENCE OF CREOLE LANGUAGES

FRANCESCA TRIA

*ISI Foundation
Turin, ITALY
fratrig@gmail.com*

VITTORIO LORETO

*Physics Dept., Sapienza University of Rome
Rome, ITALY
&
ISI Foundation
Turin, ITALY
vittorio.loreto@roma1.infn.it*

SALIKOKO MUFWENE

*Dept. of Linguistics, University of Chicago
Chicago, IL 60637, USA
s-mufwene@uchicago.edu*

VITO D.P. SERVEDIO

*Institute for Complex Systems, CNR
Rome, ITALY
Vito.Servedio@roma1.infn.it*

Creole languages (Reinecke, 1975; Holm, 2000; Holm & Michaelis, 2009; Michaelis, Maurer, Haspelmath, & Huber, 2013) offer an invaluable opportunity to study the processes leading to the emergence and evolution of Language, thanks to the short - typically a few generations - and reasonably well defined time-scales involved in their emergence. Another well-known case of a very fast emergence of a Language, though referring to a much smaller population size and different ecological conditions, is that of the Nicaraguan Sign Language (Senghas, Kita, & Özyürek, 2004). What these two phenomena have in common is that in both cases what is emerging is a contact language, i.e., a language born out of the non-trivial interaction of two (or more) parent languages. This is a typical case of what is known in biology as horizontal transmission. In many well-documented cases, creoles emerged in large segregated sugarcane or rice plantations on which the

slave laborers were the overwhelming majority. Lacking a common substrate language, slaves were naturally brought to shift to the economically and politically dominant European language (often referred to as the lexifier) to bootstrap an effective communication system among themselves (Chaudenson, 2001; Mufwene, 2001). Here, we focus on the emergence of creole languages originated in the contacts of European colonists and slaves during the 17th and 18th centuries in exogenous plantation colonies of especially the Atlantic and Indian Ocean, where detailed census data are available. Those for several States of USA can be found at <http://www.census.gov/history>, while for Central America and the Caribbean can be found at <http://www.jamaicanfamilysearch.com/Samples/1790a111.htm>. Without entering in the details of the creole formation at a fine-grained linguistic level, we aim at uncovering some of the general mechanisms that determine the emergence of contact languages, and that successfully apply to the case of creole formation.

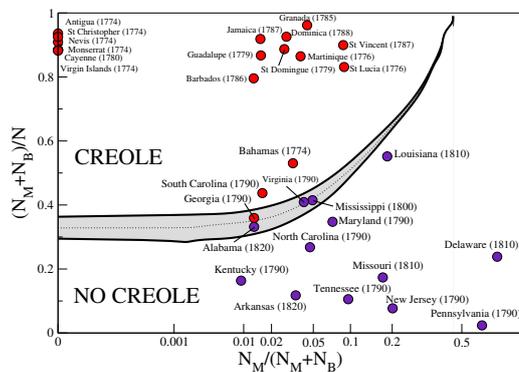


Figure 1. **Census data and model predictions.** We refer to census data about the latest decades of the 18th century and the beginning of the 19th, reporting the number of Free Whites, Free Creole Blacks, and Bozal Blacks B in several States of USA and the Caribbean. Points are the projection of the census data in a bidimensional plane. Red circles mark States where a creole language emerged while purple ones identify States where a creole language historically did not emerge. The gray stripe is the outcome of our modeling scheme and separates the regions where respectively the creole C (above the stripe) and the European E (below the stripe) represent the dominant language (i.e., spread among more than the 80% of the population) in the Mulattos and Bozal populations in the asymptotic states of the model.

We demonstrate a dynamical processes leading to the emergence and stabilization of creole languages, suggesting ways in which modeling can be used as a research tool to clarify accounts of where creoles emerged and what specific ecological factors explain why they did not emerge elsewhere. We judged the

language games (Baronchelli, Felici, Caglioti, Loreto, & Steels, 2006; Loreto & Steels, 2007; Puglisi, Baronchelli, & Loreto, 2008; Loreto, Baronchelli, Mukherjee, Puglisi, & Tria, 2011; Tria, Galantucci, & Loreto, 2012) framework as particularly suitable for this task since it simulates how a population of individuals can bootstrap linguistic consensus—on cultural timescale—out of the local interactions of pairs of individuals. Inspired by the Naming Game (NG) (Baronchelli et al., 2006), our modeling scheme (Tria, Servedio, Mufwene, & Loreto, 2015) incorporates demographic information about the colonial population in the framework of a non-trivial interaction network including three populations (Pucci, Gravino, & Servedio, 2014): Europeans, Mulattos/Creoles, and Bozal slaves. We show how this sole information makes it possible to discriminate territories that produced modern creoles from those that did not, with a surprising accuracy (Fig. 1). We submit that these tools could be relevant to addressing problems related to contact phenomena in many cultural domains (Weinreich, 1963; Thomason, 2001; Bakker & Matras, 2013; Castellano, Fortunato, & Loreto, 2009): e.g., emergence of dialects, language competition and hybridization processes like those undergone by languages with an important number of non-native speakers and more generally all processes where different cultural features come in contact. Thanks to its flexibility, we believe our modeling scheme can be easily modified and adapted to very different contact ecologies, and in this sense it could become an important tool of investigation for scholars interested in testing specific hypotheses about the emergence and evolution of Language.

References

- Bakker, P., & Matras, Y. (2013). *Contact languages: A comprehensive guide*. De Gruyter Mouton.
- Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., & Steels, L. (2006). Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics*, P06014.
- Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81, 591-646.
- Chaudenson, R. (2001). *Creolization of language and culture*. New York: Routledge - Taylor and Francis -.
- Holm, J., & Michaelis, S. M. (Eds.). (2009). *Contact languages: Critical concepts in linguistics*. Routledge.
- Holm, J. A. (2000). *An introduction to pidgins and creoles*. Cambridge University Press.
- Loreto, V., Baronchelli, A., Mukherjee, A., Puglisi, A., & Tria, F. (2011). Statistical physics of language dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(04), P04006.
- Loreto, V., & Steels, L. (2007). Social dynamics: the emergence of language. *Nat. Phys.*, 3, 758-760.

- Michaelis, S. M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.). (2013). *Apics online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Mufwene, S. S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Pucci, L., Gravino, P., & Servedio, V. D. P. (2014). Modeling the emergence of a new language: Naming game with hybridization. In W. Elmenreich, F. Dressler, & V. Loreto (Eds.), *Self-organizing systems* (Vol. 8221, p. 78-89). Springer Berlin Heidelberg.
- Puglisi, A., Baronchelli, A., & Loreto, V. (2008). Cultural route to the emergence of linguistic categories. *Proc. Natl. Acad. Sci. USA (PNAS)*, *105*, 7936-7940.
- Reinecke, J. (1975). *A bibliography of pidgin and creole languages*. University Press of Hawaii.
- Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, *305*(5691), 1779-1782.
- Thomason, S. (2001). *Language contact*. Edinburgh University Press.
- Tria, F., Galantucci, B., & Loreto, V. (2012). Naming a structured world: A cultural route to duality of patterning. *PLoS ONE*, *7*(6), e37744.
- Tria, F., Servedio, V. D., Mufwene, S. S., & Loreto, V. (2015). Modeling the emergence of contact languages. *PLoS ONE*, *10*(4), e0120771.
- Weinreich, U. (1963). *Languages in contact: findings and problems*. Mouton.

DENDROPHOBIA IN BONOBO COMPREHENSION OF SPOKEN ENGLISH

ROBERT TRUSWELL

*Linguistics and English Language, University of Edinburgh
Edinburgh, UK
rob.truswell@ed.ac.uk*

Fitch (2014) proposed the *Dendrophilia hypothesis* as a description of the ubiquity of hierarchical structures in human cognition:

‘Humans have a multi-domain capacity and proclivity to infer tree structures from strings, to a degree that is difficult or impossible for most non-human animal species.’ (Fitch, 2014, 352)

Part of Fitch’s supporting evidence concerns Fitch and Hauser’s (2004) demonstration that humans learn to recognize sequences of the forms $(ab)^n$ and $a^n b^n$, while cotton-top tamarins can only learn the former. Fitch takes this to support the Dendrophilia hypothesis because $(ab)^n$, but not $a^n b^n$, can be generated in the limit by constituency-free finite-state grammars. However, this result has been disputed on empirical and theoretical grounds, e.g. Perruchet and Rey (2005), Jäger and Rogers (2012).

This paper gives a complementary source of evidence for Fitch’s hypothesis. We examine a corpus from (Savage-Rumbaugh et al., 1993) of 660 utterances directed in parallel to a bonobo, Kanzi, and a human infant, Alia, together with descriptions of their behavior in response to those utterances. Unlike grammar induction experiments such as Fitch and Hauser (2004), these strings are paired with interpretations. We can then infer aspects of a subject’s interpretation of an utterance from their behavior, and aspects of the grammatical representation of the utterance from that interpretation. I argue that Kanzi fails to respond to requests precisely where correct interpretation requires hierarchical constituency.

Kanzi’s overall performance across the corpus (71.5% ‘correct’ responses according to Savage-Rumbaugh et al.’s criteria) is comparable to Alia’s (66.6% ‘correct’). Usually, though, a correct response could be achieved through common-sense combination of the concepts expressed by individual words, without using syntactic information (Anderson’s 2004 *semantic soup* strategy). One such example, carried out correctly by Kanzi, is *Put the backpack in the car*: few other actions involving backpacks and cars suggest themselves.

In some cases (e.g. *Put the tomato in the oil / Put some oil in the tomato*), correct interpretation requires sensitivity to linear order, but not constituency. Kanzi's accuracy on 43 such sentences in the corpus (21 pairs, with 1 example repeated) is 76.7%, in line with his 71.5% overall accuracy. This suggests that Kanzi can make use of linear order information in his understanding of spoken English.

However, Kanzi responded correctly to only 4/18 sentences containing coordinated NP objects (22.2%). When asked to *Show me the milk and the doggie*, he shows only the dog; when asked to *Give the lighter and the shoe to Rose*, he gives Rose only the lighter. Kanzi ignores the first conjunct on 9/18 trials, and ignores the second conjunct on 5/18 trials.

Despite the small number of critical sentences, this represents a highly significant drop relative to both Kanzi's baseline accuracy ($p < 10^{-4}$) and Alia's 68.4% accuracy on sentences containing the same construction ($p < 0.01$). This, then, is a species-specific, construction-specific drop in performance.

I propose that Kanzi's performance dips precisely here (and not on many other constructions of comparable length) because correct interpretation of such sentences requires reference to hierarchical constituent structure. Specifically, unlike the previous examples, Kanzi must recognize that the object of *give* is the complex phrase *the water and the doggie*, and not just, for example, the next noun. Likewise, the patient of the action of giving should be the group of objects denoted by the complex phrase, not just the denotaton of a single noun. Kanzi's generally impressive performance therefore only drops where reference to constituency is required, while Alia has no similar problem. In Fitch's terms, Kanzi is more dendrophobic than Alia.

References

- Anderson, S. (2004). *Doctor Doolittle's delusion*. New Haven, CT: Yale University Press.
- Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, 11, 329-364.
- Fitch, W. T., & Hauser, M. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303, 377-380.
- Jäger, G., & Rogers, J. (2012). Formal language theory: Refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B*, 367, 1956-1970.
- Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review*, 12, 307-313.
- Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R., Brakke, K., Williams, S., Rumbaugh, D., & Bates, E. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, 58, 1-252.

A CONSTANT RATE EFFECT WITHOUT STABLE FUNCTIONS

ROBERT TRUSWELL, NIKOLAS GISBORNE

Linguistics and English Language, University of Edinburgh
Edinburgh, UK
n.gisborne@ed.ac.uk, rob.truswell@ed.ac.uk

Many grammatical changes progress uniformly across linguistic contexts, but with different temporal offsets in different contexts. This is the **Constant Rate Effect**, or CRE (Kroch, 1989). For Kroch, the CRE reflects competition between functionally equivalent forms: within the terms of the standard equation (1) describing S-shaped change, (Bailey, 1973; Blythe & Croft, 2012), a CRE arises when s is constant across contexts, but k varies. This implies a direct link from CREs to **grammar competition**, in which competing forms realize the same function. It situates competition most naturally in adult communicative strategies, and construes such instances of language change as evolution of a population of forms which realize certain communicative functions.

$$\frac{p}{1-p} = e^{k+st} \leftrightarrow p = \frac{e^{k+st}}{1+e^{k+st}}, p \text{ being the probability of a given form} \quad (1)$$

We present a CRE that is better analysed in terms of acquisition rather than use. It concerns **headed *wh*-relatives** in English like *the person [[to whom]_i I spoke ____i]*, with a clause containing a *wh*-phrase modifying a noun phrase.

Headed *wh*-relatives emerged slowly in Middle and Early Modern English, (c. 1100–1700). The first examples had oblique and adverbial *wh*-phrases; argumental relatives with *which* followed by c.1350, with *whom*- and then *who*-relatives emerging in the 15th century. Nevertheless, the rate of change across these linguistic contexts is near-identical. Fig. 1 demonstrates this for relatives with *wh*-PPs and with NP *which*, using data from the Penn Parsed Corpora of Historical English (Kroch & Taylor, 2000; Kroch, Santorini, & Delfs, 2004).^a

Relativizer *which* has been in competition with *that* and \emptyset as strategies for relativizing on argument positions over the last c.650 years. However, there was no competing strategy for relativizing PPs when *wh*-PPs emerged (earlier relatives

^aRegression parameters are as follows: for PPs, $p = e^{0.0052t-11.8}$; for *which*, $p = e^{0.0058t-12.77}$. We omit PP data after 1550 from the regression analysis, because relatives with PP gaps after c.1550 are complicated by the emergence of preposition-stranding in English *wh*-phrases.

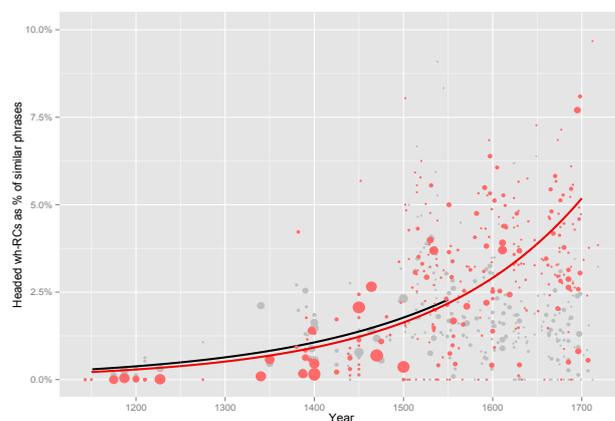


Figure 1. PPs functioning as *wh*-relativizers as a proportion of all PPs (black); subjects and object *which* functioning as *wh*-relativizer as a proportion of all subjects and objects (red). Each circle represents a single text; circle size is proportional to text size.

with demonstrative PP relativizers disappeared in Old English). Accordingly, the constant rate of change across *wh*-PP and *which*-relatives cannot reflect a similar competition process across the two construction types.

Instead, this change reflects competing functional specifications of *wh*-forms. Such competition, unlike Kroch's, is naturally located in acquisition, because a learner identifies a form before inducing a feature specification for it (Shipley, Smith, & Gleitman, 1969). This differs from competition among communicative strategies, but still maintains Kroch's logic of competition and selection.

References

- Bailey, C.-J. (1973). *Variation and linguistic theory*. Arlington, VA: Center for Applied Linguistics.
- Blythe, R., & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, 88, 269-304.
- Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1, 199-244.
- Kroch, A., Santorini, B., & Delfs, L. (2004). *Penn-Helsinki parsed corpus of Early Modern English*.
- Kroch, A., & Taylor, A. (2000). *Penn-Helsinki parsed corpus of Middle English (2nd edition)*.
- Shipley, E., Smith, C., & Gleitman, L. (1969). A study in the acquisition of language: Free responses to commands. *Language*, 45, 322-342.

NORMS FOR CONSTRUCTING LANGUAGE IN HUMANS AND ANIMALS

ROBERT ULLRICH

*Comparative Developmental Psychology, Freie Universität Berlin
Berlin, Germany
robert.ullrich@fu-berlin.de*

1. The normativity of background assumptions

When writing about the ‘language’ of non-human and human animals (henceforth: ‘animal’ and ‘human’), scientists apply the same term to different concepts. Some ascribe ‘language’ for communication others for thinking (e.g. Evans & Marler, 1995, pp. 370 – 374 vs. Scott-Phillips, 2015, pp. 79 – 83). The confusion can be explained by spelling out implicit background assumptions to the entire concept. Some assumptions follow normative attitudes. Those will be denoted as ‘norms’.

1.1. *The vocal modality of language: A norm from the recent past*

Norms signify what is ‘commonly done’ (i.e. descriptive norms), or they express what is ‘commonly approved’ (i.e. social norms). Norms may exist independently of behavior or empirical evidence, and in such cases be based on commonly shared values of a group.

Norms are involved in the ascription of behavior to a certain definition of ‘language’. That is why the process of ascription will be denoted as ‘construction’. Until the early 20th century some norms limited the ascription of ‘language’ to certain populations within the human group itself. An elucidatory example is the denial of ‘language state’ to the non-verbal communication system of deaf people. The responsible oral norm led people to identify human ‘language’ with ‘speech’. Physiological conditions were treated as precondition for the faculty. Although the aftermath of that norm are still observable in current publications, the vocal modality lost its normativity in the course of the 20th century in a shift towards a modality-free but still norm containing ‘code model’.

1.2. *Two current normative approaches: The ‘code’ and the ‘mental’*

Since the days of Darwin, there is an influential comparative evolutionary approach to ‘language’. It is often followed by excessive interest on birdsong, as well as the calling and/or gesture behavior of mammals. In this view, to investigate the animal’s ‘code’ is to find out about the biological substrate of human ‘language’. The underlying assumption is that human ‘language’ can also be understood as a system of complex codes. As such ‘language’ must have its gradual, evolutionary precursors. Since the code is the only observable entity of ‘language’ it is the only appropriate way for its investigation (e.g. Evans & Marler, 1995).

In contrast, opponents to this view deny a connection between code and ‘language’. An approach of considering ‘language’ based purely on mental capacity is often followed by interest on metacognition and/or metarepresentations. In this view, to investigate the code of ‘language’ is simply an inappropriate categorical mistake. The code and its communicative function are seen as a mere byproduct of mental capacity. Proponents of this view tend to validate ‘language’ as human unique without evolutionary precursors (e.g. Scott-Phillips, 2015, pp. 45–50).

1.3. *Incompatible facts lead to progress, incompatible norms do not*

By spelling out the normative background assumptions underlying the construction of ‘language’, one might be able to explain the resulting confusion on the use of terminology. Also, it may help to clarify the very role of norms. On one side norms motivate research in opposed areas, thus responsible for detecting incompatible *facts* in competing approaches. On the other hand, expounding on the normative assumptions of language incite discussion on incompatible normative *approaches*. However, as is intended to be demonstrated through this research, the discussion of incompatible *facts*, rather than that of normative *approaches*, can lead to progress in ‘language’ research.

References

- Evans, C. S., & Marler, P. (1995). Language and Animal Communication: Parallels and Contrasts. In H. L. Roitblat & J.-A. Meyer (Eds.), *Comparative Approaches to Cognitive Science* (pp. 341–382). Cambridge: MIT Press.
- Scott-Phillips, T. C. (2015). *Speaking our minds*. London, New York: Palgrave Macmillan.

ADDRESSEES USE ZIPF'S LAW AS A CUE FOR SEMANTICS

FREEK VAN DE VELDE & DIRK PIJPOPS

*Dept. of linguistics, University of Leuven,
Leuven, Belgium*

freek.vandevelde@kuleuven.be, dirk.pijpops@kuleuven.be

1. Zipf's size-meaning correlation

In a number of publications George Kingsley Zipf (1932, 1935, 1949) set out to describe several quantitative tendencies that hold universally in language. Apart from the famous direct inverse correlation between the relative frequency of a word and its rank in a frequency list, Zipf also noted (i) an inverse correlation between the relative frequency of a word and its phonetic substance, and (ii) a correlation between the relative frequency of a word and its level of polysemy. Combining (i) and (ii), an inverse correlation can be deduced between the phonetic substance of a word and its polysemy: shorter words tend to have more meanings (Pustet 2004). We will call this Zipf's size-meaning tendency. Zipf himself saw one over-arching principle behind the phonetic size of words: the principle of least effort. Speakers are under a constant pressure to reduce articulatory effort: frequent, and semantically less specific words can be shortened by speakers. It is often assumed that this benefit for speakers is diametrically opposed to the addressee's interest (Langacker 1977): addressees want clear articulatory distinctiveness, to overcome noise in the signal. In this paper, however, we argue that addressees can benefit from Zipf's size-meaning tendency as well. If they are able to segment the speech signal in words, they are able to work on the correlation between the phonetic size and the semantics (see also Piantadosi et al. 2011). In our study, polysemy was measured using a proxy: cross-linguistically stable asymmetries of marked-unmarked related semantic pairs. The inverse correlation between polysemy and markedness, is grounded in recent work by Winter et al. (2013), who adduce evidence from corpus data and psycholinguistic experiments. The relation between phonetic size and semantic complexity is corroborated by recent work by Lewis & Frank (subm.).

2. Research Design and Findings

We set up an experiment in which 370 native-Dutch speakers (after filtering of problematic responses) were presented with 9 pairs of visual stimuli (drawings), based on Urban's (2011) asymmetrical word-pairs (e.g. SUN – MOON) and Berlin & Kay's color hierarchy (e.g. GREEN – GREY). Together with the visual pairs 2x9 verbal targets were presented, in the form of fake words that differed in phonetic size. Subjects were asked to match the verbal targets to the visual stimuli, under the pretext that the verbal targets were actually attested in children's speech. We only selected Urban pairs of visual stimuli which did *not* conform to the expected size-meaning correlation in Dutch. This 'hyper-conservative' setup was needed to avoid interference from the known language. We put phonotactic constraints on the verbal targets and controlled for phonetic similarity to actual words in Dutch. Visual stimuli were pretested to make sure they were interpreted as intended. Additionally, a number of fillers, both of unrelated word pairs and with equally sized verbal targets were used to mask the actual research goal. Mixed-effect logistic regression was used to detect the impact of factors such as semantic domain (nature, cultural artefacts, color), and the difference in phonetic size on whether subjects coupled the visual stimuli to the verbal targets in conformity with Zipf's size-meaning tendency.

Even with the 'hyper-conservative' research set-up, and with visual stimuli of which language users are often unaware that they show semantic asymmetries, we found that test subjects indeed act on Zipf's size-meaning tendency: they significantly coupled the semantically unmarked visual Urban stimulus to the shorter verbal targets ($p < 0.0001$). We controlled for whether test subjects indicated in the debriefing that they consciously used a Zipf-like heuristic.

The regression analysis showed that the effects, though relatively small, are stronger if the difference in phonetic size in the verbal targets increases, and are more apparent in cultural artefacts (e.g. CAR_{unmarked} – TRAIN_{marked}), nature terms (e.g. SUN_{unmarked} – MOON_{marked}) and color terms (e.g. GREEN_{unmarked} – GREY_{marked}) than in body terminology (e.g. HEART_{unmarked} – BELLY_{marked}).

Our research results support the idea that Zipf's size-meaning tendency is not only under evolutionary selection by speakers, in their attempt to minimize articulatory effort, but benefits addressees as well, who can use this tendency as a cue: through their life-time experience with language, they know that in general, shorter words have more unmarked meanings, and they apply this implicit knowledge when they are confronted with a new language when other cues are absent.

Acknowledgements

We gratefully acknowledge the input of the students who collaborated on this project: Mathias De Backer, Wendy Feytons, Alek Keersmaekers, Veerle Monserez, Thomas Van Hoey, and Marie Vanwingh.

References

- Langacker, R.W. (1977). Syntactic reanalysis. In Ch.N. Li (Ed.), *Mechanisms of syntactic change* (pp. 57-139). Austin: University of Texas Press.
- Lewis, M.L. & M.C. Frank (submitted). The length of words reflects their conceptual complexity.
- Piantadosi, S.T., H. Tily & E. Gibson. 2011. Word lengths are optimized for efficient communication. *PNAS*, 108(9), 3526-3529.
- Pustet, R. (2004). Zipf and his heirs. *Language Sciences*, 26, 1-25.
- Urban, M. (2011). Asymmetries in Overt Marking and Directionality in Semantic Change. *Journal of Historical Linguistics*, 1, 3-47.
- Winter, B., G. Thompson & M. Urban. (2013). Cognitive factors motivating the evolution of word meanings: Evidence from corpora, behavioral data and encyclopedic network structure. In: E. A. Cartmill, S. Roberts, H. Lyn, & H. Cornish (Eds.), *10th International Conference on the Evolution of Language* (pp. 353-360). New Jersey: World Scientific.
- Zipf, G.K. (1932). *Selected Studies of the Principle of Relative Frequency in Language*. Harvard: Harvard University Press.
- Zipf, G.K. (1935). *The Psycho-biology of Language. An Introduction to Dynamic Philology*. Boston: Houghton Mifflin Company.
- Zipf, G.K. (1949). *Human behavior and the principle of least effort*. Cambridge, (Mass.): Addison-Wesley.

**A CONTINUUM OF HUMAN COGNITIVE-LINGUISTIC
EVOLUTION**

OLGA VASILEVA

*Psychology Department, Simon Fraser University,
Vancouver, Canada
ovasilev@sfu.ca*

The current research explores a novel evolutionary hypothesis linking abnormal phenotypes of human cognition to ontogenetic variation in language acquisition styles.

A number of theories suggested that abnormal cognitive development in humans might have evolutionary underpinning with abnormal conditions being byproducts of specific human cognitive adaptations, that can be seen as windows into evolutionary development of our species in general, and language evolution in particular (e.g. Benitez-Burraco & Boeckx, 2014).

Schizophrenia and autism are two conditions that are frequently linked to human cognitive evolution in general and to language evolution in particular. A growing body of evidence suggests that these conditions are not only related, but present potentially opposite sides of a spectrum (Crespi & Badcock, 2008), with some symptoms manifested in the opposite manner in each condition.

Ontogenetic communicative development in humans presents another interesting continuum – so-called language-acquisition styles, with children exhibiting these styles being called “referential” and “expressive”. Cross-cultural research in North America and Russia (e.g. Bates et al., 1991; Dobrova, 2009) indicates that these styles can, first of all, be clearly defined only in a population segment where both clearly defined styles represent extreme points of a continuum; and second of all, main features of the styles are manifested in the opposite manner. For example, referential children are characterized by a fast rate of vocabulary growth and clear speech, while expressive children, on the contrary, by a slower vocabulary growth and less clear speech. While language acquisition styles were first distinguished based mainly on linguistic characteristics, further research showed that expressive and referential children also differ in social factors (socioeconomic status, parental education, family composition), biological factors (brain hemisphere dominance, gender) and certain cognitive parameters (imitation, ability for generalization) These same factors also play a significant role in abnormal cognitive phenotype development. The current submission proposes a link between abnormal phenotypes and language acquisition styles, as a closer analysis suggests that there are some

similarities between abnormal phenotypes and language acquisition styles respectively (schizophrenia – expressive style; autism – referential). The hypothesis proposes that there are not two independent spectrums of abnormal linguistic phenotypes and language acquisition styles, but a single continuum of human cognitive and linguistic development with abnormal phenotypes at the extreme points of it and normal variation in language acquisition in the central majority of the population. Existing evidence suggests that there are some similarities between abnormal phenotypes and language acquisition styles. For example, autistic children exhibit decreased self-reference; these referential children tend to discuss themselves in third person longer than typical children (delayed first-person reference).

This continuum might be present in both ontogeny and phylogeny, where existing variation can be explained by the interplay between the initial linguistic and cognitive biases common in humans and specific choices that an organism makes based on the cues from the environment for the former and specific biases in human cognition and language developed over the course of species evolution for the latter.

My presentation will review the main properties of autism and schizophrenia in relation to each other and to language evolution; analyze various linguistic, cognitive and social parameters common for two language acquisition styles; and show connections between abnormal phenotypes and their normal counterparts, the language acquisition styles, with precise examples for each step of the analysis. Implications for the evolution of human cognition and language will be discussed.

Acknowledgements

The author is grateful for the continuing guidance of senior supervisor Timothy Racine; and for helpful feedback and unwavering support from colleagues in the Human Evolutionary Studies Program.

References

- Bates, E., Bretherton, I., & Snyder, L. (1991). From first words to grammar: Individual differences and dissociable mechanisms. New York: Cambridge University Press.
- Benítez-Burraco, A. & Boeckx, C. (2014). Language disorders and language evolution: Constraints on hypotheses. *Biological Theory*, 1–6.
- Crespi, B. & Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. *Behavioral and Brain Sciences*, 31, 241-261.

Dobrova, G. R. (2009). On variability of language ontogenesis: Referential and expressive strategies of language acquisition. *Questions in Psycholinguistics*, 2009 (9).

LANGUAGE EVOLUTION IN ONTOGENY AND PHYLOGENY

OLGA VASILEVA

*Psychology Department, Simon Fraser University,
Vancouver, Canada
ovasilev@sfu.ca*

Research in language evolution frequently turns to studies of language development for generating hypotheses about possible evolutionary trajectory of communicative development in humans. One of the problems faced by such hypothetical models is the necessity to explain both universal aspects and individual variation in language development, which is essential for determining the core characteristics of language capacity shared by all humans and estimating the time of their emergence in evolution.

The current submission introduces a potentially fruitful tool, the Theory of Functional Systems (TFS). It was developed by P. Anokhin, and is supported by existing research (e.g., Toomela, 2010; Sudakov, 1997). This theory can contribute to our understanding of language development by examining variation in human communicative and linguistic abilities and its relation to the evolution of our species.

The TFS explains the development of complex behaviors both in evolution and in individuals by treating them and their supporting cognitive substrates as complex systems, and postulating three levels of systemogenesis. In primary systemogenesis, core abilities forming the basis of behavior are created; in accompanying secondary systemogenesis, these abilities are modified through the individual's interaction with its environment and preserved through brain plasticity in ontogenesis. If the environment creates stable selective pressure, certain behavioral systems formed during individual development may prove advantageous, and their maturation in the next generation's ontogenetic development may take place earlier. This third process is referred to as evolutionary systemogenesis.

Evolutionary systemogenesis relies, among other things, on specific mechanisms of gene expression in the nervous system. Growing evidence suggests that genes expressed in an individual's brain at the time of early Central Nervous Systems (CNS) development are often expressed again in the adult brain during various learning experiences. Moreover, the same genes can be involved in different aspects of behavioral development and learning (e.g. Tokarev et al., 2011). TFS has a potential for explaining (i) such genetic - behavioral

interconnection, (ii) the process of transition from individual adult learned behavior over multiple generations into pre-adult species-specific behavior, (iii) the links between phylogeny and ontogeny at the behavioral level, as well as (iv) the time sequence of emergence of certain abilities during development.

Application of TFS to the problem of language evolution requires examining genes associated with language development. For example, research in rodents demonstrated that one of the most prominent language associated genes - FoxP2 - is apparently involved not only in vocal development, but in certain aspects of learning in adult brains (Enard et al., 2009).

In ontogenetic language development, the core systems which have experienced strong evolutionary pressure will mature earlier and build the foundation for the organism's tuning to the specific environment through plasticity and learning, allowing further development of linguistic capacity. Thus, those aspects of language that develop early in ontogeny are most likely to have experienced strong selective pressure during evolution. Aspects of language development that are variable might help us identify specific environmental cues that are essential for an infant's growing linguistic system in secondary systemogenesis, and their developmental timing may unravel the evolutionary history of language. A TFS-based model can explain why human infants are capable of acoustic differentiation of speech sounds well before reaching motor independency and learning other aspects of language. Candidate abilities for primary and secondary systemogenesis are sound discrimination, word-object recognition, gestural communication and language acquisition styles. The ontogeny and phylogeny of language development has been a struggle to elucidate; TFS may greatly advance our understanding of human linguistics.

Acknowledgements

The author is grateful for the continuing guidance of senior supervisor Timothy Racine; and for helpful feedback and unwavering support from colleagues in the Human Evolutionary Studies Program.

References

- Enard, W. et al. (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell*, 137, 961.
- Sudakov, K. V. (1997). The theory of functional systems: General postulates and principles of dynamic organization. *Integrative Physiological and Behavioral Science*, 32(4), 392-414.
- Tokarev, K., Tiunova, A., Scharff, C., Anokhin, K. (2011). Food for song:

Expression of C-Fos and Zenk in the zebra finch song nuclei. *PloS_One*,
6:e21157.

Toomela, A. (2010). Biological roots of foresight and mental time travel.
Integrative Psychological and Behavioral Science, 44(2), 97-125.

CONSTITUENT ORDER IN PICTORIAL REPRESENTATIONS OF EVENTS IS INFLUENCED BY LANGUAGE

ANU VASTENIUS, JORDAN ZLATEV, JOOST VAN DE WEIJER

*Centre for Languages and Linguistics, Lund University,
Sweden*

anu.vastenius@semiotik.lu.se

The origin of word order in human language has been addressed in recent years in empirical research, and in some studies SOV has been found to be the most basic or default order. Goldin-Meadow et al. (2008) conducted a study to test how speakers of languages with different word orders represent events with pictures and gestures. The results showed that the participants predominantly used the order Actor-Patient-Act (ArPA) in their nonverbal representations, irrespective of their native language. Based on this, Goldin-Meadow et al. (2008: 9167) concluded: “there appears to be a natural order that humans, regardless of the language they speak, use when asked to represent events non-verbally”.

Later on, other studies have thrown doubt on the universality of such a “natural order” (e.g. Schouwstra & de Swart, 2014). To investigate this issue, we replicated the experiment by Goldin-Meadow et al. using a slightly modified design. In the replication, no gestures were used, as they are intrinsically more related to language than pictures (Kendon, 2004), and therefore possibly more easily influenced by the native-language word order. Furthermore, contrary to the original study, the pictures were printed on separate, non-transparent cards, which needed to be placed in a particular order so as to produce a representation of the event. In the original study, the pictures were printed on transparencies, which always resulted in the same final product regardless of the order in which they were placed. Consequently, no consistent strategy of ordering was required. In our study, participants performed the task on a transversal plane with a sagittal directionality (from furthest to closest to them). More specifically, the participants had to place the picture cards below one another on a 13 x 52 cm

board, with the narrow side facing them. The intention was that, in this way, they would be minimally influenced by the direction of motion shown in the pictures.

Twenty-six native speakers of Kurdish (SOV) in the Kurdish region of Iraq and twenty-seven speakers of Swedish (SVO) were presented with 36 video clips showing the events. Half of each language group were asked to describe the event prior to ordering the pictures, and the other half only to order the pictures after each video.

The results showed that, unlike in the original study, the constituent order of the native-language did have an impact on the order of the pictorial representations when using this experimental design. The speakers of Swedish were less consistent in using the ArPA order than the speakers of Kurdish, and this tendency was stronger for the participants who described the events verbally before representing them pictorially. This can be interpreted as a moderate version of linguistic relativity, such as Slobin's (1996) thinking-for-speaking, stating that language modulates the cognitive representations that are recruited during the process of language use. It is likely that the explicit linear order in which the pictures had to be placed was more analogous to word order, and hence was more easily influenced by it, than in previous designs.

References

- Goldin-Meadow, Susan, Mylander, Carolyn, So, Wing Chee, and Özyürek, Asli (2008) The natural order of events: How speakers of different languages represent events nonverbally. *PNAS*, 105: 9163-9168.
- Kendon, Adam (2004) *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press
- Schouwstra, Marieke., de Swart, Henriëtte (2014) The semantic origins of word order. *Cognition* 131: 431–436.
- Slobin, Dan I. (1996) From “thought and language” to “thinking for speaking”. In J.J.Gumperz & S.C Levinson (Eds.) *Rethinking linguistic relativity* (p.70-96). Cambridge: Cambridge University Press.

ICONICITY, NATURALNESS AND SYSTEMATICITY IN THE EMERGENCE OF SIGN LANGUAGE STRUCTURE

TESSA VERHOEF¹, CAROL PADDEN¹ & SIMON KIRBY²

*¹Center for Research in Language, UC San Diego,
La Jolla, USA
tverhoef@ucsd.edu, cpadden@ucsd.edu*

*²Language Evolution and Computation Research Unit, University of Edinburgh,
Edinburgh, UK
simon.kirby@ed.ac.uk*

Systematic preferences have been found for the use of different iconic strategies for naming man-made hand-held tools (e.g. Padden et al., 2014; Ortega et al., 2014) in both sign and gesture: HANDLING (showing how you hold it) and INSTRUMENT (showing what it looks like) forms are most frequently used for tools. Within those two, sign languages vary in their use of one strategy over the other (Padden et al., 2013). Nevertheless, despite having overall preferences, what variation exists tends to be conditioned by meaning. In ASL signers, handling forms are more likely to be used for actions and instrument forms for objects (Padden et al., 2014). These lexical preferences across different sign languages provide an ideal test case for understanding the emergence of conventions in language in which multiple types of bias are at play. Specifically, we argue that there may be distinct biases operating during production and interpretation of signs on the one hand, and learning a conventional system of signs on the other. It is crucial we understand how these distinct biases interact if we are to explain the emergence of systematicity in a linguistic system with iconic underpinnings.

We present three experiments that together help to form a picture of the interplay between naturalness, iconicity and systematicity in the origin of linguistic signals. The first experiment (N=720 participants, all non-signers) maps out the initial natural biases people have for pairing ACTION and OBJECT concepts related to tools (e.g. ‘using a toothbrush’ and ‘a toothbrush’) with HANDLING and INSTRUMENT forms in three different tasks, conducted

online. Each participant only responds to one item from one of these tasks, allowing us to rule out any influence of task learning or item order. Task (1): Choosing the right gesture (video) for a given concept. (2): Choosing the right concept for a given gesture. (3): Mapping two concepts and gestures. In line with earlier findings (Padden et al., 2014), we show that non-signers have a strong preference for HANDLING forms in task (1). We also find a strong bias for ACTION concepts in task (2) and a strong bias for mapping HANDLING to ACTION and INSTRUMENT to OBJECT in task (3), demonstrating differences in naturalness of particular iconic strategies. Either mapping would be iconic, but clearly people are biased and find one option more natural.

The second experiment (N=27 non-signers) investigates the effects of these biases on the learnability of artificial languages. In addition to reflecting naturalness on an item by item basis, languages can also vary in systematicity across sets of items (i.e. the extent to which all ACTIONS pattern the same way, and all OBJECTS pattern the same way). Three different languages were designed: (1) congruent with natural bias and systematic, (2) incongruent with bias and systematic, (3) random. As expected, we found languages in category (3) to be harder to learn than those in category (1). Surprisingly, languages in category (2) seem just as learnable as languages in category (1), even though the mapping runs completely counter to the strong naturalness bias we found in experiment 1. A closer look at the performance over time for participants in the different conditions reveals that participants who are exposed to (2) seem to need only a few examples before they detect and accept the unexpected pattern. The results show that even non-signers quickly detect a pattern for which they need to categorize abstract iconic gesture strategies; the handling-instrument distinction cannot be understood by simply relying on differences in form.

The third experiment looks in more detail at the flexibility of participant's biases when they are exposed to data and whether even minimal exposure can nevertheless result in responses that are the reverse of the ones we saw in the first experiment. We exposed non-signers (N=864) to two example tools for which the form-meaning mapping was either (1) congruent with the bias for both, (2) incongruent with the bias for both, (3) one congruent and one incongruent. After this they were asked to respond to one of the three tasks taken from the first experiment for a third tool. Our findings show that, even after exposure to just two examples, the pattern of responses changes strongly, demonstrating that the bias for systematicity operating across sets of items can completely overturn the bias for naturalness operating on individual items.

Our experiments help to understand the subtle interplay between learning biases and mapping biases and how these may shape the emergence of language.

References

- Ortega, G., Sümer, B., & Özyürek, A. (2014). Type of iconicity matters: Bias for action-based signs in sign language acquisition. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society (CogSci 2014)* (pp. 1114-1119). Austin, Tx: Cognitive Science Society.
- Padden, C., Meir, I., Hwang, S.-O., Lopic, R., & Seegers, S., & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture* 13(3): 287-308.
- Padden, C., Hwang, S.-O., Lopic, R., & Seegers, S. (2014). Tools for language: Patterned iconicity in sign language nouns and verbs. *TiCS* 7: 81-94.

LANGUAGE EVOLUTION AND LANGUAGE ORIGINS IN TEACHING LINGUISTICS AT THE UNIVERSITY LEVEL

SŁAWOMIR WACEWICZ^{1,2}, PRZEMYSŁAW ŻYWICZYŃSKI^{1,2}
& ARKADIUSZ JASIŃSKI¹

¹*Center for Language Evolution Studies, ²Department of English,
Nicolaus Copernicus University, Toruń, Poland
cles@umk.pl*

Despite the vigorous development of language evolution research over the last three decades, very little of this progress has trickled into teaching linguistics: so far this important area of the academy has failed to accommodate the bulk of the empirical and theoretical advances. In this paper we report the results of a survey of fourteen popular introductory-level linguistic textbooks, which – with rare exceptions – show that the teaching of language evolution to students of general linguistics rests on out-dated and largely inadequate conceptual frameworks, and fails to communicate major theoretical breakthroughs and empirical results. Based on the feedback from the community of language origins researchers, we formulate an inventory of key messages that should be incorporated into textbooks and curricula.

We have evaluated fourteen introductory academic textbooks on general linguistics published in the last decade. We have established that:

— a majority of textbooks (e.g. Akmaijan et al., 2010) either fail to mention language origins/evolution completely, do it only perfunctorily, or anchor the discussion in inadequate theoretical contexts, such as the classifications by Max Müller or Charles Hockett (see Wacewicz & Żywiczyński, 2015, for why this latter influential framework should be abandoned);

— some of the recently published textbooks, including the most popular ones (e.g. Denham & Lobeck, 2013), show improvement in their coverage of language origins/evolution relative to their previous editions; this, however, mostly applies to their presentation of empirical findings rather than the theoretical backbone.

Overall, despite visible progress, the subject of the evolutionary emergence of language tends to receive inadequate treatment in linguistic textbooks. This

leads to the propagation of such misconceptions as the continuity of language with monkey alarm calls, and the failure to understand the most fundamental prerequisites for the evolutionary language emergence, most notably those related to cooperativeness.

We call for a greater as well as more systematic representation of interdisciplinary language evolution research in basic level linguistic instruction. In particular, the following central messages should be included into teaching materials and curricula:^a

- the newly constituted status of language evolution research, with its inherently interdisciplinary nature, methodological pluralism, and a growing reliance on empirical research (see e.g. Hurford, 2012);
- the cooperative underpinnings of language (e.g. Tomasello, 2008);
- the cognitive and socio-cognitive pre-adaptations (Dor et al., 2014);
- the role of cultural evolution (Smith & Kirby, 2008) and modelling approaches for simulating the emergence of linguistic structure (e.g. Steels, 2011);
- the very nature of language, as seen from the “origins” perspective.

Finally, linguistic textbooks would also benefit from showcasing some of the ways in which tools developed by linguists are applied in other related disciplines; for example, to analysing the compositional structure or turn-taking structure of monkey vocal signalling.

References

- Akmajian, A., Demers, R., Farmer, A., Harnish, R. (2010). *Linguistics: An Introduction to Language and Communication, 6th edition*. Cambridge, MA: MIT Press.
- Dor, D., Knight, C. & Lewis, J., ed. (2014). *The social origins of language: Early society, communication and polymodality*. Oxford: Oxford University Press.
- Hurford, J. R. (2012). Linguistics from an evolutionary point of view. In R. Kempson, T. Fernando and N. Asher (Eds.), *Philosophy of Linguistics* (pp. 473–498). Amsterdam: North Holland/Elsevier.
- Denham, K. & Lobeck, A. (2013). *Linguistics for Everyone. An Introduction*. Boston, MA: Wadsworth Cengage Learning.

^a This list will be supplemented with the results of a survey conducted among the participants of the Protolang language origins conference, 24-26 September 2015, Rome.

- Smith, K. & Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society of London B*, 363(1509), 3591–3603.
- Steels, L. (2011). Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4), 339–356.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Wacewicz, S. & Żywiczyński, P. (2015). Language Evolution: Why Hockett's Design Features are a Non-Starter. *Biosemiotics*, 8(1), 29–46.

LANGUAGES PREFER ROBUST PHONEMES

ANDREW WEDEL
Department of Linguistics
Tucson, AZ, USA
wedel@email.arizona.edu

BODO WINTER
Cognitive and Information Sciences Department, Univ. California
Merced, CA, USA
bwinter@ucmerced.edu

The signaling systems of all human languages are characterized by combinatorial structure: individually meaningless elements are combined to form meaningful units, words. A phoneme's overall contribution to conveying meaning is constrained by the degree to which it is perceptually distinct from other phonemes. If language evolution is influenced by pressures to transmit semantic information efficiently (e.g., Graff, 2012; Piantadosi et al., 2012), spoken languages should evolve to favor more perceptually distinctive phoneme types. Existing theoretical work suggests that those phonemes categories that are more perceptually distinct are transmitted more robustly across linguistic populations (Blevins, 2004; Winter, 2014).

In this paper, we present a test of this prediction: do phoneme inventories evolve to favor more robustly transmitted phonemes? To operationalize robustness, we use Graff (2012)'s symmetric phoneme confusability index, which is based on an English phoneme perception study from Miller and Nicely (1955). For each consonant phoneme within the English phoneme inventory, we ask whether its distinctiveness predicts how likely the phoneme will be found in other languages. And indeed, using the 1,672 languages represented in the PHOIBLE database (Moran et al., 2014), we show that the confusability of a phoneme is inversely correlated with its cross-linguistic frequency.

The evolution of language sound-systems as a whole has been proposed to be driven by changes in individual words, which then spread by diffusion through the lexicon (e.g., Wang, 1969, Wedel 2012). If phoneme inventories

evolve via transmission of phonemes through words, more robust phonemes should be on average found in more words (cf. Graff, 2012). We provide evidence in favor of this prediction by showing that scores on the Graff confusability index also significantly correlate with a set of within-language phoneme lexical frequencies, computed for Turkish, Mutsun, Mawukakan/Mahou, English, Dutch, German and Nama (7 languages from 5 different families): More robust phonemes occur in more words. Finally, the frequency with which consonant phonemes appear across the languages in PHOIBLE is also significantly correlated with within-language lexical frequencies. Specifically, the more languages in PHOIBLE that a phoneme appears in, the more words it tends to appear in within a given language. These results survive controls for language area and language family.

Taken together, our findings support the prediction that phoneme inventories evolve indirectly via a pathway involving transmission of individual words, and that the transmission of phonemes within words is constrained by their perceptual robustness. Our results firmly fit within a view that sees phoneme systems as evolving to meet the functional demands of language use (Wedel et al., 2013), and a view that sees language as more broadly evolved to support robust communication (Winter, 2014).

References

- Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- Graff, P. (2012). Communicative Efficiency in the Lexicon. PhD thesis, Massachusetts Institute of Technology.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122: 280-291.
- Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, 27: 338-352.
- Moran, S., McCloy, D., & Wright, R. (eds.) (2014). *PHOIBLE Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://phoible.org>, Accessed on 2015-08-08.)
- Wedel, A. (2012). Lexical contrast maintenance and the development of sublexical contrast systems. *Language and Cognition*, 4: 319-355.
- Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128: 179-186.
- Winter, B. (2014). Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays*, 36: 960-967.

**RETHINKING ZIPF'S FREQUENCY-MEANING
RELATIONSHIP:
IMPLICATIONS FOR THE EVOLUTION OF WORD MEANING**

BODO WINTER

*Cognitive and Information Sciences, University of California, Merced
Merced, U.S.A.
bodo@bodowinter.com*

DAVID H. ARDELL

*Quantitative Systems Biology, University of California, Merced
Merced, U.S.A.
dardell@ucmerced.edu*

In 1945, George Zipf discovered that more frequent words have more meanings listed in dictionaries than less frequent words. For example, a look at Google Ngram (Michel et al., 2011) reveals that the verb “follow” has a relative frequency of ~0.01%, 100 times more frequent than the verb “adorn” (0.0001%). On macmillandictionary.com, “follow” has 14 senses, “adorn” only 1.

Numerous proposals within cognitive and functional linguistics argue that words acquire new meanings through contextual re-interpretation (e.g., Evans & Wilkins, 2000; Traugott & Dasher, 2001). If a word occurs often enough in a novel context, meaning imbued by the context may become conventionalized. We argue that Zipf's frequency-meaning relationship is in fact reflective of this fundamental mechanism by which semantic systems evolve over time.

Word frequency is positively correlated with contextual diversity (Adelman et al., 2006), a measure of the number of different contexts that a word occurs in. In this paper, we use contextual diversity measures from Google Ngram (the number of different Ngrams a word occurs in) as well as from SUBTLEX-US (the number of different movies a word occurs in, Adelman et al., 2006). We take dictionary data from the MacMillan Online Dictionary as well as from WordNet (Miller, 1995; Fellbaum, 1998).

We perform a series of statistical analyses (negative binomial regressions) of the frequency-meaning relationship, showing that if contextual diversity is

controlled for, word frequency is, in fact, not positively correlated with the number of dictionary senses. Zipf's frequency-meaning relationship is driven by contextual diversity, consistent with cognitive/functional accounts of the evolution of word meaning.

This general result holds for a number of different ways of operationalizing contextual diversity and sense counts. Moreover, previous findings on the distribution of dictionary meanings can be shown to be consistent with our approach: Verbs have more dictionary senses than nouns (Fellbaum, 1990), which our data suggests is due to verbs being more contextually diverse. We also show that the more morphemes a word has (adding semantic specificity), the weaker the relationship between contextual diversity and senses: Words with many morphemes (e.g., "antidisestablishmentarianism") are less prone to acquiring new senses in novel contexts, presumably because they are too specific to be reinterpreted in novel contexts. Finally, we show that contextual diversity from 200 years ago (Google Ngram) predicts present dictionary senses.

Our results suggest a re-interpretation of Zipf's frequency-meaning relationship, and they suggest avenues for novel computational models of evolutionary semantics. We outline implications for models of the evolution of vocabularies (Smith, 2004), as well as for models of Zipfian distributions.

References

- Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17(9), 814-823.
- Evans, N., & Wilkins, D. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, 76, 546-592.
- Fellbaum, C. (1990). English verbs as a semantic net. *International Journal of Lexicography*, 3(4), 278-301.
- Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge: MIT Press.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., & Nowak, M. A., Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331, 176-182.
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38, 39-41.
- Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology*, 228(1), 127-142.
- Traugott, E.C., & Dasher, R. B. (2001). *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of General Psychology*, 33, 251-256.

THE STRUCTURE OF ICONICITY IN THE ENGLISH LEXICON

BODO WINTER

*Cognitive and Information Sciences, University of California, Merced
Merced, U.S.A.
bodo@bodowinter.com*

LYNN K. PERRY

*Department of Psychology, University of Miami
Coral Gables, U.S.A
lkp36@miami.edu*

MARCUS PERLMAN, GARY LUPYAN

*Department of Psychology, University of Wisconsin-Madison
Madison, U.S.A.
mperلمان@gmail.com, lupyan@wisc.edu*

Many theories of language evolution argue for a trajectory from iconic communication systems, where signals resemble their meanings, towards systems with more arbitrary relationships between words and meanings (e.g., Armstrong & Wilcox, 2007; Imai & Kita, 2014). In spoken languages, the end result of this process is often assumed to be a lexicon that is dominated by arbitrariness, with Indo-European languages, especially English, often cited as examples of arbitrariness par excellence (e.g., Vigliocco, Perring, & Vinson, 2014). But, to what extent are the lexicons of present-day spoken languages truly arbitrary? Are there particular kinds of meanings that are more prone to being expressed iconically? Finally, are words either iconic or not, or is iconicity a graded quality of words?

Using the methods presented in Perry et al. (2015), we collected iconicity ratings for 1,952 English words from 705 participants who rated each word on a scale from -5 (“words that sound like the opposite of what they mean”) to +5 (“words that sound like what they mean”). Overall, the distribution of iconicity ratings skewed towards the iconic end of the scale ($+0.9$, $t(1951)=35.5$,

$p < 0.0001$). Hartigan's dip test shows no evidence for bimodality in the iconicity ratings ($p > 0.1$), suggesting that iconicity is indeed a continuous rather than categorical notion. In line with Perry et al. (2015), iconicity varied between lexical categories ($F(4,1881) = 35.41$, $p < 0.0001$), with verbs and adjectives receiving higher iconicity ratings. These results mirror the patterning of ideophones across the meaning space of languages (Dingemanse, 2012), i.e., they preferentially express manner of movement and sensory perceptions.

We further tested whether meanings related to specific sensory modalities are more or less prone to iconicity by using perceptual strength ratings from Lynott and Connell (2009), who asked participants to judge how much an object property such as "yellow" or "loud" is perceived through each of the five senses (seeing, hearing, touch, taste, smell). Overall perceptual strength was positively associated with iconicity ($F(1,415) = 6.88$, $p < 0.01$). Predominantly auditory words ("rustling", "buzzing", "muffled") received the highest iconicity ratings (+2.3), followed by haptic (+1.8, "sticky", "soft"), visual (+1.21, "shiny", "yellow"), olfactory (+1.04, "fishy", "aromatic") and gustatory words (+0.8, "acidic", "tasty").

These results reveal that the English lexicon harbors a considerable amount of iconicity in its sound structure, something that native speakers can pick up on. Moreover, words rated as the highest in iconicity correspond to meanings that are commonly encoded as ideophones in the world's languages. Vocal iconicity is particularly concentrated in sensory meanings, especially those relating to the auditory, haptic and visual senses, but less to the chemical senses. This suggests that in early communication systems, vocal iconicity may have been more useful for expressing some meaning categories compared to others.

References

- Armstrong, D. F., & Wilcox, S. (2007). *The gestural origin of language*. Oxford: Oxford University Press.
- Dingemanse, M. (2012). Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass*, 6, 654-672.
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130298.
- Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41, 558-564.
- Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and Its Relation to Lexical Category and Age of Acquisition. *PLoS ONE*, 10(9), e0137147.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 707-784.

SIGNAL AUTONOMY IS SHAPED BY CONTEXTUAL PREDICTABILITY

JAMES WINTERS, SIMON KIRBY, KENNY SMITH
Centre for Language Evolution, University of Edinburgh

At the heart of any communication system is the goal of reducing uncertainty about the intended meaning of the speaker. In achieving this aim, speakers not only rely on the conventional meaning of linguistic forms, but also on how these forms interact with the contextual information at hand. In short, when context is known and informative, it helps in reducing uncertainty about the intended meaning (Piantadosi, Tily & Gibson, 2012). This relationship between context, meaning and uncertainty has important consequences for how cultural evolutionary processes shape the structure of linguistic systems. A recurrent observation is that languages vary in their *signal autonomy*: the degree to which a signal can be interpreted in isolation, without recourse to contextual information (Wray & Grace, 2007). One hypothesis is that signal autonomy is causally related to *contextual predictability*: to what extent can a speaker estimate and therefore exploit the contextual information that a hearer is likely to use in interpreting an utterance.

To investigate these claims, we experimentally simulate the relative pressures from speakers and hearers in a communication game, with the main manipulation being to the *referential context*: the relationship between a target object and a set of distractor objects, and how these impact upon the task of discrimination (Winters, Kirby & Smith, 2015). For the training phase, participants were trained on an artificial language, which consisted of randomly generated sets of 2-3 syllable signals. These signals were then assigned to four images that differed from each other on the dimensions of shape and colour (e.g., *blue blob*, *grey oval*, *red square*, *yellow star*). The trained language was therefore ambiguous with respect to whether the signals referred to colour, shape, or both colour and shape. Participants were then assigned fixed roles of either a *speaker* or a *hearer* for the communication phase. In each trial, speakers typed a signal for a target image, and hearers used this signal to discriminate the target from a set of three distractors (the context). There were a total of 16 target images a speaker needed to convey over three blocks of 32 trials.

To test for the effect of referential context on signal autonomy we manipulated two variables: (i) context-type (across trial predictability) and (ii) access to context (within trial predictability). Context-type is the extent to which a partic-

ular dimension (e.g., shape) is relevant for discrimination across successive trials. For the *Shape Different* referential contexts, the context-type remains consistent across trials, as targets and distractors differ in shape, but share the same colour. *Mixed* context-types vary across trials: half of the trials consist of contexts in which the target and distractors differ in shape (but share the same colour) and half in which they differ in colour (but share the same shape). We also manipulated whether the speaker had knowledge about the relevant distinctions needed to communicate with the hearer. In the *Shared* conditions, speakers had access to the context (i.e., the target and distractors that hearer was confronted with), whereas in the *Unshared* condition speakers only saw the target in isolation (although the hearer's task remained the same: to distinguish a target from a set of three distractors). This gives us four conditions: *Shape-Different Shared*, *Shape-Different Unshared*, *Mixed Shared*, *Mixed Unshared*. By decreasing contextual predictability within and across trials we predict that speakers will respond by creating more autonomous signals (and vice versa).

Our results show that context does shape the degree of signal autonomy: when the context is predictable, languages are organised to be less autonomous (more context-dependent) through combining linguistic signals with context to reduce uncertainty. When the context decreases in predictability, speakers favour strategies that promote autonomous signals, allowing linguistic systems to reduce their context dependency. For the *Shape-Different Shared* condition, which was maximally predictable in terms of context-type and access to context, speakers only conveyed the shape dimension in their linguistic systems, leaving out the colour dimension as this was irrelevant to communicative success (resulting in low autonomy). Conversely, in the *Mixed Unshared* condition, which had the lowest contextual predictability, speakers consistently opted for strategies that promoted compositional structure: this allowed for autonomous systems that specified both colour and shape within the linguistic system. For the conditions in-between these two extremes of contextual predictability – *Shape-Different Unshared* and *Mixed Shared* – speakers were more heterogeneous in their strategy choice, with the resulting systems varying in their degree of autonomy. Taken together, these results show how pragmatic factors can play a salient role in the cultural evolution of language, with manipulations to contextual predictability shaping the types of systems that emerge over repeated interactions between speakers and hearers.

Bibliography

- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3): 280-291.
- Wray, A., & Grace, G.W. (2007). Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117: 543-578.
- Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, 7(3): 415-449.

THE CULTURAL CO-EVOLUTION OF LANGUAGE AND MINDREADING

MARIEKE WOENSDREGT

KENNY SMITH, CHRIS CUMMINS, SIMON KIRBY

*Centre for Language Evolution, School of Philosophy, Psychology & Language Sciences
University of Edinburgh, Edinburgh, United Kingdom
m.s.woensdregt@sms.ed.ac.uk*

One defining feature of human language is its heavy reliance on reasoning about the mental states of others (e.g. Scott-Phillips, 2014). The development of such mindreading skills was a crucial step in hominin evolution because it allowed for the expression and recognition of communicative intentions, thereby paving the way for the cooperative information sharing that we find in humans today. The ability to recognize and infer communicative intentions also plays an important role in language development on an ontogenetic time-scale, as evidenced by studies correlating mindreading skills and word learning (e.g. Parish-Morris, Hennon, Hirsh-Pasek, Golinkoff, & Tager-Flusberg, 2007).

This relationship between language and mindreading may be reciprocal; the acquisition of language has been shown to unlock further levels of mindreading development in the individual (e.g. Lohmann & Tomasello, 2003; Pyers & Senghas, 2009). Furthermore, Heyes and Frith (2014) have argued that mindreading is a skill that has developed (partly) through cumulative cultural evolution - just like language (Kirby, Tamariz, Cornish, & Smith, 2015).

In this paper we present a computational model that investigates the implications of such a bidirectional interaction between the capacities of language and mindreading. This new agent-based model builds on previous models of cross-situational word learning (Siskind, 1996), but extends this framework by allowing agents to learn a way of inferring an interlocutor's communicative intentions, simultaneously with learning the lexicon.

In the model, agents communicate (probabilistically) about objects that are proximate to them. The learner's task is to establish the relations between the words and the objects (i.e. the lexicon), and the perspective of the speaker, which may be the same as the learner's own or may differ. The communicative intentions of the speaker are a product of the context that the agents find themselves in and the speaker's perspective (which renders some objects more salient than others). Utterances in turn are a product of a speaker's communicative intention

and lexicon.

The data that a learner has access to in order to learn a speaker's lexicon and perspective consists of the speaker's word use in context. From these two observable variables the learner has to infer two unobservable variables simultaneously: the speaker's lexicon and their perspective. This learning happens through Bayesian inference, where accumulating evidence allows the learner to weigh different combinations of lexicon and perspective hypotheses against each other - based on the likelihood of the incoming data given a specific hypothesis.

Although a speaker's utterances are as much a product of their perspective as they are of their lexicon, simulations with this model show that given consistent input, a learner is able to infer both the correct lexicon and the correct perspective from scratch, by reciprocally bootstrapping the learning of the one variable with partial knowledge of the other. The learning trajectory that is revealed in these simulations is one where acquiring a bit of the lexicon helps the learner infer the speaker's perspective, which in turn allows the learner to acquire the rest of the lexicon.

We will discuss the dynamics of this model on two different levels: exploring the emergence of language and mindreading capacities both within individual agents and across generations of a population. This model thus gives insight into the effects of an individual-level interaction of cognitive capacities on population-wide dynamics such as establishing and maintaining a stable signalling system; thereby connecting proximate and ultimate causes of language evolution.

References

- Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science (New York, N.Y.)*, *344*(6190), 1243091.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.
- Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*(4), 1130–1144.
- Parish-Morris, J., Hennon, E. a., Hirsh-Pasek, K., Golinkoff, R. M., & Tager-Flusberg, H. (2007). Children with autism illuminate the role of social intention in word learning. *Child development*, *78*(4), 1265–87.
- Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from learners of a new sign language. *Psychological science*, *20*(7), 805–812.
- Scott-Phillips, T. (2014). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Palgrave MacMillan.
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2), 39–91.

A GAME THEORETIC ACCOUNT OF SEMANTIC SUBJECTIFICATION IN THE CULTURAL EVOLUTION OF LANGUAGES

EVA ZEHENTNER, ANDREAS BAUMANN, NIKOLAUS RITT
& CHRISTINA PRÖMER

*Department of English Studies, University of Vienna,
Vienna, Austria
eva.zehentner@univie.ac.at*

Linguistic research from the past decades has revealed a pathway in semantic change by which cultural transmission causes word meanings to become increasingly subjective, i.e. increasingly based in speakers' beliefs and attitudes (see Davidse, Vandelanotte & Cuyckens 2010 and the references therein).

A prototypical – but far from the only – case of subjectification is the rise of 'epistemic' meanings (1b) in 'deontic' modals (1a) (Traugott 1989).

- (1) a. *John must work hard to survive.* (objective necessity)
b. *John looks tired. He must be working hard.* (speaker's subjective certainty)

While subjectification is often taken to reflect the need of speakers to express their inner selves, we consider this hypothesis as shallow and little informative. Instead, we propose an account in terms of evolutionary game theory and take subjectification to emerge through sender-receiver interactions where senders may attempt to manipulate receivers (e.g. by altering their construal of reality), while receivers may exploit signals for reading speakers' minds (i.e. beliefs, goals and intentions) (cf. Dawkins & Krebs 1984).

In our model, interlocutors may intend or interpret a message as either objective (about external reality) or subjective (about beliefs etc.). They may be cooperative or uncooperative (at a proportion that we fix *a priori* at $q \in (0,1)$). Cooperative speakers are honest, uncooperative ones lie. Cooperative listeners are credulous, uncooperative ones disregard the encoded message, but try to infer hidden speaker beliefs.

The evolutionary dynamics of the populations of subjective and objective interlocutors are modeled as an asymmetric role game (Hofbauer & Sigmund 1998: 122ff.) with two positions (speaker and listener) and two strategies (subjective and objective), yielding four different behavior types (subjective speaking & subjective listening; objective speaking & subjective listening, etc.). This yields a 4-by-4 game with 16 different encounter types.

Payoffs resulting from pairwise speaker-hearer interactions are divided into four ordinal categories (no benefit/loss, small benefit/loss, medium benefit/loss, and large benefit/loss), which are numerized from 0 to ± 3 . Information about external reality is taken to be more valuable when true (and more harmful when false) than information about speakers' intentional states.

For each combination of cooperative or uncooperative individuals choosing one of the available strategies in one of the two positions the payoff is determined heuristically and weighted according to the assumed proportions of cooperative and defective players.

An analysis of the resulting dynamics reveals two qualitatively different evolutionary outcomes: if the proportion of cooperative players does not exceed a certain threshold ($q < 1/\sqrt{2}$), the behaviour type 'objective speaking & subjective listening' represents the only evolutionarily stable strategy-combination. Otherwise, i.e. if the proportion of cooperative speakers is extraordinarily large, the replicator dynamics exhibit a cyclic behavior where speakers switch periodically from one strategy to the other, followed by subsequent periodic listener-strategy adaptations.

We take this to suggest that subjectification is driven by listener's interest in (potentially hidden) beliefs and intentions of speakers rather than by speakers' desire to express their inner selves. At the same time, our account shows that concepts developed in the study of animal communication can be productively applied in the study of language diachrony as well.

References

- Davidse, K., L. Vandelanotte & H. Cuyckens (eds.) (2010). *Subjectification, intersubjectification and grammaticalization*. Berlin: De Gruyter Mouton.
- Hofbauer, J.; Sigmund, K. (1998): *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.
- Krebs, J. R. & R. Dawkins. (1984). Animal signals: mind reading and manipulation. In J. R. Krebs & N. B. Davies (eds.), *Behavioural ecology: An evolutionary approach*, 380–402. Sunderland, Mass.: Sinauer Associates.
- Traugott, E. C. (1989). On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65(1). 31–55.

DEEP LEARNING MODELS OF LANGUAGE PROCESSING AND THE EVOLUTION OF SYNTAX

WILLEM ZUIDEMA

*Institute for Logic, Language and Computation
Amsterdam Brain & Cognition
University of Amsterdam
Amsterdam, the Netherlands
zuidema@uva.nl*

Jackendoff's (2002) observation that 'Your theory of language evolution depends on your theory of language' is nowhere more true than in accounts of the origins of syntax and grammar. Researchers arguing that no specialised cognitive machinery is needed for syntactic processing, naturally locate the explanation for the emergence of syntactic language somewhere in the hominin lineage in non-linguistic domains such as changes in social structure, brain capacity or cooperativity. Researchers that favor connectionist or constructionist models of syntax and grammar with language-specific parameters & mechanisms, naturally argue for gradual, evolutionary adaptation of a preexisting brain structures. And finally, researchers that assume discrete, symbolic models of syntax, tend to reject gradualist accounts in favor of a 'saltationist' view, where the crucial machinery for learning and processing syntactic structure emerges in a single step (as a side-effect of known or unknown brain principles or as a macro-mutation). A key argument for the last position is that the vast amount of work on describing the syntax of different natural languages makes use of symbolic formalisms like minimalist grammar, HPSG, categorial grammar etc., and that it is difficult to imagine a process of any sort that moves gradually from a state without to a state with their key components. The most often repeated version of this argument concerns 'recursion' or 'merge', which is said to be an all or nothing phenomenon where you can't have just a little bit of it (Reuland, 2009).

In my talk I will review this argument in the light of recent developments in computational linguistics where a class of models has become popular that we can call 'vector grammars', that include the Long-Short Term Memory networks, Gated Recurrent Networks and Recursive Neural Networks (e.g., Goller & Küchler, 1996; Hochreiter & Schmidhuber, 1997; Socher, Manning, & Ng, 2010; Socher, Bauer, Manning, & Ng, 2013; Karpathy, Johnson, & Fei-Fei, 2015; Rohrmeier, Zuidema, Wiggins, & Scharff, 2015). These models have over the

last two years swept to prominence in many areas of NLP and shown state of the art results on many benchmark tests including constituency parsing, dependency parsing, sentiment analysis and question classification. They have in common that they are much more expressive than the older generation of connectionist models (roughly speaking, vector grammars are just as expressive as grammar formalisms popular in theoretical linguistics), but replace the discrete categorical states from those formalisms with points in a continuously valued vector space and hence quite easily allow accounts of gradual emergence (as exploited when these models are trained on data using variants of backpropagation).

I will conclude that these new type of models, although mostly evaluated in a language engineering context, hold great promise for cognitive science and language evolution. For theories of the evolution of grammar and syntax they effectively undermine a key argument: that the only explanatory models are incompatible with gradualist evolution; rather, they show that syntactic category membership, recursion, agreement, long-distance dependencies, hierarchical structure etc. can all gradually emerge, both in learning and in evolution.

References

- Goller, C., & Küchler, A. (1996). Learning task-dependent distributed representations by backpropagation through structure. In *International conference on neural networks* (pp. 347–352).
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- Jackendoff, R. (2002). *Foundations of language*. Oxford, UK: Oxford University Press.
- Karpathy, A., Johnson, J., & Fei-Fei, L. (2015). *Visualizing and understanding recurrent networks* (Tech. Rep.). arXiv preprint arXiv:1506.02078.
- Reuland, E. (2009). Imagination and recursion: issues in the emergence of language. In F. C. Sophie de Beaune & T. Wynn (Eds.), *Cognitive archaeology and human evolution*. Cambridge University Press.
- Rohrmeier, M., Zuidema, W., Wiggins, G. A., & Scharff, C. (2015). Principles of structure building in music, language and animal song. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1664), 20140097.
- Socher, R., Bauer, J., Manning, C. D., & Ng, A. Y. (2013). Parsing with compositional vector grammars. In *Proceedings of the ACL conference*.
- Socher, R., Manning, C. D., & Ng, A. Y. (2010). Learning continuous phrase representations and syntactic parsing with recursive neural networks. In *Proceedings of the NIPS-2010 Deep learning and unsupervised feature learning workshop*.

LANGUAGE-BIOLOGY COEVOLUTION FIXATION TIMES

BART DE BOER

*AI-lab, Vrije Universiteit Brussel,
Brussels, Belgium
bart@ai.vub.ac.be*

In order to understand language evolution, we need to understand the interaction between biological and cultural evolution (e. g. Deacon, 1997). This paper presents a modification of a standard approach from theoretical biology (the Moran process, Moran, 1958, explained below) for calculating how quickly biological specializations to culturally transmitted traits can evolve: the fixation time. This addresses two issues that thwart the analysis of language-biology coevolution. The first issue is that of the speed of biological and cultural evolution. Although it is often assumed that languages change much faster than biological evolution, (cultural) language change may be very slow in some cases (Pagel, Atkinson, & Meade, 2007, although see Greenhill, Atkinson, Meade and Gray 2010 for a different perspective on the speed of language change) while biological evolution can operate rapidly in small populations. The second issue is that evolution operates in finite populations, so randomness plays a role. This means that one cannot just look for fitness advantages, but must calculate the probabilities of the spread of a trait.

The Moran process is used to mathematically model evolution in finite populations. It describes the change in the number of mutants in an otherwise uniform population of fixed size. The state of a population can then be fully described by a single number. The Moran process only allows for an increase or decrease by one of the number of mutants. Thus evolution can be modeled as a simple Markov process, consisting of a single sequence of states, allowing the calculation of the probability that a mutant will spread and the time it will take to spread.

By allowing multiple (instead of one) sequences of states – each sequence modeling a different type of culture – and new transitions between states with different culture, this paper extends the Moran process to cases where biology coevolves with culture. This is applied to an example (modeled after Chater, Reali, & Christiansen, 2009) in which there are two cultural states (neither with an intrinsic fitness advantage) and two biological types of agents: generalists and specialists. Generalists learn both types of culture equally well, whereas specialists learn one type better than the generalists (giving a fitness advantage) and

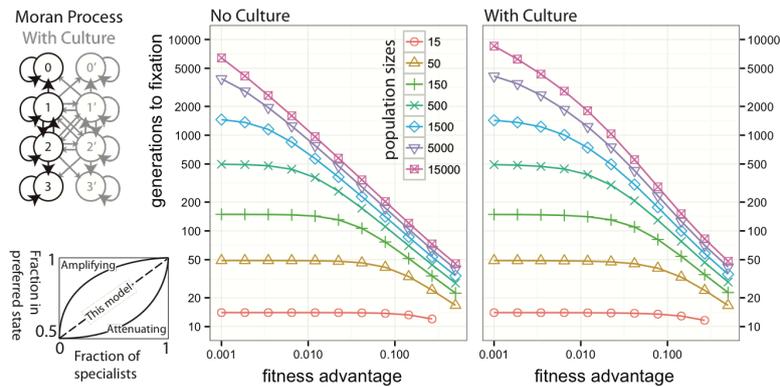


Figure 1: Top left Panel: Moran process with extra states and transitions needed for incorporating culture indicated in grey. Numbers indicate the number of mutants in a population state, arrows indicate possible transitions. Bottom left panel: graphs illustrating amplifying and attenuating culture. In amplifying culture, a small proportion of specialists causes the culture to spend a large fraction of time in the preferred state. Middle and right panels: fixation times in generations for systems without culture (middle) and systems with culture (right). The x-axis shows the effect of the fitness advantage associated with being a specialist in the preferred culture. Different line markers indicate different population sizes. Culture is neither amplifying nor attenuating. Note that fixation times in the presence of culture are somewhat longer, but not extremely so.

one type worse. At the same time, the number of specialists determines the fraction of time the culture spends in the state preferred by the specialists. This is illustrated in the lower left panel of figure 1. If the fraction of specialists is 0, one can expect the population to spend half of the time in each cultural state (when both states are truly equivalent). If the fraction of specialists increases, one expects the fraction the population spends in the preferred state to increase as well. The relation between the two fractions can have different forms, but here only the linear relation is explored.

Using a mathematical technique developed by Antal en Scheuring (2006) the time it takes for a mutation for specialization to a specific cultural variant to reach fixation was calculated. It turns out that for a range of fitness advantages and a range of population sizes, fixation times in the presence of culture are not much longer than when there is no culture. This indicates that it is possible for biological adaptations for arbitrary culture to evolve whenever the number of possible cultural states is low, and when the number of specialists for a certain type of culture determines the fraction of time a population spends in that type of culture. Also, it does not necessarily mean that such adaptations will be strong (i.e. individuals can only learn their preferred type of language) or easily observable, as work by Thompson, Smith & Kirby (2012) shows that whenever a weak bias evolves, the amplifying effect of culture causes any advantage of stronger biases to disappear. However, my results do show that we should not a priori give up hope of finding cognitive mechanisms that evolved for language.

References

- Antal, T., & Scheuring, I. (2006). Fixation of Strategies for an Evolutionary Game in Finite Populations. *Bulletin of Mathematical Biology*, 68(8), 1923–1944.
- Chater, N., Reali, F., & Christiansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences*, 106(4), 1015–1020.
- Deacon, T. W. (1997). *Symbolic species, the co-evolution of language and the human brain*. The Penguin Press.
- Greenhill, S. J., Atkinson, Q. D., Meade, A., & Gray, R. D. (2010). The shape and tempo of language evolution. *Proceedings of the Royal Society of London Series B - Biological Sciences*, 277, 2443–2450.
- Moran, P. A. P. (1958). Random Processes in Genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(1), 60–71.
- Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449, 717–721.
- Thompson, B., Smith, K., & Kirby, S. (2012). Cultural Evolution Renders Linguistic Nativism Implausible. In T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, & J. R. Hurford (Eds.), *The Evolution of Language: Proceedings of the 8th international conference (EVO LANG9)* (pp. 557–558). Hackensack (NJ): World Scientific.

CATEGORY LEARNING IN AUDITION, TOUCH, AND VISION

SABINE VAN DER HAM, BILL THOMPSON, BART DE BOER

*Artificial Intelligence Lab, Vrije Universiteit Brussel
Brussels, Belgium
sabine@ai.vub.ac.be*

The evolution of linguistic cognition is a notoriously difficult problem: as all learning mechanisms are intertwined with human development and behavior, it is hard to tease apart which aspects of language are the result of cultural processes and which are evolved cognitive traits (De Boer, 2015). We explore whether humans have specializations related to language, and specifically, whether statistical learning of categories is specialized for language-like signals. While it is well-known that statistical learning is not restricted to language learning, the way the mechanism operates in different domains may not be the same, and may be affected by perceptual and cognitive constraints (Conway & Christiansen, 2005).

We present an experiment in which participants learn, categorize, and reproduce signals in the acoustic, visual, and tactile modalities. Participants are trained on bimodal distributions of tones, images and buzzes with some variation in duration, resulting in a ‘long’ and ‘short’ category (Maye, Werker, & Gerken, 2002). After training, participants rate individual signals on a 6 point scale from ‘definitely short’ to ‘definitely long’. The production task consists of creating 3 signals for both categories by pressing the mouse button. The signal is presented as long as the button is pressed. If there is indeed a specialization for learning language-like signals, then we expect that participants reveal: a) more certainty in the categorization task in the auditory and visual conditions, and b) better estimation of the peaks in the distributions, resulting in more reliable reproductions of the categories in the auditory and visual conditions. Alternatively, in normal hearing adults there may be a linguistic training effect, for instance from distinctions between long and short vowels in their native language, in which case better performance in the auditory condition compared to the other conditions is expected, following earlier perception studies (Jones, Poliakoff, & Wells, 2009).

A within subjects design provided us with categorization and production data from 29 participants in all modalities. Results from a logistic regression reveal three interesting trends. First, statistical learning of categories and (reliably) reproducing them is possible in all domains, including the tactile modality. Second, and most importantly, the categorization and production processes are remark-

ably similar in the tactile and auditory domains (odds ratios 0.99 and 1.10; 95% confidence intervals 0.84-1.15 and 0.94-1.28, respectively), but not in the visual domain (OR = 1.30; CIs = 1.12-1.53), suggesting that there is no cognitive specialization for learning language-like signals, nor that there is an effect from previous language experience. Finally, comparing the categorization and production data reveals an interesting tension: on the one hand, we demonstrate that across all three modalities participants were able to induce representations of distinct categories that respect the statistics of the training distributions. On the other hand, we consistently find more variation in duration among participants' productions than was present in the training distributions. This pattern of variation differs between modalities, with more variation in the visual modality. We fitted a Bayesian model of inference for Gaussian distributions to the data in order to investigate whether this variation reflects a meaningful component of the learning process, and to provide a quantitative characterization of the pattern of differences in how participants formed categories in this experiment. The model correctly predicts the overall pattern of categorization behaviour that we find in the empirical data.

Our study further explores the extent to which statistical learning is domain general, as it is becoming increasingly clear that we cannot approach it as a unitary mechanism (Frost, Armstrong, Siegelman, & Christiansen, 2015). Teasing apart differences between modalities will help us find out how the perceptual biases of each modality potentially affect these learning mechanisms, which is necessary for understanding whether certain forms of learning are specialized for language.

Acknowledgements

This research was supported by European Research Council Starting Grant number 283435, awarded to B. de Boer.

References

- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 24.
- De Boer, B. (2015). Biology, culture, evolution and the cognitive nature of sound systems. *Journal of Phonetics*, 53, 79–87.
- Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in cognitive sciences*, 19(3), 117–125.
- Jones, L. A., Poliakoff, E., & Wells, J. (2009). Good vibrations: Human interval timing in the vibrotactile modality. *The Quarterly Journal of Experimental Psychology*, 62(11), 2171–2186.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111.

Index

Żywicznyński Przemysław, 594

Abramova Ekaterina, 331

Ackermann Hermann, 20

Aguilar Elliot, 451

Alday Phillip M., 22

Ames Audra, 334

Anderson Noelle, 337

Anton Jean-Luc, 483

Archbold Jahmaira, 382

Ardell David, 337, 599

Ardila Alfredo, 339

Asano Rie, 342, 345

Atkinson Mark, 350

Barcelo-Coblign Lluís, 31

Baumann Andreas, 40, 609

Bayne Ted, 49

Behme Christina, 352

Benítez-Burraco Antonio, 365

Bentz Christian, 59

Berdicevskis Aleksandrs, 70

Bergmann Till, 79

Bertello Alice, 483

Bickel Balthasar, 388

Blythe Richard A., 362

Boeckx Cedric, 365

Bonnard Mireille, 490

Bowling Daniel Liu, 419, 422

Braun Mario, 490

Brentari Diane, 523

Briesemeister Benny, 490

Brink Tila, 490

Broadway Megan, 368, 472

Bryant Greg, 371

Buch Armin, 190

Burtenshaw Tiarn, 466

Byrne Richard, 385, 435

Cameron Dianne, 334

Carr Jon W., 374, 558

Cartmill Erica, 376

Cavicchio Federica, 379

Ceolin Andrea, 190

Cheng Leveda, 541

Christodoulopoulos Christos, 539

Chu Becky, 412

Clay Zanna, 382

Cochet H el ene, 385

Collier Katie, 388

Congdon Jenna V., 419

Coppola Marie, 132, 523

Cornish Hannah, 374

Coulon Olivier, 483

Culbertson Jennifer, 391, 454

Cummins Chris, 607

Cuskley Christine, 394, 397

Cuthbertson Cory, 399

Dale Rick, 79

de Boer Bart, 182, 407, 419, 422,

468, 613, 616

de Swart Henri ette, 563

Dediu Dan, 401, 445, 448

Delgado Tania, 520

Leemor Livnat, 379
 Lenzen Daniel, 502
 Lepic Ryan, 502
 Lestrade Sander, 171
 Levelt Clara C., 433
 Levy Roger, 227
 Lewis Molly, 464
 Lister Casey, 466
 Little Hannah, 182, 407, 468
 Longobardi Giuseppe, 190
 Loreto Vittorio, 394, 397, 570
 Love Scott A., 483
 Lupyan Gary, 403, 470, 511, 556,
 602
 Lyn Heidi, 368, 472

 Mühlenbernd Roland, 97, 236
 Müller Gudrun, 427
 Macuch Silva Vinicius, 475
 Manser Marta B., 388
 Margiotoudi Konstantina, 483
 Marie Damien, 483
 Massaro Dominic, 511
 Mccrohon Luke, 478
 Mcloughlin Michael, 480
 Meguerditchian Adrien, 483
 Michaud Jérôme, 199
 Michelioudakis Dimitris, 190
 Micklos Ashley, 209
 Mills Gregory, 485
 Miranda Eduardo, 480
 Mitchell Dominic, 487
 Mitchener William, 217
 Moisik Scott, 401, 445, 448
 Monechi Bernardo, 397
 Montant Marie, 490
 Morgan Emily, 227
 Morgan Hope, 491
 Motamedi Yasamin, 493

 Nölle Jonas, 437
 Naccache Albert, 245, 496

 Nakai Tomoya, 499
 Namboodiripad Savithry, 502
 Nazarian Bruno, 483
 Newen Albert, 419, 422
 Niederhut Dillon, 254
 Nielsen Alan, 505
 Noad Michael, 480
 Noble Bill, 262

 Ocklenburg Sebastian, 419, 422
 Ohan Jeneva, 466
 Okanoya Kazuo, 499, 507
 Oudeyer Pierre-Yves, 288

 Padden Carol, 591
 Paukonis Andrius, 419
 Pepperberg Irene M., 509
 Perlman Marcus, 403, 511, 602
 Perry Lynn, 511, 602
 Pijpops Dirk, 580
 Pika Simone, 427, 514
 Pleyer Michael, 271, 437, 517
 Pleyer Monika, 271
 Ponz Aurélie, 490
 Power Camilla, 163
 Prömer Christina, 40, 609
 Progovac Ljiljana, 280

 Radford Andrew N., 388
 Radkevich Nina, 190
 Rasilo Heikki, 312
 Ravnani Andrea, 520, 543
 Reber Stephan, 419
 Rendell Luke, 480
 Renton Jessica, 362
 Richie Russell, 132
 Ridley Amanda R., 405
 Rissman Lilia, 523
 Ritt Nikolaus, 40, 415, 609
 Roberts Gareth, 526
 Roberts Sean, 140, 448, 475
 Rogers Shane, 410
 Rosselló Joana, 440, 528

Roth Muriel, 483
 Russell Andrew, 568

 Saldana Carmen, 530
 Salikoko S. Mufwene, 570
 Samuelson Mystera, 472
 Sandler Wendy, 379, 554
 Santacreu-Vasut Estefania, 124
 Santana Carlos, 533
 Schel Anne, 541
 Schouwstra Marieke, 391, 493, 536,
 563
 Schueller William, 288
 Seinhorst Klaas, 297
 Senghas Annie, 523
 Serafin Billie, 368
 Servedio Vito, 570
 Shamay-Tsoory Simone, 379
 Shi Edward Ruoyang, 345
 Shurley Joleana, 560
 Silvey Catriona, 539
 Slocombe Katie, 541
 Smith Kenny, 350, 374, 415, 454,
 457, 493, 505, 530, 536,
 548, 550, 552, 605, 607
 Smith Scott, 509
 Sonnweber Ruth, 543
 Spierings Michelle, 546
 Spierings Michelle J., 433
 Spike Matt, 417
 Spike Matthew, 548, 550
 Stadler Kevin, 552
 Stamp Rose, 554
 Stangl Katelyn, 425
 Sturdy Christopher B., 419
 Sulik Justin, 470, 556

 Tallerman Maggie, 304
 Tamariz Monica, 558, 560
 ten Cate Carel, 433, 459, 546
 Theofanopoulou Constantina, 365
 Thompson Bill, 312, 563, 616

 Tinitis Peeter, 437
 Tkachman Oksana, 566
 Townsend Simon, 541, 568
 Townsend Simon W., 388, 405
 Tria Francesca, 570
 Trueswell John, 412
 Truswell Robert, 574, 576

 Ullrich Robert, 578

 Van de Velde Freek, 580
 Van de Weijer Joost, 589
 van der Ham Sabine, 616
 Vasileva Olga, 583, 586
 Vastenius Anu, 589
 Verhoef Tessa, 502, 591

 Wacewicz Sławomir, 594
 Wahle Johannes, 236
 Walker Bradley, 466
 Wallace Emma, 541
 Watson Stuart, 541
 Wedel Andrew, 597
 West Victoria, 541
 Wichmann Soeren, 149
 Wicker Bruno, 490
 Wielandt Sara, 334
 Wilke Claudia, 541
 Winter Bodo, 337, 448, 511, 597,
 599, 602
 Winters James, 605
 Wittig Roman M, 427
 Woensdrecht Marieke, 607
 Wonnacott Elizabeth, 415

 Yanovich Igor, 97, 321

 Zehentner Eva, 609
 Ziegler Johannes, 490
 Ziegler Wolfram, 20
 Zilber-Izhar Katia, 509
 Zlatev Jordan, 589
 Zuberbuhler Klaus, 382
 Zuidema Willem, 611