THEORIA
ET HISTORIA SCIENTIARUM
VOLUME XI

MULTIPLE PERSPECTIVES
ON LANGUAGE ORIGINS

Issue editors:
Sławomir Wacewicz, Przemysław Żywickiński

Guest editors:
Piotr P. Chruszczewski, Katarzyna Buchek,
Aleksandra R. Knapik, Jacek Mianowski

Toruń 2014
Editors
Sławomir Wacewicz, Przemysław Żywiczyński, Tomasz Komendziński

Editorial Board
James Hurford, University of Edinburgh, UK
Włodzisław Duch, Nicolaus Copernicus University, Poland

Advisory Editorial Board
David M. Armstrong, University of Sydney, Australia
Mario Bunge, McGill University, Canada
Piotr Chruszczeński, University of Wrocław, Poland
Nathalie Gontier, University of Lisbon, Portugal
Stevan Harnad, Princeton University, USA
Michał Heller, Pontifical Academy of Theology, Poland
Michael Hoskin, Cambridge University, England
Ray Paton, University of Liverpool, England
Steven Pinker, Massachusetts Institute of Technology, USA
Vilayanur S. Ramachandran, University of California, San Diego, USA
Nicholas Rescher, University of Pittsburgh, USA
John Searle, University of California, USA
Piotr Stalmaszczyk, University of Łódź, Poland
Jordan Zlatev, Lund University, Sweden

PL ISSN 0867-4159

© Copyright by Uniwersytet Mikołaja Kopernika
Toruń 2014

The editors would like to express gratitude to the following individuals for their help in the editorial process: Katarzyna Rogalska, Natalia Palka, Monika Boruta, Krzysztof Kaszyński, Marek Placiński, Karolina Kniewska, Marta Sibierska, Julia Trzeciakowska, Monika Jasińska, Anna Lewandowska, Izabela Sznajder, Kamila Szybakowska.

A publication of Center for Language Evolution Studies (CLES), NCU Toruń
http://www.cles.umk.pl/

Editorial address: Theoria et Historia Scientiarum
Center for Language Evolution Studies (CLES)
Nicolaus Copernicus University
ul. Bojański 1
87-100 Toruń, POLAND

Publisher: Uniwersytet Mikołaja Kopernika
ul. Gagarina 11, 87-100 Toruń

Print: Wydawnictwo Naukowe UMK
Edition: 300 copies
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sławomir Wacewicz, Przemysław Żywczyński</td>
<td>5</td>
</tr>
<tr>
<td>From the narrow to the broad. Multiple perspectives on language evolution</td>
<td></td>
</tr>
<tr>
<td>Michael Pleyer, James Winters</td>
<td>19</td>
</tr>
<tr>
<td>Integrating Cognitive Linguistics and language evolution research</td>
<td></td>
</tr>
<tr>
<td>Maciej Pokornowski</td>
<td>45</td>
</tr>
<tr>
<td>The fourth V, as in evolution: How evolutionary linguistics can contribute to data science</td>
<td></td>
</tr>
<tr>
<td>Ines Adornetti, Francesco Ferretti</td>
<td>63</td>
</tr>
<tr>
<td>The pragmatic foundations of communication: An action-oriented model of the origin of language</td>
<td></td>
</tr>
<tr>
<td>Sławomir Wacewicz, Przemysław Żywczyński, Luke McCrohon</td>
<td>81</td>
</tr>
<tr>
<td>Linguistic politeness from an ethological perspective: Theoretical questions and empirical issues</td>
<td></td>
</tr>
<tr>
<td>Daniela Lenti Boero</td>
<td>99</td>
</tr>
<tr>
<td>Sounds of protolanguages: Some preliminary insights from developmental psychology</td>
<td></td>
</tr>
<tr>
<td>Magdalena Igras, Stanislaw Kacprzak, Mariusz Mąsior, Mariusz Ziółko</td>
<td>117</td>
</tr>
<tr>
<td>The acoustic diversity in the phoneme inventories of the world’s languages</td>
<td></td>
</tr>
<tr>
<td>Till Nikolaus von Heiseler</td>
<td>129</td>
</tr>
<tr>
<td>How language evolved as a backchannel between two feedback loops</td>
<td></td>
</tr>
<tr>
<td>Joel Parthemore</td>
<td>159</td>
</tr>
<tr>
<td>The case for protoconcepts: Why concepts, language, and protolanguage all need protoconcepts</td>
<td></td>
</tr>
<tr>
<td>Agnieszka Dębska</td>
<td>179</td>
</tr>
<tr>
<td>Pointing in the right direction. A discussion of The evolution of social communication in primates, eds. M. Pina and N. Gontier</td>
<td></td>
</tr>
<tr>
<td>Sławomir Wacewicz</td>
<td>191</td>
</tr>
<tr>
<td>The shades of social. A discussion of The social origins of language, ed. Daniel Dor, Chris Knight and Jerome Lewis</td>
<td></td>
</tr>
</tbody>
</table>
From the narrow to the broad.
Multiple perspectives on language evolution.

Abstract. Although many of the recent controversies in the field of language evolution research are empirical, the deepest divides seem to remain theoretical in nature. Specifically, defining language in incompatible ways has led to radically different views on language evolution as a programme, including evaluation of its current success and future progress. Despite recent manifestos from the “narrow” camp (Hauser et al. 2014; Bolhuis et al. 2014), who along the lines of Hauser, Chomsky and Fitch (2002) equates language with the syntactic processor, the rival approach seems to be gaining momentum. It embodies a move in exactly the opposite direction, by understanding language broadly and assuming an inclusive perspective on its origins, which results in ongoing expansion of the field of language evolution. New areas of academic reflection (such as normativity) are being brought to bear, new areas of linguistics are being included (such as pragmatics or linguistic politeness); and, interestingly, existing linguistic methods are now being used to inform animal communication studies.

Keywords: language evolution, evolution of language, faculty of language, FLN, FLB, biolinguistics, biosemiotics
1. Introduction

Many of the recent controversies in the field of language evolution – the modern study of language origins – bear the hallmarks of empirical problems amenable to the methods of the Kuhnian normal science. This day-to-day “puzzle solving” may not be able to give us any “big” answers in the near future, but it is yielding a wealth of evidence to evaluate the competing hypotheses and weigh them against each other in an informed way. Were the Neandertals able to speak? The debate continues, but converging evidence (e.g. Johansson 2012, Dediu and Levinson 2013) systematically favours that possibility. Was there a gestural rather than vocal protolanguage? Semiotic experiments such as those by Fay et al. (2013) combined with data from e.g. primate communication (e.g. Slocombe et al. 2011) provide good reasons to believe that the first language-like communicative system could indeed rely on a significant visual component.

Interestingly, however, some of the deepest controversies in language evolution are not due to the shortage of good evidence, but rather result directly from the foundational theoretical problems. “Your theory of language evolution depends on your theory of language”, the title of a chapter by Ray Jackendoff (2010), may also be the most accurate summary of the status quo in this field of research: the deepest divide lies in defining the central notion of interest. In “Discussing the evolution of the assorted beasts called language”, Rudolf Botha (2000) offers a useful point in case: reading through the volume resulting from the first Evolang conference, he lists thirteen different conceptualisations of language, from a “process” to a “skill”, to an “activity”, to a “contract”. If we wanted a unifying dimension for organising all this diversity and the resulting difference of opinion, probably the most useful systematising polarity would be that between a narrow and an inclusive views.

2. Narrowing language down

The biolinguist has a quick but powerful fix. Prototypically, science deals with problems that are well-defined and tractable, so if we want language origin to enjoy the status of a scientifically legitimate and respectable question, we must pare down the all-encompassing and thus unwieldy

---

1 Which, notably, are informative in their own right, that is independently of the “origins” question.
From the narrow to the broad. Multiple perspectives on language evolution

notion of language to something of manageable proportions, preferably open to formalisation. In other words, the scientific way is to narrow language down to its bare computational bones, making it roughly synonymous with syntax\(^2\). The conceptualisation of choice is distinctly along the lines of the Chomskyan linguistic tradition, and the most important programmatic text is the influential 2002 paper by Hauser, Chomsky and Fitch (Hauser et al. 2002). On this view, language (narrowly defined) is the “computational core” responsible for “discrete infinity”—the recursively combinatorial potential characteristic of human verbal communication. Primarily, it is a cognitive capacity, but in an evolutionary sense the “language faculty” can be understood as a genetic constitution of an organism that gives rise to the cognitive capacity in the course of normal ontogenetic development. In this evolutionary sense, FLN, or language faculty narrowly understood, is a phylogenetically novel, biological trait.

What is important, Hauser et al. (2002) is commonly but erroneously thought to define FLN as “what is both uniquely human and uniquely linguistic”. In fact, the definition of FLN in the 2002 text is unequivocally as “the computational core”, with additional hypotheses that this computational core may be uniquely human and may consist only of recursion. The “uniquely human + uniquely linguistic” definition of FLN was only proposed three years later in a follow-up text by the same authors (Fitch et al. 2005). The fact that those authors never retracted the 2002 version lead to the two definitions functioning side by side, resulting in a regrettable terminological mess—see Wacewicz (2012) for discussion.

To someone interested in staking out the scientifically legitimate territory by equating it with the study of FLN, this definitional move has two parts. Firstly, it tells us what language is: the combinatorial machinery responsible for generating the discrete and hierarchically structured infinity of human linguistic expression, the computational core that is relatively tangible and formalizable. The biolinguistics formalisation almost obligatorily refers to the apparatus of the Minimalist Programme, leading critics to suspect “that ‘biolinguistics’ may turn out to be merely a more scientific-sounding term for generative minimalism” (Bickerton 2014: 74).

---

\(^2\) Cf. e.g. Bolhuis, Tattersall, Chomsky and Berwick (2014): “In our view, for the purposes of scientific understanding, language should be understood as a particular computational cognitive system, implemented neurally, that cannot be equated with an excessively expansive notion of <language as communication>.” “In place of a complex rule system or accounts grounded on general notions of <culture> or <communication>, it appears that human language syntax can be defined in an extremely simple way that makes conventional evolutionary explanations much simpler.”
But equally if not more important is the negative proposal: what language is not. Such a stance generates a sizeable list of “FLB” capacities (extended working memory, extended Theory of Mind, mental time travel and displaced representation, vocal imitation, cooperation with non-kin, etc.) and other phenomena that supposedly do not qualify into the lawful bounds; that is, at least not under the term of the evolution of language.

Not surprisingly, this negative stance is well reflected in the criticism of the mainstream language evolution research, to which the “narrow” (“biolinguistics”) camp is committed. As a particularly telling example, Hauser, Chomsky, Lewontin and others (Hauser et al. 2014) posit that:

“(1) studies of nonhuman animals provide virtually no relevant parallels to human linguistic communication, and none to the underlying biological capacity; (2) the fossil and archaeological evidence does not inform our understanding of the computations and representations of our earliest ancestors… (3) our understanding of the genetics of language is so impoverished that there is little hope of connecting genes to linguistic processes any time soon; (4) all modeling attempts have made unfounded assumptions, and have provided no empirical tests, thus leaving any insights into language’s origins unverifiable.”

Many of the problems highlighted by the recent statements from the Chomskyan camp (Hauser et al. 2014; Bolhuis et al. 2014) seem to be, in one way or another, a throwback to the influential text by Richard Lewontin (1998), “The evolution of cognition: Questions we will never answer”. Over a decade ago, Lewontin formulated a sceptical manifesto, pointing to numerous weaknesses in the chain of inferences from available evidence to adaptive (selection-based) accounts of the evolutionary emergence of particular cognitive traits. The reservations included poor definition of most cognitive traits, poor mapping to the biological (neuronal, genetic) correlates, their loose correspondence to the criteria for natural selection (such as heritability or measurable fitness advantage), and lack of direct evidence for function; plus scant fossil evidence and limited and partly inadequate comparative evidence. Symptomatically, a rhetorical

---

3 This general sentiment is probably best summarized in Noam Chomsky’s words at a UCL seminar, 10 October 2011: “There is a field called evolution of language, which has a burgeoning literature, most of which in my view is total nonsense.” (see e.g. http://linguistlist.org/issues/22/22-4631.html or http://www.evolutionarylinguistics.org/home/chomsky-on-language-evolution for reference)
question that he asks near the end of the text is “Why is anyone interested in the evolution of human cognition?”

But this sceptical attitude, if treated seriously, could write off a majority of sciences whose inferences go beyond strict formalisation. In 2014, the genetic and fossil evidence is much richer, including well established evidence of fire domestication from at least 700 kya (e.g. Roebroeks and Villa 2011), contra Lewontin’s (1998: 124) estimate of 100 kya, and comparative evidence is unimaginably more complete than the 1998 state of art. As to the definitions and operationalization of cognitive traits, this pertains equally well to any other cognitive science, but if anything, it is scientific theory that has a duty to track everyday meaning rather than vice versa. Scientific theory has every right to coin new technical terms but not to reach out to natural language to tamper with and impose on its established use, precisely what the syntactocentric theory is trying to do with “language”. Finally, the “why bother” question is just as relevant to the studies in great many other fields of basic research, so it cannot be used as an argument, unless we are prepared to let go of such disciplines as history or palaeontology.

3. The broad way – language evolution as an expanding field

Unsurprisingly, the “narrow”, or biolinguistic, perspective as outlined above, is now a minority position in the field. This is related to its major drawback: outside of a narrowly defined perspective, this restrictivenotion of language is simply not particularly interesting. Put differently, even a relatively complete answer about the emergence of FLN would hardly serve as a satisfactory explanation of the language origins problem in the form that is interesting to most people. When the question of the origin of language is posed as one of the foundational human questions, the notion of language there is definitely not the narrow sense appropriated by a syntactician. Features or phenomena such as common ground and online coordination, intersubjectivity, collective intentionality, massively cooperative exchange of information, mimetic communication, and multimodal proto-sign, may not be narrowly linguistic but are certainly preconditions for verbal communication, and are definitely the kinds of things about whose origins we would like to know. It is not clear what cognitive gains are supposed to result calling syntax “language” and ignoring the remaining aspects of human verbal communication (rather than, for example, studying the emergence of syntax and the emergence of language as parallel but interrelated projects).
Importantly, however, the “broad” view is getting broader. Insisting on the lexicon or aspects of phonetics and phonology as inalienably linguistic (and possibly unique to humans), as Pinker and Jackendoff (2005) did, is just a first step. Capacities that are inherently linguistic do not stop at syntax and its semantic/intentional and sensorimotor interfaces. Pragmatics is one example. Although – for reasons mostly related to theoretical prejudice – until recently it was neglected in the origins context\(^4\), a growing number of authors turn our attention to the basic fact that understanding of human communication rests on a rich pragmatic scaffolding (Scott-Phillips 2014, Adornetti and Ferretti, this volume). Linguistic politeness, an area that the standard view would see as totally peripheral, is also calling for attention from a naturalistic, evolutionary perspective (Wacewicz et al., this volume).

Going beyond FLN does not mean simply adding more individual-level capacities, but something more fundamental: stepping outside a single person’s mind and body to acknowledging language as a super-individual phenomenon. What Chomsky would call “E-language” may be external, but not epiphenomenal. There is a growing conviction, for example, that cultural evolution (e.g. Kirby et al. 2008) is not something second-order, kicking in only after we have biologically evolved (to use Arbib’s [2006] term) a “language-ready brain”; rather, language itself, as much as an individual capacity, is a cultural selection niche.

All of this is now being complemented with a move in the other direction. So far, language evolution accounts have benefitted enormously from a wealth of data on animal communication and cognition from comparative biologists and psychologists. But now, specifically linguistic techniques and methodologies are being deployed to analyze animal communication (and cognition). Examples include an application of Wierzbicka’s Natural Semantic Metalanguage to the study of conceptual repertoires in chimpanzees (Wierzbicka 2014: 156–182), examination of patterns of turn-taking in monkey vocalisations (Takahashi et al. 2013), semantic-pragmatic analysis of roots and affixes in Campbell’s monkey’s alarm calls (Schlenker et al., in press), and the general controversy regarding the applicability of a linguistic-morphological analysis to monkey calls (cf. Ouattara et al. 2009, Barceló-Coblijn and Gomila 2012).

4. Multiple perspectives on language origins

The present issue of *Theoria et Historia Scientiarum* is comprised of texts which – both individually and as a collective – unambiguously vote in favour of the latter approach. This is particularly visible in the first part of the volume, where each of the papers supports the inclusive perspective on language evolution directly, by making a case for the relevance of a particular linguistic aspect, phenomenon, or methodology. Michael Pleyer and James Winters (2014) put forward a case for a closer integration of Cognitive Linguistics into language evolution research, and Maciej Pokornowski (2014), in turn, calls for a closer rapprochement between language evolution and data science, specifically big text data. Ines Adornetti and Francesco Ferretti (2014) advance a pragmatic, action-oriented account of language emergence which takes into consideration the embodiment and embeddedness of human communication. Finally Sławomir Wacewicz *et al.* (2014) show the phenomenon of linguistic politeness, a neglected one in language evolution research, to clearly belong into the realm of naturalistic, evolutionary study.

The “inclusive” line continues in the second part of the volume, with the text by Daniela Lenti-Boero (2014), who demonstrates how early infant vocalization can be useful in informing research on the origins of speech. Speech is also the topic of Magdalena Igras *et al.* (2014), who propose a new approach to using phoneme inventories for tracing back language phylogenies. Till von Heiseler (2014) and Joel Parthemore (2014) both present new hypotheses: the former on the role of storytelling in language evolution, the latter on *protoconcepts*, onto- and phylogenetically basic units of prelinguistic but conceptual thought. The volume is concluded with two reviews – by Agnieszka Dębska (2014) and by Sławomir Wacewicz (2014) – both of books that explicitly refer to the social context and social origin of human communication.

4.1. New directions

Pleyer and Winters (2014) put forward the point that “we can gain important insights on the evolution of language and cognition by integrating evolutionary linguistics and the framework of Cognitive Linguistics [(CL)]]”. Their discussion focuses on two trends in CL: usage-based approaches, which view language structures as arising through abstraction from individual communicative events (Bybee 2010; Ellis 2013), and construction grammar with its notion of language as “a structured inventory or network
of constructions [...] defined as form-meaning pairings with varying degrees of schematicity [...]” (Goldberg 2006, Hilpert 2014). When interpreted as complex adaptive systems (Beckner et al. 2009: 2), the description of linguistic process afforded by usage-based approaches and Construction grammar can, according to Pleyer and Winters, help understand the evolution of language in the three scales identified by Kirby (2012, mentioned earlier in Kirby and Hurford 2002) – the ontogenetic timescale of individuals acquiring language, the glossogenetic timescale of historical language change and the phylogenetic timescale of the evolution of the species. In drafting their proposal for the synthesis of CL and evolutionary linguistics, the authors identify the common interest of the two fields in “the cognitive machinery [...] of language and use”, which helps account for the change of language systems and, if studied with regard to its biological prerequisites, can – as Pleyer and Winters hope – shed light on how language emerged in the first place.

In a similar vein, Pokornowski (2014) argues for a “closer interaction between data science and evolutionary linguistics”. He observes that the emerging field of Big Data has not yet benefitted from the interest of evolutionarily inclined linguists, and that the vast and quickly growing resources of digitalised language are becoming an important but relatively unexplored type of material for testing evolutionary hypotheses. Pokornowski then turns to Twitter as a paragon example of a massive open-access body of authentic language dataripe for investigation, including evolutionarily-inspired investigation. His claim is substantiated by putting forward a model, inspired by the Iterated Learning tradition, for studying the quasi-evolutionary changes at the level of topic-specific sublanguages on Twitter, with possibilities for extension to other social media sites.

Also in line with a broad and inclusive perspective on the study of language origins, Adornetti and Ferretti (2014) speak out against narrow views on language, such as the Chomskyan syntactocentric tradition or the code model. These authors underscore the grounding of language in context: to human users, language is not an abstract, disembodied processing machine, but a tool for interaction of real, flesh-and-blood agents, which they use in a specific external environment to achieve pragmatic communicative success. On Adornetti and Ferretti’s analysis, embracing a less abstract and more action-oriented and embedded account of language goes hand in hand with arguments for the gestural rather than vocal origin of human symbolic communication. Their claims are fleshed out with references to rich empirical work on the mirror system in primates.

Wacewicz, Żywieczański and McCrohon (2014) turn their attention to linguistic politeness, a phenomenon almost non-existent in evolutionary-
linguistic literature. They point to the fact that linguistic politeness acts as an interface between language and the “core questions of human ethology and human behavioural ecology”, such as normativity, cooperation and group hierarchy. While they do take note of the substantial difficulties in investigating politeness – for example, problems with an unambiguous definition, operationalization and measuring of politeness, and respecting the vast intercultural differences – they conclude that linguistic politeness deserves a place on a naturalistically oriented research agenda. Among the most interesting questions Wacewicz et al. point to is that of the evolutionary stability of politeness, which seems to be a consistently beneficial signal despite being “cheap” and manipulable.

4.2. New solutions

Lenti Boero proposes how arguments from developmental psychology, specifically those that relate to human vocal behaviours, can inform the research on early, protolinguistic, beginnings of language. Careful to avoid the recapitulationist fallacy, Lenti Boero examines the developmental pathway of human infants – from the earliest stage characterized by crying through musibabling (her own concept referring to infants’ engagement in protophonic productions) until the onset of the universal speech perception at the age of six months – in search for the vocal precursors of language. In doing so, she appeals to the arguments both from parental selection (Locke 2006) and sexual selection (as envisaged by Darwin 1871 and later by Mithen 2006, 2009). Her paper also contains an interesting discussion on vocal mimicry and the viability of onomatopoeic protolanguage in the light of neurocognitive research (Hashimoto et al. 2006).

The paper by Igras et al. (2014) deals with measuring phonemic inventories in languages and the possibilities of applying this knowledge to inform research on language phylogenies. This approach was pioneered by Atkinson (2011), who suggested that the notion of founder effect from population genetics could be applied to tracing back migrations of populations of language users across considerable time scales. Igras et al. compiled a database with audio material from over 3500 languages, relying mostly on the material from the Global Recording Network but complemented with other sources. They have been developing methods of automated segmentation of this data into phonemes and computing distances between languages; despite initial equivocal results, they are aiming to bring this methodology to bear on the linguistic serial founder debate.
Von Heiseler faces a tough task of forming a unified evolutionary explanation that would account for most of human socio-cognitive uniqueness. He envisages a scenario of super-fast evolution in the hominin line, based on two interlocking mechanisms: group-level selection through bellicose conflict between hostile bands, and sexual selection for the bravest warriors (feeding back into the group’s military performance). As the females cannot observe the brave deeds directly, they must rely on narrative descriptions, which in time might lead to selection for the most proficient storytellers. There is no doubt that Von Heiseler’s account is speculative and controversial. But while it turns on several risky assumptions, the paper also contains a number of daring and innovative proposals.

Parthemore (2014) considers the problem of primacy between language-like communication (which apparently relies on concepts) and conceptual thought (which on some influential accounts derives from language). Parthemore rejects the radical positions of Fodor’s nativism and McDowell’s conceptualism to propose his own moderately nativistic Protoconcept Hypothesis, on which human cognition comes innately equipped with a small set of protoconceptual categories, which then give rise to progressively more complex conceptual repertoires. In short, we are innately biased to experience and interact with the world in some ways and not in others, beyond what our embodiment contributes. Basic protoconcepts correspond to some of the most superordinate categories organizing our experience, such as proto-objects, proto-happenings, and proto-properties (in English mapping to the grammatical categories of nouns, verbs, and adjectives and adverbs). Complementing them with several quasi-logical operators would yield “a minimally structured logic both far simpler and at the same time far more expressively powerful than existing formal logics” – a system capable of grounding protolanguage.

4.3. The social dimension

Increasingly many recent publications on language origins refer to the social perspective (e.g. Scott-Phillips 2014, Dunbar et al. [ed] 2014). Dębska (2014) reviews one of them, The Evolution of Social Communication in Primates (ed. Pina and Gontier 2014). Rather than offering a point-by-point report on the book, Dębska identifies and developssome of the threads in the book that “define the key points of the current debate on language evolution”, such as the flexibility vocalisation in apes or the extent of their Theory-of-Mind capacities. Dębska’s specific focus is on the surprisingly complex communicative behaviour that pointing turns out to be. She
discusses the differences of perspective between contributors to the volume, mirroring those in the field at large, as to whether (untrained) chimpanzees and other apes are capable of declarative pointing or of understanding human declarative pointing. Dębska concludes that in this subfield of communication theory “there is a need to develop at least minimal empirical criteria for basic concepts”.

Finally, Wacewicz (2014) reviews *The social origins of language* (ed. Dor, Knight and Lewis 2014), which in his view “aspires […] to the role of an important programmatic statement”. After contemplating the possible definitions of “the social”, he turns to the major threads of the book: the problem of honest, cooperative signalling in humans, of cultural evolution, and of the multifaceted nature of language. This last point is particularly important. Wacewicz fully endorses the inclusive outlook on the study of language (origins), which is represented in the reviewed volume and is best captured in a quote from the introduction by the editors (Dor, Knight and Lewis 2014: 2): “more and more of what we learn about language remains confined to specialized professional circles. However, to understand the origin of language requires a move in the opposite direction—a large-scale, collective interdisciplinary effort at theoretical synthesis. […] Everything counts.”

**Acknowledgements**

This research was supported by grant UMO-2012/07/E/HS2/00671 from the Polish National Science Centre.

**References**


Dor D., C. Knight and J. Lewis (eds.). (2014). *The social origins of language*.


Abstract. In this paper we argue that we can gain important insights on the evolution of language and cognition by integrating evolutionary linguistics and the framework of Cognitive Linguistics. In Cognitive Linguistics, language is seen as tightly integrated with cognition as a whole. Construction Grammar and usage-based approaches are closely related to the Cognitive-Linguistic paradigm. Construction Grammar proposes that knowledge of language can be defined as the knowledge of form-meaning pairings of different degrees of schematicity and complexity, whereas usage-based approaches stress that language acquisition and processing are based on instances of actual language usage. As we demonstrate in this paper, concepts from Cognitive Linguistics, construction grammar, and usage-based approaches can help in elucidating the cognitive and interactional factors involved in language evolution. The paper will focus on two main areas: In evolutionary linguistics, language is seen as a complex adaptive system whose structure emerges out of the interaction of three other complex adaptive systems at three different timescales: ontogeny, glossogeny, and phylogeny. Cognitive Linguistics can help in specifying common cognitive factors and processes that play a role on all three of these timescales. Secondly, a Cognitive-Linguistic and constructionist, usage-based perspective can shed light on the cognitive factors underlying the origin of the division of labour between contentful (“lexical”) and procedural (“grammatical”) constructions in language structure. In a Cognitive-Linguistic perspective, this development can be related to the trade-off between the cognitive factors of learnability and expressivity.
In this paper, we argue that there is a wealth of relevant research and theorizing in Cognitive Linguistics that can make important contributions to the study of the evolution of language and cognition. Specifically, we will argue that CL is very well-suited to specify the complex set of the underlying cognitive skills, capacities and processes that language use, structure, acquisition and evolution depend on. Furthermore, we argue that Cognitive Linguistics and evolutionary linguistics are ideally suited for integration as they both adopt a view of language as a complex adaptive system. In evolutionary linguistics, the complex adaptive system of language is seen as consisting of three interacting timescales. These in themselves each are complex adaptive systems, and they are all relevant to the emergence of the complex adaptive system of language: ontogeny (the level of the individual learning a language), glossogeny (the level of historical language change in populations) and phylogeny (the level of the biological evolution of the species) (Beckner et al. 2009; Kirby 2012; Kirby and Hurford 2002). A Cognitive-Linguistic perspective can help identify common patterns and factors that are relevant to all three timescales. In this way, Cognitive Linguistics can make significant contributions to evolutionary linguistics by showing how cognitive and usage factors play a role in all three complex adaptive systems that are involved in the emergence of language. In adopting a complex adaptive system perspective on language, CL can thus help to investigate cognitive, socio-cultural, embodied, conceptual and other factors involved in the evolution of human language on all three levels.

Cognitive Linguistics (CL hereafter) is a school of linguistic theory and practice that sees language as an integral part of cognition and tries to explain linguistic phenomena with relation to general cognitive capacities (e.g. Evans and Green 2006; Geeraerts and Cuyckens 2007). Language, in this view, is seen to draw on mechanisms and principles that are not language-specific but general to cognition. CL is closely related to so-called usage-based approaches (e.g. Barlow and Kemmer 2000) as well as to construction grammar (e.g. Hoffmann and Trousdale 2013). In combination with CL, these approaches have important implications for evolutionary linguistics, which we will address in this paper.
Integrating Cognitive Linguistics and language evolution research

Usage-based approaches explain language structure as being based on processes of abstraction and schematization from instances of actual language use in context. In these approaches, the formation of linguistic structure is argued to proceed via the repetition and entrenchment of patterns in language use in richly social interactive contexts which get conventionalized in a community. Cognitive and communicative processes employed in interpersonal feed into and shape the emergence of linguistic structure (Bybee 2010; Ellis 2013). Usage events and the strategies and cognitive processes employed by language users in actual language use thus become the key focus of interest in this approach. This has important ramifications for research on language evolution, which similarly sees language as emerging ‘bottom up’ through the interaction of multiple complex adaptive systems. One of the key questions from a usage-based perspective then becomes how linguistic constructions emerge through language use and build a constructional network.

In construction grammar, and constructionist approaches more generally, language is seen as a structured inventory or network of constructions, which are defined as form-meaning pairings with varying degrees of schematicity and abstraction (Goldberg 2006; Hilpert 2014).

Such form-meaning pairings include fully specified constructions such as morpheme constructions (e.g. anti-, un-, -ing), word level constructions (e.g. bear, banana, but), as well as partially filled constructions (such as [AGENT] crane [AGENT]+POSS neck, cf. Langacker 2008) and idioms (e.g. kick the bucket, jog X’s memory). At the most abstract pole of the continuum of constructions they include fully abstract constructions such as the Ditransitive Construction, which has an abstract, schematic meaning of transfer (SUBJ V OBJ1 OBJ2, e.g. He baked her a vegan cupcake), or the Caused-Motion-Construction (S V OBJ ADV, e.g. She sneezed the napkin off the table) (Goldberg 2006).

Cognitive-Linguistic research is driven by a number of core assumptions, which also have important implications for studies of the evolution of language (cf. Geeraerts 2006). First of all, in CL language is seen symbolic and conceptual: Language expresses and evokes conceptualizations, using particular cultural models and cognitive resources (Fauconnier 2004; Langacker 2008; Croft and Cruse 2004). Linguistic utterances thus function as instructions or “prompts” for the dynamic construction of a mental representation by the hearer (e.g. Evans and Green 2006). Following from this, CL also focusses on the interactive nature of language. The main function of language on this view is the dynamic, collaborative construction of meaning in interaction. Finally, the conceptual and interactive function
of language also entail that language is fundamentally perspectival: all linguistic expressions and conceptualizations are always tied to a specific perspective or point of view. In CL, this general phenomenon is captured with the concept of construal: when speakers structure a conceptualization for purposes of expression, they structure it in a specific manner (Langacker 1987: 126). Language enables speakers to express different conceptual perspectives on the same referent, situation, or event. The structured inventory of constructions which language users possess enables them to construe a situation in many different ways and from multiple perspectives (Geeraerts 2006). Speakers have available different construal operations in their language that enable them to assign salience to different aspects of a conceptualization. In this way, the same conceptual content can be viewed and construed in different ways (Langacker 2008: 43f.). For example, the same conceptual content, e.g. a glass that contains water, can be linguistically construed in different ways:

(1) the glass with water in it designates the container; (2) the water in the glass designates the liquid it contains; (3) the glass is half-full designates the relationship wherein the volume occupied by the liquid is just half of its potential volume; and (4) the glass is half-empty designates the relationship wherein the volume occupied by the void is just half of its potential volume. (Langacker 2008: 43)

One of the main goals of CL is to uncover the cognitive capacities and mechanisms that support and underlie the symbolic, conceptual, interactive, and perspectival nature of language. It is in regard to this that CL can make significant contributions to evolutionary linguistics (cf. Pleyer 2012). Some recent work has already begun to apply Cognitive-Linguistic insights to questions of language evolution and change (e.g. Bybee 2012; Sinha 2009; Tomasello 2008).

As we will outline, one particularly fruitful point of contact is that in evolutionary linguistics language is treated as a complex adaptive system (henceforth, CAS; Steels 2000; Mufwene 2001). Importantly, this perspective of language as a CAS has also been adopted in Cognitive-Linguistic and usage-based approaches (Beckner et al. 2009; Winters et al. 2010). On this view, language is seen as an emergent product of the interaction of multiple dynamical systems.

There are two areas closely related to the complex adaptive system perspective in which we will show that insights from CL can help in tackling problems in evolutionary linguistics.
The first concerns the problem of linkage between our biological capacity for language and the structure of language. This problem can be solved by looking at the cognitive factors involved in learnability and expressivity in combination with the dimension of sociocultural transmission.

The second avenue where CL can help in elucidating problems in evolutionary linguistics is the origin of the contentive/functional divide in language. Here, we argue, constructionist concepts of grammaticalization and constructionalization can give important insights into how this divide might have originated in the evolution of language.

In the next section (2), we turn to the notion of complex adaptive systems and illustrate the implications of research in CL and usage-based approaches for the ontogenetic (2.1), glossogenetic (2.2) and phylogenetic (2.3) timelines of the complex adaptive of system of language. Then we will turn to two specific case studies that show how CL and usage-based approaches offer a promising perspective to the study of language evolution. First, we will turn to the problem of linkage and the cognitive factors that influence the relationship between language strategies and language systems (3). Then we will address the evolution of the division of the contentive/functional divide (4), before offering some concluding remarks (5).

2. The three complex adaptive systems of language

The term complex adaptive system (CAS) was first coined by Holland (1992). CAS are similar to other complex systems in that they exhibit emergent properties as a result of multiple interconnected elements. But it is their capacity to evolve, learn and adapt that ultimately differentiates them from other complex systems: past experiences filter through, or influence, future states of the system due to a cumulative process (Deacon 2010). Like evolutionary (Whitacre and Bender 2010) and developmental (Edelman and Gally 2001) systems, language displays all the hallmarks of a CAS:

(a) The system consists of multiple agents (the speakers in the speech community) interacting with one another. (b) The system is adaptive; that is, speakers’ behavior is based on their past interactions, and current and past interactions together feed forward into future behavior. (c) A speaker’s behavior is the consequence of competing factors ranging from perceptual mechanics to social motivations. (d) The structures of language emerge from interrelated patterns of experience, social interaction, and cognitive processes. (Beckner et al. 2009: 2)
A CAS perspective can show how general properties emerge from local interactions via an amplification dynamic: Large ensembles of interacting elements become expressed as system-wide characteristics as a result of interacting constraints and biases (Deacon 2010: 124). As we discuss in section 3, solving local communication problems through short-term strategies, such as usage events, can lead to the emergence of long-term patterns and structures that form language systems. In turn, these language systems then constrain the types of strategies that speakers employ. This idea of feedback loops that exist on multiple timescales is shared in much of the current work in evolutionary linguistics. Here, language arises through the interactions of three complex adaptive systems operating on different timescales (Kirby 2012):

1. The *ontogenetic* timescale of individuals acquiring language
2. The *glossogenetic* timescale of historical language change
3. The *phylogenetic* timescale of the evolution of the species.

In this section, we will show that a Cognitive-Linguistic perspective can direct attention to some of these common factors.

### 2.1. The ontogenetic timescale of individuals acquiring language

On the *ontogenetic timescale*, Cognitive-Linguistic and usage-based approaches focus on the cognitive factors and social scaffoldings influencing the acquisition and learning of language.

In other words, these approaches focus on the factors underlying learnability and expressivity in acquisition, which will be dealt with more thoroughly in sections 3 and 4.

Two general kinds of factors are seen to support the acquisition of language: *sociocognitive capacities and motivations* on the one hand and *domain-general cognitive mechanisms* on the other. The sociocognitive capacities that support language acquisition include processes of mutual coordination and cooperation, joint attention, shared intentionality, and perspective-taking (cf. Clark 1997; Tomasello 1999, 2008). Joint attention is the ability to attend to the same situation together with another person in triadic engagement. In this form of engagement attention is directed at both the other and an event in the outside world. One particular area where the importance of sociocognitive capacities is especially evident is that of word learning. For example, human infants reliably use gaze following to learn about objects and events from 12 months onwards (Flom and Johnson 2011). At 18 months of age, children learn to associate a word with the object
the adult is looking at, not the one they are looking at (Baldwin and Moses 2001). At 24 months, children associate a new word with an object that is new from the adult’s perspective, but not from their own (Akhtar et al. 1996). As experimental evidence from infants as young as 14 month olds shows, this latter ability rests on a rudimentary but steadily developing understanding of perspectives in interaction (Moll and Tomasello 2007; Tomasello and Haberl 2003).

Overall then, the sociocognitive foundations of language acquisition are already evident in the production and comprehension of declarative pointing in prelinguistic infants. This is especially interesting as these capacities appear not to be present in non-human primates to the same degree (Tomasello 2008; Miklosi and Soproni 2006). Infants at the age of 12 months already begin to show some species-unique ways of directing and sharing attention, such as holding things up so they can show them to others, or pointing to interesting situations and events. These pave the way for the acquisition of language as they establish a referential triangle between a “me,” a “you” and an “it” analogous to speaker, listener, and topic (Tomasello 2007: 1092). As Ibbotson (2011: 332) points out, this view of development “predicts there should be a close correspondence between language emergence and social-cognitive abilities and indeed research shows children’s emerging linguistic skills are predicated on their ability to engage in nonlinguistically mediated joint attentional activities.” These sociocognitive foundations thus can be seen as a crucial evolutionary foundation for the emergence of language. Another important sociocognitive foundation that is already present at 14 months of age in humans, but absent in non-human species, is shared intentionality. Tomasello and colleagues have argued that shared intentionality is one of the most crucial foundations for language acquisition. It can be defined as the skills and motivations to form joint goals and intentions with others and to share psychological states with them (Tomasello et al. 2005; Tomasello 2008). Shared intentionality is tightly connected to another important sociocognitive prerequisite for learning and using language, namely understanding and establishing common ground. Understanding common ground means being able to understand others’ perceptions, engage with them in joint attention and create a common conceptual ground for joint understanding and cooperation. Their understanding of pointing gestures in cooperative situations shows that starting from 14 months onwards, young children become increasingly proficient in this task. In the context of language acquisition, this is particularly significant, as cooperative pointing rooted in common ground
creates different conceptualizations or construals of things. These presage the ability of linguistic creatures to place one and the same entity under alternative different ‘descriptions’ or ‘aspectual shapes’, which is one of the hallmarks of human conceptual thinking, but it does this without the use of any conventional or symbolic vehicles with articulate semantic content (Tomasello 2014: 57).

Pointing gestures thus already have the potential to embody different construals. For example, depending on the context and common ground, pointing at a piece of wood can construe it as firewood, an obstacle that needs to be removed, a crutch if you just twisted your ankle, a suitable weapon for a pretend play swordfight, etc. (cf. Tomasello 2014: 57). These sociocognitive foundations are thus closely related to the notion of construal and a more generalised understanding of perspectives, which are crucial to children’s language acquisition and sociocognitive development in general (Clark 1997; Moll and Tomasello 2007; Pleyer 2014).

From an evolutionary perspective, these sociocognitive capacities thus present important prerequisites for the emergence of language.

The general cognitive mechanisms employed by children when learning a language include, for example, statistical learning (e.g. Romberg and Saffran 2010), categorization (Taylor 2012; Bybee 2010), generalization and schematization (e.g. Langacker 2000; Ibbotson 2011), analogy (e.g. Gentner and Christie 2010), entrenchment (e.g. Lieven 2010), chunking and automatization (Bybee 2010).

In a usage-based perspective, children use their general cognitive abilities for pattern-finding to abstract concrete and specific items from instances of actual language use around them and then start generalizing over these concrete instance to arrive at more and more abstract constructional schemas (Tomasello 2003). Starting from 18 months onwards, children start using combinations of words, thus partitioning an experiential situation into several symbolic units. At the same age, young children also partition scenes conceptually by means of systematized pivot schemas. These kinds of multi-word utterances are organized locally around particular concrete words with one abstract slot, e.g. More___: more juice, more play, etc. (Tomasello 2011: 244). Starting around their second birthday, children start using some verbs in item-specific frames that are more complex, e.g., Draw ___; Draw ___ on ___; Draw ___ for ___; ___ draw on ___. In contrast to pivot schemas, these already employ some syntactic marking (Tomasello 2011: 245). In their second year of life children then begin to develop more abstract constructions that
contain even less particular lexical items. In these constructions, the abstract slots have a clearly specified functional role and the constructions as a whole have specific communicative functions (Tomasello 2011: 246f.).

Taken together, these factors enable children to internalize the linguistic abstractions they make from instances of actual language use in context. That is, with the support of these mechanisms they acquire linguistic constructions that are grounded in context and common ground established through cooperative activity (Clark 1996; Tomasello 2014). Their knowledge of language, then, consists of a network of related perspectival constructions “in which the same event is construed from different perspectives so that speakers can choose the construction that is most appropriate to realize their communicative intention in a particular situation” (Diessel 2013: 357).

2.2. The glossogenetic timescale of historical language change

On the glossogenetic timescale, cultural transmission and historical language change in dynamic populations are determined by social and cognitive factors as well as emergent properties of the transmission process (e.g. Deacon 2010; Hruschka et al. 2009; Kirby 2012). As will be discussed below in more detail, languages are shaped by the brain in such a way that they are learnable (Christiansen and Chater 2008; Deacon 1997; Brighton, Kirby and Smith 2005). Thus, languages adapt to be transmitted and learned through domain-general cognitive capacities and constraints. There is a multiplicity of constraints and factors, including constraints of body, brain, environment, and culture that play a role in this process. Given these constraints that influence the interactions and usage events of individual speakers, the process of transmission then leads to languages changing and gives rise to language structure. Generally speaking, in addition to the factors listed above, language structure and usage are “shaped around human learning and processing biases deriving from the structure of our thought processes, perceptuo-motor factors, cognitive limitations, and pragmatic constraints” (Christiansen and Chater 2008: 490). The specific nature of the process of cultural transmission and selection also plays a role in the shaping of language structure (Kirby 2012; Steels 2011), as do emergent properties related to cultural conventions, discourse factors, and semiotic constraints (Beckner et al. 2009).

The existence of domain-general biases also highlights the overall importance of development on the ontogenetic timescale. Humans show a great degree of developmental flexibility, which itself is seen as an evolutionary system (Deacon 2010). In human development, domain-general
mechanisms act as slight biases, which allow developmental pathways to explore the functional space of the cultural environment.

As we will further illustrate in sections 3 and 4, CL and usage-based approaches can help to unravel the interaction of general cognitive mechanisms and cultural transmission in influencing language change and the emergence of structural patterns (e.g. Beckner et al. 2009; Bybee 2010; Christiansen and Chater 2008; Winters et al. 2010). Through the repeated interaction between individuals we can observe regular patterns of change at the glossogenetic level. Language emerges as a product of its underlying speech community, but also adapts to the very dynamics from which it emerged (Beckner et al. 2009). As Diessel (2012: 1609) points out, the cognitive processes of analogy, entrenchment, and categorization play an important role both in language acquisition and language change. These factors thus are not only crucial for language acquisition; they also enter into the glossogenetic emergence of linguistic structure. The cognitive factors discussed in this and the previous sections also interact with and influence construal: Speakers use their ability for perspectival construal when they use particular linguistic strategies to categorize and linguistically structure situations in specific ways. These perspectival choices influence linguistic structure through processes of entrenchment and conventionalization (see section 3, cf. Evans and Green 2006: 110). Entrenchment denotes “the strength of autonomy or representation of a form-meaning pairing at a given level of abstraction in the cognitive system.” (Blumenthal-Dramé 2012: 4) Through usage events, linguistic structures that occur frequently in discourse and are salient become consolidated in memory, are processed sequentially and are stored in automated chunks (Bybee 2010). Again, these cognitive factors that affect the mental representation and processing of linguistic structure are domain-general in nature. This is stressed, for example, by Langacker:

Automatisation is the process observed in learning to tie a shoe or recite the alphabet: through repetition or rehearsal, a complex structure is thoroughly mastered, to the point that using it is virtually automatic and requires little conscious monitoring.” Automatisation leads to the progressive entrenchment of a structure, which “eventually becomes established as a unit. Lexical items are expressions that have achieved the status of units for representative members of a speech community (Langacker 2008: 16f.).

Through processes of analogy and categorization linguistic structures instantiated in usage events can also become more schematized and abstract.
Overall then, on this view domain-general processes serve as the basis for the emergence of grammar and linguistic structure (Bybee 2012).

Even though historical contingency is considered an important factor in diachronic language change (Lass 1997), there are also language-external biases and constraints that influence the continual cycle of innovation, amplification and fixation (Croft 2000). Still, these paths of change are far from pre-determined, with the trajectories of change being much more similar than the resulting states (Beckner et al. 2009: 7). Overall then, structural patterns emerge through processes such as grammaticalization (e.g. Beckner et al. 2009; Bybee 2010; Traugott and Trousdale 2013; see section 4).

Explaining the emergence of systematic structure through the processes of social transmission and interaction has become a central goal in the study of the cultural evolution of language (Scott-Phillips and Kirby 2010). This is why insights from CL and usage-based approaches could be highly profitable for this enterprise. In section 4, we will turn to one particular proposal how general cognitive mechanisms can influence language change, namely the division of labour in the construction between contentive and functional items (Kirby 2013).

2.3. The phylogenetic timescale of the evolution of the species

On the phylogenetic timescale, CL focuses on the biological evolution of the species and of the uniquely human multi-component suite of skills that enables language learning and production (e.g. Tomasello 2003, 2008; Christiansen and Chater 2008). Consequently, CL can help specifying the cognitive and representational capacities that had to evolve beyond those found in other animals in order to support language and the interactive and dynamic processes of meaning construal fundamental to linguistic interaction. From a phylogenetic perspective one of the key questions is to what extent the cognitive capacities discussed in sections 3.1 and 3.2 have homologues and analogues in other animals, especially the other great apes. This enterprise is well underway. Steels (e.g. 2004), Hurford (2012), Arbib (2012) and Bybee (2012), for example, have all adopted a constructionist or usage-based perspective on the question of language evolution and have specified some of the cognitive prerequisites that needed to evolve to make the human brain “language-ready” (Arbib 2012; cf. Pleyer and Lindner 2014). On the one hand, these concern the domain-general processes crucial to language acquisition and transmission discussed in sections 2.1 and 2.2. In addition, there are other important factors that such approaches can shed light on, such as the evolution of the cognitive capacity for “massive storage”.
(Hurford 2012: 261ff.) of constructions and exemplar representations (Bybee 2012; Taylor 2012). In CL, meaning and linguistic knowledge are seen as encyclopaedic, i.e. they are tightly connected to our general knowledge of the world and other conceptual domains (Langacker 1987, 2008). Related to this, constructions are hypothesized to be stored much in the same way as other types of knowledge (Goldberg 1995; Bybee 2010; Traugott and Trousdale 2013). This means that from a phylogenetic perspective, the evolution of the ability for the massive storage of all types of representations was a key development enabling language-readiness. The domain-general capacities underlying language acquisition thus are also crucial for the constitution of a language-ready brain. These include statistical processing that is sensitive to frequency effects, abilities for categorization, generalization and schematization, which enable the recognition of analogical similarities between stored exemplars, and the processes involved in entrenchment, and neuromotor automation. Other evolutionary requirements are the ability to voluntarily retrieve form-meaning pairings from long-term memory and increased hierarchical processing capacity (Hurford 2012; Tallerman 2009; Pleyer and Lindner 2014). These and other domain-general abilities such as the ability to make inferences based on context and common ground, and the ability to cross-modally associate meanings with vocal or manual signals, also can be found to different degrees in non-human primates and other animals. However, “the human cognate capacities are much richer, both in detail as well as degrees of abstractness” (Bybee 2012: 536).

On the other hand, the cognitive prerequisites for language concern the socio-cognitive foundations of language discussed in the previous sections. Among these, the capacities for construal and reciprocal perspective-sharing in discourse characterized by shared intentionality seem to be among the most important. As Tomasello (2014: 68) puts it: “Communicators conceptualizing or perspectivizing things in different ways […] and then recipients comprehending the intended perspectives through socially recursive inferences, is not the result of becoming a language user, but rather its prerequisite.” Similarly, Tomasello et al. (2005: 690) emphasize that:

Saying that only humans have language is like saying that only humans build skyscrapers, when the fact is that only humans (among primates) build freestanding shelters at all. Language is not basic; it is derived. It rests on the same underlying cognitive and social skills that lead infants to point to things and show things to other people declaratively and informatively, in a way that other primates do not do, and that lead them to engage in collaborative
Integrating Cognitive Linguistics and language evolution research

and joint attentional activities with others of a kind that are also unique among primates.

From a phylogenetic perspective, the evolution of the shared intentionality infrastructure is thus of crucial importance. These and other sociocognitive proclivities are already evident in humans from a very early age on and together with domain-general pattern-finding capacities build the foundations for language acquisition (Tomasello 1999, 2003, 2008). Evidence from comparative psychology suggests that other primates also possess a rich set of socio-cognitive capacities in the domain of understanding goal-directed actions, head orientation, eye orientation, and gaze. Moreover, great apes also exhibit at least rudimentary perspective-taking abilities. For example, “chimpanzees, like humans, understand that others see, hear and know things” (Call and Tomasello 2008: 190). However, the structure of social perspective-taking and -setting in humans goes well beyond the capacities exhibited by other primates. Humans do not only understand and take other people’s perspectives, but, in contrast to chimpanzees, they make use of their perspective-taking capabilities in fundamentally cooperative, declarative, and informative kinds of communication (cf. Tomasello 2008). Chimpanzees, for instance, have trouble understanding perspectives in non-competitive, cooperative interactions and neither produce nor understand declarative, perspective-sharing pointing gestures. As has been outlined in section 2.1, in human children, on the other hand, this ability and motivation is a crucial foundation for the acquisition of language (cf. Tomasello 2003, 2008). This means that the evolution of the human drive to share perspectives and psychological states with others was of fundamental importance in the evolution of language. Biological factors underlying these prosocial cooperative motivations and the drive to share perspectives and attitudes are thus among the key evolutionary adaptations that enabled the evolution of human language. Overall, then, on the phylogenetic dimension, CL directs attention to the evolution of the factors underlying the learnability and expressivity-functions of a network of constructions, a point that we will elaborate on in the next sections.

3. Solving the problem of linkage

In developing an explanatory framework in linguistics, we need to take seriously the disconnect between *explanans* (our biology) and *explanandum* (language structure): that is, we are faced with a *problem of linkage* (Kirby
Rather than there being a straightforward link between our individual cognitive machinery and the features we observe in language, we are instead faced with an additional dynamical system: socio-cultural transmission. Treating language as a CAS (see section 2) solves this problem of linkage because we can consider how short-term language strategies (Evans and Green 2006: 110) can give rise to language systems through long-term patterns of learning and use (Bleys and Steels 2009; Steels 2012).

Language strategies are heuristics an idiolect draws upon to convey their interactional goal and include the use of speech acts and our choices over constructions (Evans and Green 2006). Take the relatively simple example of referentially signalling a particular pool ball. Besides non-linguistic strategies, such as pointing, we can draw upon a whole host of referential labels, such as ball. However, these choices are constrained by the usage events and contexts in which they are situated. For example, in fig. 1 the labels ball and pool ball would be useless, unless additional information is provided. This can be done through compounding (e.g., pass me the cue ball) or by providing referentially useful information on another dimension (e.g., pass me the white).

A language system, then, is a collection of these strategies, which form a group of paradigmatic choices (Steels 2012). For instance, German features a case marking system made up of the paradigmatic choices of nominative, accusative, dative, and genitive. Importantly, not only do these strategies influence the emergence of linguistic systems, the linguistic systems themselves act as constraints on the choice of strategies (ibid.).
What strategies are used depends on a whole host of contextual factors that are particular to each usage event. This explains why we observe a high degree of historical contingency in the cultural evolution of language; that is, change proceeds generally via localised interactions (Lass 1997; Beckner et al. 2009). Still, even if we grant historical contingency a large role in language, we cannot avoid observations of systematic regularities both in how it is structured and how it changes over time. Predictable patterns of change, such as grammaticalization (see section 4), are argued to underpin general, cross-linguistic patterns of similarity (cf. Evans and Levinson 2009; Beckner et al. 2009). What we want to claim for the rest of this section is that the factors governing linguistic strategies, and the subsequent emergence of linguistic systems, can broadly be classified under pressures for learnability and expressivity (Smith, Tamariz and Kirby 2013).

Learnability refers to our limited exposure to input data as well as domain-general limitations in our memory and processing capabilities (Christiansen and Chater 2008): Patterns that are difficult to learn are harder to reproduce (Brighton, Kirby and Smith 2005). However, language is not merely a task of passively remembering and reproducing a set of constructions. Language is also a social and interactional phenomenon, whereby the role of usage, communication and coordination are salient pressures on the system (see also: Tomasello, 2008; Bybee, 2010; Winters, Kirby and Smith, 2014). This is the expressivity pressure.

It is only when there is a balance between these two pressures of learnability and expressivity that we do observe the emergence of systematic structure (Smith, Tamariz and Kirby 2013). If we remove the pressure for expressivity, then the learnability pressure gradually leads to a degenerate system consisting of one form for every meaning. Conversely, if we remove the learnability pressure, then we tend to get holistic systems where there is a non-systematic (arbitrary) one-to-one mapping between forms and meanings.

The take home point we want to stress is that these pressures of learnability and expressivity are acting locally on individual constructions and that they are strategically deployed in context. These strategies then become system-wide characteristics and influence the long-term emergence of language structure. Unifying cognitive and evolutionary linguistics provides a framework in which researchers can identify the cognitive machinery and processes underpinning learning and use, how linguistic systems change over time, and ultimately how they evolved in the first place. This allows us to ask the question how language evolved given its biological and cognitive machinery. That is, we can examine how languages pattern
synchronously, as well as how they change diachronically, and use this to infer what biological features underpin these processes.

4. Division of labour in the constructicon

Common to many theories in linguistics is the notion of a division of labour between *contentful* ("lexical") and *procedural* ("grammatical") constructions (e.g., Talmy 2000). Generally speaking, contentful constructions are used referentially, whereas procedural constructions are normally considered more abstract, and signal linguistic relations, perspectives and deictic orientation (Traugott and Trousdale 2013): that is, they “contribute information about how to combine [...] concepts into a conceptual representation” (Terkourafi 2011: 358-359).

There is considerable debate over the nature of these divisions and what constitutes a member of one category over another (e.g., Cann 2001; Boye and Harder 2012; Traugott and Trousdale 2013). For instance, even though there is overlap in the descriptive terms such as *open class* and *closed class*, *lexical* and *grammatical*, *contentive* and *functional*, *contentful* and *procedural*, these terms are not always used in a strictly synonymous manner. One example is the gradation found in English adverbs. Manner adverbs, such as *foolishly* and *fast*, pattern on the contentful end of the continuum, while focus marking adverbs (*only* and *even*) and degree adverbs (*very* and *quite*) cluster on the procedural end (Traugott and Trousdale 2013).

Irrespective of how we choose to classify the division, and the degree to which these divisions are categorical or gradient, the point remains that there are observable patterns in the structure of language. With this in mind, Kirby (2013) considers the division of labour to be a major transition in the cultural evolution of language (alongside *combinatorial phonotactics* and *compositionality*), and posits that the emergence of these new replicators, with a specifically syntactic function, might be driven by processes we observe in language change, namely *grammaticalization*: “What remains an open question at this stage is if we can demonstrate that the same process leads to the emergence of the very first functional elements in the longer term emergence of language. In other words, did the process of grammaticalization lead to the original split in the lexicon in the first place?” (Kirby 2013: 135).

For Heine and Kuteva (2007) and Bybee (2010) the answer is in the affirmative: they see no *a priori* reason why the processes underpinning language change should have been fundamentally different for the emergence of grammatical categories. Much of the modern debate on grammaticalization
Integrating Cognitive Linguistics and language evolution research centers around two definitions (cf. Kiparsky 2012). The first of these sees grammaticalization as a process through which “the parts of a constructional schema come to have stronger internal dependencies” (Haspelmath 2004). This definition encompasses the set of changes in morphosyntactic form: Here, we observe a unidirectional process towards increasingly tightly bonded units that show loss of autonomy (Lehmann 2004). An example of this type of grammaticalization would be the following change: postposition > clitic or suffix (Kiparsky 2012). The second definition sees grammaticalization as an expansion of semantic-pragmatic, syntactic, and collocational range (Himmelmann 2004). An example of this type of change is when an epistemic modal acquires a deontic meaning (Kiparsky 2012).

Taking a constructionist and usage-based perspective, Traugott and Trousdale (2013) show how constructional changes take place within a network model, with these changes being characterised by the creation of new constructional nodes as well as the reconfiguration of links between nodes. These networks not only consist of micro-constructions, they also are hierarchically organised with there being groupings of nodes (schemas and subschemas). In terms of Traugott and Trousdale’s diagnostics, grammatical constructionalization is associated with an expansion of productivity (expansion both of type frequency and token frequency) and an increase in schematicity (shift towards procedural function and changes in schemas), but a reduction in the transparency between the form and meaning mapping of a micro-construction (compositionality). Importantly, grammatical constructionalization is seen as an outcome of the changes described, and is not considered a process (see also Joseph 2001). As such, the two definitions of grammaticalization as either expansion or reduction are not necessarily orthogonal in Traugott and Trousdale’s account: here, grammatical constructionalization involves an expansion in the construction types and their range of use and the chunking and fixing of forms within the linguistic system. As they note, “[E]xpansion is the logical outcome of attrition resulting from repetition and chunking” (p. 147).

Given that constructions are influenced by these basic cognitive mechanisms of learning and use, we can see how repetition, chunking and expansion provide the raw material through which procedural functions can arise. The motivation for the evolution of these procedural functions, then, is the need to expand the expressivity of the constructicon, without detrimentally impacting our learnability. Such problems become inevitable once we begin using language across a larger range of cognitive, contextual and socio-cultural niches. These notions echo those of Bybee (2010: 203):
Note that in this view, the first language or languages are thought not to be the same as present day languages. They would have had lexical items but not grammatical items or constructions. Grammar developed gradually as language was used and as the capacities of humans or our ancestors increased to accommodate a large vocabulary, more abstract categories and many automated sequences.

One tentative hypothesis we might draw from these accounts is that procedural functions are the result of language adapting to its own internal dynamic to solve pressures from learnability and expressivity. If language is expanding into new socio-cultural contexts and niches, and our strategic goal is to reduce uncertainty about the intended meaning whilst maximising our expressive capacity, then a construction made up of purely contentful constructions is constrained by the burdens of memory and processing as well as the need to coordinate on a shared system. In short, humans cannot go on indefinitely creating new contentful constructions. Procedural constructions provide a simple solution to this problem by expanding the ways in which a linguistic system can package information (see *Information Structuring*: cf. Goldberg 2014): that is, through a small set of procedural constructions we can greatly enhance a language’s expressivity, without significantly impinging on learnability.

Through reusing a pre-existing inventory of contentful constructions, and applying procedural constructions as a way of modifying and signalling relations in a predictable way, we solve two problems simultaneously. First, the language is highly expressive, as pre-existing signals can be used to express new content. These *reuse strategies* are evident in any situation where we observe ambiguity and are advantageous in communication systems skewed towards hearer inference over speaker effort (Levinson 2000). In short, when the context is known and informative, it is a useful resource in decreasing uncertainty about the intended meaning (Piantadosi *et al.* 2012; Winters *et al.* 2014). Employing a reuse strategy allows a population of speakers to make use of contextual relevance, and our powerful inferential capacities, to expand the expressivity of the system without detrimentally impacting upon our capacity to coordinate a linguistic system across a community of speakers. Second, the *learnability problem* is solved, as the reuse of constructions allows for rate-limited growth in the construction. By limiting the number and types of new constructions, the burden on our memory and processing apparatus is significantly reduced.

Our account for the emergence of procedural constructions makes several predictions about the cognitive system and mechanisms underpinning
this division of labour. The first requirement is an inferential capacity that makes use of common ground and the communicative context to expand a pre-existing system of constructions (for recent accounts see Smith and Höfler 2014; Scott-Phillips 2014). Once this inferential capacity is in place, the process of cultural evolution frees up a communication system’s ability to explore new functional spaces where contentful constructions take on a more procedural role. What pushes these previously contentful constructions into these new procedural spaces are the well-documented pressures of learnability and expressivity acting on multiple timescales to expand the construction types and their range of use within the linguistic system.

The account we have briefly sketched here offers some tentative hypotheses regarding the division of labour in the construction. Future work should aim to explicitly test these hypotheses through a combination of analytical models, computational simulations and laboratory experiments (Irvine et al. 2013).

5. Conclusion

There are many convergences and similarities between CL and evolutionary linguistics and the two disciplines can profit from interdisciplinary integration. Language exists at three timescales that dynamically interact with one another – it is a Complex Adaptive System. Integrating this view with research from CL can give crucial insights into important problems in evolutionary linguistics.

Unifying cognitive and evolutionary linguistics provides a framework in which researchers can identify the cognitive machinery and processes underpinning learning and use, how linguistic systems change over time, and ultimately how they evolved in the first place. This allows us to frame what evolved by asking how language evolved given its biological and cognitive machinery. That is, we can examine how languages pattern synchronically, as well as how they change diachronically, and use this to infer what biological features are necessary.

As CL sees the complex adaptive system of language as well as its evolution as relying on general cognitive capacities and factors, it also actively seeks to integrate converging evidence from other disciplines in cognitive science (Evans and Green 2006). This feature of CL thus makes this paradigm highly amenable to interdisciplinary integration and presents another reason why a synergetic dialogue between CL and language evolution research seems worthwhile. As we have shown, adopting a usage-based
and constructionist perspective, CL can shed light on the relation of socio-cognitive and domain-general factors on the one hand, and the emergence, acquisition, and use of language on the other. Specifically, CL can help elucidating the cognitive principles implicated in all three complex adaptive systems that are involved in the complex phenomenon of language: ontogeny, glossogeny, and phylogeny.

A crucial question, then, concerns the relative roles of cultural evolution and biological evolution in accounting for the underlying structural hallmarks of language. In this paper, we have alluded to a possible explanation of the contentful-procedural divide in terms of a trade-off of cognitive pressures for learnability and expressivity. This perspective can thus help us to gain insight into the cognitive foundations as well as the processes that influence language learning and use, as well as the emergence of language more generally.

By showing how general cognitive capacities can interact with cultural evolutionary processes, we might be able to extrapolate from well-attested processes observed in historical language change to the evolution of language. CL thus can make a significant contribution to the highly interdisciplinary study of language evolution.

References


Integrating Cognitive Linguistics and language evolution research


The fourth V, as in evolution: How evolutionary linguistics can contribute to data science

Abstract. The paper explores the importance of closer interaction between data science and evolutionary linguistics, pointing to the potential benefits for both disciplines. In the context of big data, the microblogging social networking service – Twitter – can be treated as a source of empirical input for analyses in the field of language evolution. In an attempt to utilize this kind of disciplinary interplay, I propose a model, which constitutes an adaptation of the Iterated Learning framework, for investigating the glossogenetic evolution of sublanguages.

Keywords: Data science, evolutionary linguistics, natural language processing, Twitter, glossogeny, Iterated Learning framework

1. Introduction

This paper shows the importance of closer interaction between data science and evolutionary linguistics, pointing to the potential benefits for both disciplines. To substantiate the claim concerning these profits, a model is put forward for investigating the glossogenetic evolution of sublanguages emerging on social networking services (SNS), such as Twitter.

Users of social networking services generate unprecedented amounts of text, which constitutes one of the foundational factors of the big data revolution. For linguists this presents both new opportunities and challenges.
On the one hand, these data can be viewed as corpora that are especially valuable if one is to consider their size and actuality. On the other hand, challenges surface, as it is one of the roles of language experts to provide insight that will enable developing tools for efficient execution of natural language processing (NLP) tasks. The optimization of these tasks constitutes one of the principal demands in the emerging field of data science. Despite the relatively satisfactory state of syntactic processing tools, the semantic and pragmatic analyses of texts are still in their infancy (Cambria and White 2014). Similarly, tasks associated with language dynamics await scientific tackling in order to bring algorithms closer to natural language understanding.

2. The three V’s of Big Data

Recent developments in digital technology, including the virtually ubiquitous access to the Internet via mobile devices, coupled with the advent of social media platforms, lead to the generation of unparalleled amounts of data. Many intuitive interpretations of the buzzword big data overlook the significance of the phenomenon in question. The key aspect is the results that can be achieved through analyses conducted over these enormous sets of information, hence the demand for specialists able to work within the emerging academic discipline, i.e. data science. Its inherent interdisciplinary nature renders it rather difficult to be defined unequivocally (Provost and Fawcett 2013: 2–3). In a similar vein, the skills and expertise expected from a “data scientist” usually exceed competencies of any single person; the minimum requirement seems to be the ability to develop codes, operate data analytical tools and, at the same time, have a solid background in the relevant domain (e.g. medicine, marketing, etc.), which defines the type of data for analysis (Davenport and Patil 2012: 73–74). Since at present there are not many academic institutions that offer degrees in data science (Davenport and Patil 2012: 74), the only currently viable solution is forming interdisciplinary groups of experts with appropriate labor distribution. Indeed, big data analyses drive progress in many diverse domains of social activity, including: health care, science, politics, social engineering, economy, and logistics (Conte et al. 2014: 326). This demand opens up a niche also for linguists.

In its contemporary interdisciplinary context (Jackendoff 2002, 2007a, 2007b), linguistics plays a pivotal role in the numerous challenges that lie ahead of computational social science. After all, in 2007, a third of all digital data comprised of text (Hilbert 2014). Thus, mainly via the applied end of its disciplinary spectrum, namely natural language processing, the study
of language provides important base-knowledge for conducting analyses on big text data. In general, digitalized language can be considered one important type of structured data, *i.e.* data whose analysis is possible with the currently available tools, such as key word recognition or parsing. However, language data interpretability depends on the particular NLP task and so, for instance, sentiment analysis, relation extraction or textual entailment still constitute computational operations that yield low levels of accuracy, being associated with semantic and pragmatic processing (Cambria and White 2014: 55–56). Looking from the opposite perspective, great amounts of annotated, authentic texts have been the holy grail of linguists (at least, in the usage-based or empirically-driven approaches) for many years now. Thus, processing big data should be understood as having access to basically limitless amounts of information (provided, of course, that the processing yields correct metadata tagging), enabling empirical research that is superior to “rules and logic [that] miss frequency and language dynamics” (Bengfort 2013).

As any relatively new term entering the scientific discourse, big data has been causing some controversy in terms of definition. In a brief review of some of the most important definitions available, Chen *et al.* point out that explanations pertaining exclusively to size are insufficient since they always remain relative to the particular domain of research (consider the sizes of sets in fields such as: genetic sequencing, astronomy, or social science) (2014: 173). Hence the importance of focusing on the remaining two aspects: velocity and variety – following the commonly quoted Doug Laney’s “3Vs model” (see Table 1). The authors of the review propose the following synthesis: “In general, big data shall mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time” (Chen *et al.* 2014: 173), due to either their volume and/or variety.

Table 1. *The three aspects of big data recognized in Laney’s model*

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>scale and size of data sets, relative to a particular domain and to the computational power required to process them that are available to a particular processing entity</td>
</tr>
<tr>
<td>Velocity</td>
<td>pace at which data is produced and the speed required for its efficient processing</td>
</tr>
<tr>
<td>Variety</td>
<td>data type diversity in a given stream (text, video, audio, static image, etc.); also differences in data processability(structured, semi-structured, unstructured data)</td>
</tr>
</tbody>
</table>
3. The missing V

In spite of its great potential, data science is still at an early stage of its development as an independent discipline, and there are as many challenges to face as there are opportunities. In a recent review paper, Jagadish et al. (2014) identify some of the most significant problems for the developing field: heterogeneity, inconsistency and incompleteness, scale, timeliness, privacy and ownership, human cooperation and comprehension. Some of these issues, for instance heterogeneity, relate directly to problems of language processing if one is to consider features such as diversity of natural languages, genres, spelling conventions, and levels of linguistic analysis. Providing a comprehensive report on the state of the art in the NLP research, Cambria and White (2014) identify the tasks that lie ahead of big-data-driven language processing, that is interpreting texts beyond the purely syntactic level and focusing on the much in-demand semantic and pragmatic information in order to execute: sentiment analysis, emotion recognition, relation extraction, linguistic summarization, knowledge representation, word sense disambiguation, co-reference resolution, question answering and grammatical evolution. The successful management of these NLP tasks will render big language data structured across all levels of analysis, and thus, ultimately make it usable in the context of data science.

To gain a more comprehensive perspective on the types of problems occurring in the field of data science, let us turn to one of the most commonly quoted examples of big data implementations: Google Flu Trends (GFT). GFT is an algorithm developed to conduct spatiotemporal predictions of influenza pandemics based on analyses of Internet users’ queries typed into the Google search engine. The accuracy of GTF has been recently questioned by Lazer et al. (2014) in a paper that reports significant discrepancies between the algorithm’s predictions and factual data obtained from public health care institutions. The authors conclude that the progressing miscalculations in GFT prognoses are a direct result of Google engineers underestimating the dynamics that govern both users’ behavior and the technologies/platforms being used:

Twitter, Facebook, Google and the Internet more generally are constantly changing because of the actions of millions of engineers and consumers. Researchers need a better understanding of how these changes occur over time. Scientists need to replicate findings using these data sources across time and using other data sources to
ensure that they are observing robust patterns and not evanescent trends. [...] More generally, studying the evolution of socio-technical systems embedded in our societies is intrinsically important and worthy of study. (Lazer et al. 2014: 1205; bold fonts: MP)

Hence the importance of taking into consideration the fourth V, i.e. evolution, when conducting any kind of big data analyses. Language, being one of the systems mentioned above, requires investigation that would embrace its dynamic nature manifesting itself within a relevant time-frame. This is the task, whose accomplishment can be brought closer through the insights stemming from the Evolutionary Linguistics research.

4. Selecting appropriate data sources

One of the key factors affecting any given set is the source of data. Let us consider a question: which platform should one use for data crawling to achieve highest accuracy? After all, search engine queries, SNS posts or microblogging entries differ in terms of form as a direct result of users’ interactions with various functionalities provided by a particular platform.

This platform-output dependency can be illustrated with reference to the GFT discussion. In their response to the GFT critique mentioned above, Broniatowski et al. (2014) quote their own influenza surveillance studies that were significantly more successful in terms of prediction accuracy while relying on data from Twitter. The authors claim that the main advantage of Twitter data over Google’s is the lack of replicability constraint: the latter are proprietary, which restricts any research based on the query data only to scholars employed directly by the Mountain View giant, whereas the former are provided as part of the Twitter’s open access policy (Broniatowski et al. 2014), which stimulates unmonopolized scientific progress. In fact, the Twitter-based prognoses came satisfactorily close to the actual flu metrics compared against the figures obtained from public health care institutions – an effect stressing the importance of source choice in big data analyses.
5. Twitter characteristics

Twitter is a microblogging social network service that enables its users to post short text entries, called *tweets*, of maximum length not exceeding 140 characters, optionally including also image attachments and/or URLs. When considering the contemporary Social Networking Services with the highest populations, Twitter has a number of features that render it particularly interesting for researchers working in Social Network Analysis (SNA). Firstly, unlike, for instance, Facebook, Twitter does not impose reciprocity in user relations. In other words, user A can follow user B’s posts without user B being obliged to follow user A back. Secondly, contrary to the majority of social media, whose evolution is driven solely by their developers, with innovations being introduced in a top-down fashion, Twitter functionalities originate spontaneously as community-based implementations, hardcoded into the platform only once they surface as users’ adaptive behavior (Bruns and Burgess 2011: 2). Apart from the original length constraint, virtually all of the now-established Twitter functionalities emerged in a bottom-up manner. The driving force behind such innovations was interaction enhancement. This way, users of Twitter adopted a number of in-text markers that annotate tweets with specific discursive functions. Table 2 explains the canonical usage of these markers; example tweets, carrying a particular tag, are also provided.

The use of the above markers affects Twitter streams leading to the emergence of diverse information diffusion chains. Two main types of communicative systems are recognized on Twitter: hashtag-based topical discussions and community (follower-followee) conversations (Rossi and Magnani 2012: 563; Bruns and Burgess 2011: 6). Each of these communicative networks exhibit different properties.

---

1 An in-depth presentation and discussion of all Twitter functionalities lies beyond the scope of this paper. However, it has been already investigated and exposed in numerous sources; thus, I shall focus only on the features that are relevant for the current discussion. For detailed information on Twitter functionalities, I redirect the reader to the Twitter Glossary Section (http://goo.gl/daqhAp) as well as Kwak et al. 2010 and Liu et al. 2014: perhaps the two most comprehensive studies on Twitter so far.
The fourth V, as in evolution: How evolutionary linguistics...

Table 2. Markers used in tweets

<table>
<thead>
<tr>
<th>Name</th>
<th>Marker</th>
<th>Function</th>
<th>Example tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hashtag</td>
<td>#</td>
<td>marks a tweet as belonging to a discussion on a particular topic</td>
<td>The moment #EuroMaidan started. Late afternoon, November 21st 2013. I got there a bit later, it was dark already...</td>
</tr>
<tr>
<td>Reply</td>
<td>@user_name (initial position)</td>
<td>a public message addressed to a particular user</td>
<td>@Fluid_trm I am looking into it now, i haven't found him yet, but if i do find him you will be the first to know :-)</td>
</tr>
<tr>
<td>Mention</td>
<td>@user_name (non-initial position)</td>
<td>mentioning a user; not necessarily for communicative reasons</td>
<td>LIVE NOW: @rustyrockets and @ABFalecbaldwin talk about their kiss on @KeiserReport WATCH HERE: <a href="http://rt.com/on-air/">http://rt.com/on-air/</a></td>
</tr>
<tr>
<td>Retweet</td>
<td>RT</td>
<td>quotation; indicates that the content of a tweet is a direct replication of another tweet (adding one's own comment is optional or depends on the available space left)</td>
<td>“RT @mlcalderone: 48 journalists attacked this month in Ukraine demonstrations; over 100 this year: <a href="http://bit.ly/19sw6u0%E2%80%9D">http://bit.ly/19sw6u0”</a></td>
</tr>
</tbody>
</table>

Community networks, where users are linked via the “follow relations”, constitute a prototypical mode of conversation available on Twitter. This means that one user can directly address another user or a number of users. If the addressees reply, a conversation unfolds, exhibiting features expected to be found in most instances of Computer Mediated Communication (CMC). Honeycutt and Herring (2009) corroborate this view: their study confirms that, despite its design function as a microblogging service, Twitter has evolved as a platform capable of hosting conversation. One of the most comprehensive analyses of Twitter so far, conducted by Liu et al. (2014) (data sample: 37 billion tweets gathered over a period of 7 years) further proves the conversational capabilities of Twitter. The authors report that the @ marker, used for addressing, is present in about 50% of all contemporary tweets. Interestingly, although Twitter can potentially host conversations with as many as ten participants (Honeycutt and Herring 2009) or more, Macskassy shows that, within his data sample, 92% of all conversational interactions were between two users only and that there is a tendency for responsiveness to decrease as the number of conversation participants grows (2012: 231). Finally, in terms of graph generation, community networks also
exhibit diversity between themselves depending on the particular use of the @ marker: mention vs. reply (Cogan et al. 2012).

Topical networks, on the other hand, conglomerate around a particular hashtag instead of in between user interaction. Hashtags designate topics, which usually appear in relation to events in the external reality, whether in anticipatory, ad hoc or post hoc manner (Bruns and Burgess 2011: 7). Crucially, unlike in the case of @-based conversations, the users involved in hashtag discussions do not have to maintain follower-followee relations. Bruns and Burgess note that such topical networks do not form communities:

The term ‘community’, in our present context, would imply that hashtag participants share specific interests, are aware of, and are deliberately engaging with one another, which may not always be the case; indeed, at their simplest, hashtags are merely a search-based mechanism for collating all tweets sharing a specific textual attribute, without any implication that individual messages are responding to one another. (Bruns and Burgess 2011:5)

The two types of communicative networks on Twitter also differ in term of size. Depending on a particular hashtag, the topical discussions can generate up to millions of tweets, involving a comparable amount of users (Kwak et al. 2010: 597). By contrast, Honeycutt and Herring (2009) showed that, within their sample, the number of tweet exchanges per conversation (user-to-user interactions) ranged 2–30, for participant number ranging 2–10. Of course, conversations between users can also carry hashtags, which mark personal-level debates within large-scale topical discussions (Bruns and Burgess 2011: 4).

Apart from their prototypical use as topical markers, hashtags are also used for emphasis (highlighting keywords), emotive expression (as Internet memes, e.g. #facepalm), or backchannelling (Bruns and Burgess 2011: 3–5). The latter refers to a situation when Twitter serves as an official communicative platform during a particular event. For such occasions, a special hashtag is coined by the event organizers and then propagated among the participants, or anyone interested in the happening, to coordinate and gather all event-related communication. This solution is used mostly for broadcast events, where the viewers are encouraged to take part in a discussion (e.g. #xfactor) or during conferences, as means of information exchange (see Letierce et al. 2010 and Weller 2011).
6. Related work – Twitter and Evolution of Language

While the linguistic evolution of Twitter content *per se* has not been investigated so far, related dynamic phenomena occurring on the platform did receive academic attention. These analyses encompass patterns of information diffusion (Kwak *et al.* 2010) or the evolution of the service itself, *i.e.* available functionalities, users’ behavior, network structures (Liu *et al.* 2014).

To my knowledge, the paper by Cunha *et al.* (2011) constitutes the only attempt to approach questions of language evolution in relation to Twitter content. Although not explicitly referring to a particular framework or model in evolutionary linguistics, the researchers investigate the adaptive behavior of hashtags. When a new topic appears, Twitter users are free to coin a suitable hash marker to tag their tweet as being a contribution to a discussion on the particular event or theme. This unconstrained production of tags usually leads to a situation where, initially, a number of competing hashtags are in use simultaneously. Later, through competitive mechanisms, the most successful markers prevail and conglomerate most of the topic-related content. The authors report that there is a set of patterns that govern hashtag evolution. One such regularity is the correlation between the number of characters used for a given hashtag and its success – the tendency being that longer hashtags exhibit lower popularity/prevalence. Another insight is that Twitter hashtag evolution follows the “rich-get-richer” pattern: “in some stems, the popularity of the most common items tends to increase faster than the popularity of the less common ones. It generates a further spread of the forms that achieve a certain prestige.” (Cunha *et al.* 2011: 61). Although indicating an interesting direction in the evolutionary study of Twitter content, for now, the work is limited to the # markers only; also it does not constitute a strictly linguistic (morpho-syntactic) investigation of the problem, focusing solely on length and orthography.

Also worth mentioning is that Zappavigna (2011b) presents modes of semantic data visualizations for Twitter input, in which she refers to logogenesis, ontogenesis and phylogenesis as part of the adopted terminology. Although the immediate connotations are those with the field of evolutionary linguistics (see Hurford 1990), here, the terminological overlap is only superficial: the author represents the school of Systemic

---

2 The only exceptions are situations when there is an official hashtag (see backchannel-nelling, section 5).
Functional Linguistics and uses these concepts as defined by Halliday (see Halliday 1993 in Zappavigna 2011b).

7. The Iterated learning framework

The Iterated Learning framework (IL), largely developed by Simon Kirby, aims at investigating language evolution at the level of glossogeny, *i.e.* the evolutionary changes in the general structure of the communicative code; changes of cultural nature, where the replicators constitute arbitrarily defined linguistic units (Wacewicz 2013: 1). The term *Iterated Learning* refers to a process that is argued to govern, among other phenomena, the cultural transmission of human language: this, in turn, drives the changes occurring in the linguistic code itself (Kirby et al. 2014: 108–109). In the process “an individual acquires a behavior by observing a similar behavior in another individual who acquired it in the same way” (Kirby, Cornish, and Smith, 2008: 10681). Through computational and mathematical modelling as well as psychological experiments that simulate the process of cultural transmission *via* iterated learning, Kirby and his colleagues investigate the evolution of simplified language-like codes, or *evolects* (see Jasiński in preparation). By means of generalization over the data obtained, the IL researchers draw conclusions regarding the glossogenetic evolution of human language. In general, the framework constitutes a valuable tool for acquiring data for the study of the adaptive dynamics of language, thus, confronting one of the main challenges in evolutionary linguistics: the scarcity of empirical evidence.

Over a decade of IL research points to the fact that it is the linguistic codes themselves, rather than their users, that adapt towards compositionality and structure. Across all experimental settings (that is, in computer and mathematical modelling, but also with human agents) a similar pattern prevails: the input code, unstructured and comprising of randomized form-meaning mappings, becomes increasingly regular and structured after a specific number of iterations (in other words, transmissions or generations) (Kirby et al. 2014).

The proposed explanation for the adaptive mechanisms observable in IL research is that, in this way, the codes react to the narrow bottlenecks they encounter on their evolutionary path (Kirby et al. 2014: 108–109).

---

3 Experimental artificial languages or mathematical constructs used for modelling processes of structure emergence and evolution. (Jasiński: personal communication)
In a nutshell, bottlenecks constitute various constraints that mediate inter-generation transmissions (Kirby et al. 2004; Kirby and Hurford 2002), rendering the post-transmission state of a given code different from its pre-transmission state. Naturally, the process repeats itself with every consequent transmission. In the context of natural language, these constraints correspond to aspects such as the “poverty of stimulus” problem in learning or the human limited working memory capacity in processing.

8. Evolutionary perspective in microblogging discussions

If the findings discussed above are correct, it would be interesting to investigate analogous phenomena in language data of higher resolution. Therefore, I propose a model for studying glossogeny at the level of topic-specific sublanguages emerging on Twitter. Although the model refers to the Iterated Learning framework in a number of theoretical assumptions, it has different focus, methodology and aims.

While the IL framework focuses on investigating glossogenetic phenomena in evolects, the proposed model approaches similar questions, yet, in the context of authentic language data. Twitter topical discussions provide a convenient source of naturally occurring data, wherein topic-specific sublanguages\(^4\) can be extracted via a particular hashtag. As noted above, such topical discussions constitute pools of time-bound utterances relating to a particular theme, yet, by default, not being a result of conversations within communities. Crucially, rather than being a variety of conversational analysis in the CMC context, this approach allows for the investigation of sublanguages, evolving over time as micro-systems, in abstraction, to a certain degree, from the speakers involved. Also, the choice of topical discussions for analysis, significantly increases the sizes of data sets. However, the shift of focus from evolects to natural language is a profound one: we move from dealing with emerging artificial constructs to sublanguages originating within already fully-developed, highly complex linguistic systems.

As in the IL framework, the proposed model also assumes a number of bottlenecks on the process of code transmission. The transmission

---

\(^4\) Though initially I opted to use the term “sociolect” as semantically closest to what is meant here, its definition implies the existence of a particular community using such a variety of language, which could be misleading in the context of topical discussions on Twitter, see section 5.
bottleneck(s) should be understood as a set of constraints that limit the number of possible features which would otherwise freely prevail throughout the evolution of a given code. The expected bottlenecks can be classified as relating to one of the three relevant aspects of Twitter discourse: content, platform and agents. Figure 1 presents a portion of all possible pressures and biases that affect an emerging sublanguage on Twitter.

One bottleneck relating to content is the topic-hashtag association: any given topic is normally discussed under a number of alternative hashtags within Twitter (#TOPIC and #TOPIC_ALT), or without an overt marker, or even as part of private messaging between Twitter users, who can then contribute to the public stream again. Moreover, the same topic is usually discussed in parallel across different on-line and off-line media (Facebook, forums, blogs), whose users contribute and process content between the different services. At the same time, the course of the events that relate to or, in fact, constitute the topic develop in reality. This dispersion of topic-related information affects a given isolated hashtag stream: the particular channels generate a pool of possible content to enter the stream. (see Figure 1)

Figure 1. The complexity of a prototypical topical network evolving on Twitter. All arrows indicate possible channels for content stream.
Then, agents (i.e. users and/or contributors in general) constitute a bridge for these separate channels to feed each other – here the second bottleneck becomes visible. Agents can either process the topical information and then post to the stream of interest, or replicate fragments as well as complete texts within (RT in Figure 1) or across platforms. Thus, we see that content-related bottlenecks interact with agents’ biases.

One of such biases is the status of a user. Based on their SNA study, Wu et al. (2011) categorize Twitter account holders into “elite” and “ordinary”: the former exhibit extremely high followee numbers (up to millions) – the latter have relatively lower audience counts. A user’s status determines the visibility of her posts (Wu et al. 2011: 706), which means that the elite users generate content accessible to a significantly larger audience. Thus, it is this content that will have greater adaptive power, i.e. higher probability for replication (via RT), or simply more significant impact on viewers.

Another bias also relates to the categorization of users, yet under a different criterion. The development of Twitter as well as web technologies in general spawned the so-called bots, that is, algorithms programmed to post content automatically across different Internet services. Imagine the unattainable human workload necessary for any of the major news media to post a link to their story across multiple social networks. Therefore, the redundancy of such tasks is handled by bots, which also begin to manage some portion of Twitter accounts. Importantly, the output generated by non-human agents has been reported to exhibit differences in content attraction (Edwards et al. 2014).

Perhaps the most obvious bias rests on the assertion that agents, if human, have linguistic systems already entrenched and that this cognitive makeup will influence any evolving sublanguage (Pokornowski and Rogalska 2014). A number of studies conducted indicate that users’ behavior differs depending on the language they use in tweets, or their proficiency in it (Hong et al. 2011). Similarly, within one language, the choice of orthographic and/or morphological strategies varies depending on a particular variety or dialect (Gouws et al. 2011). Extra-linguistic factors, such as users’ geopolitical context or cultural background can also constitute a major bias. For instance, Chen et al. (2013) describe the linguistic strategies used for avoiding censorship by users of the Chinese microblogging platform, Weibo.

Finally, platform-related bottlenecks will encompass all the pressures that stem from Twitter’s design as a social networking service. For instance, the microblogging length limit of 140 characters per tweet already makes any Twitter corpus a collection of texts of a very specific genre, which, in turn, explains the stylistic variation found in Twitter stream (see Hu et al.
2013). Other pressures relate to specific functionalities, such as the retweet option, enabling either exact (embedded retweet functionality) or partial (via the RT marker) replication of particular content. One important factor affecting the Twitter stream is of course the evolution of the platform itself: new implementations, whether user-driven or provided by the developers, immediately shape information diffusion patterns within the entire network (Liu et al. 2014).

The bottlenecks and biases mentioned above need to be accounted for when investigating the evolution of sublanguages emerging around particular topics on Twitter. If carefully controlled, this matrix of pressures and constraints will enable the discovery of crucial adaptive mechanisms that shape the unfolding linguistic code under investigation.

9. Conclusion

The model proposed in this paper constitutes a methodological approach to the study of glossogenetic evolution of sublanguages on the microblogging platform Twitter, or possibly through further modification, on other social networking services. Although related to the Iterated Learning framework, this model has the advantage over the IL approach in its reliance on large masses of authentic language data. For the field of evolutionary linguistics this provides a possibility to arrive at more accurate conclusions concerning the mechanisms that govern language evolution at the glossogenetic level. In turn, a better understanding of the inner dynamics of Twitter content can aid NLP research in the development of processing tools that will be capable of embracing any natural changes occurring within a given stream of data, hence, increasing the accuracy of big data analyses that can have significant implications for our societies.

References:


Liu, Y., C. Kliman-Silver, and A. Mislove. (2014). The Tweets They are a-Changin’: Evolution of Twitter Users and Behavior. In *ICWSM*


Ines Adornetti
Department of Philosophy, Communication and Visual Arts
Roma Tre University
&
Department of Human Sciences
University of L’Aquila
ines.adornetti@uniroma3.it

Francesco Ferretti
Department of Philosophy, Communication and Visual Arts
Roma Tre University
francesco.ferretti@uniroma3.it

The pragmatic foundations of communication:
An action-oriented model of the origin of language

Abstract. In this paper we propose an action-oriented model of language origins that is compatible with the embodied perspective of mind. Specifically, we maintain that a crucial issue for the investigation of the origin of human communication is how language is grounded in the context. In support of such a proposal, we maintain that: a) the grounding of language is tied to a specific aspect of the embodied mind, namely the aspect that emphasizes the embedded character of cognition; b) the emphasis posed on this particular character fits well with a model of language origins founded on gesture. The connection of embedded cognition and the gestural model allows us to propose an explanation of language origin capable of overcoming the problems in the classical symbolic framework of cognitive sciences.

Keywords: embodied cognition, gestural origin of language, grounding, mirror neurons, pragmatics
1. Introduction

In this paper we propose a naturalistic model of language based on an action-oriented paradigm. By suggesting such a model, we intend to overcome the conceptual contradictions that the classical symbolic framework, of which Chomsky’s Universal Grammar (hereafter UG) is a paradigmatic example, have with the Darwinian perspective. Contrary to the classical symbolic point of view, we maintain that a naturalistic model of language has to be evolutionary plausible. In order to justify the evolutionary plausibility of our proposal of a naturalistic model of language, we highlight the strong connection between the embedded perspective elaborated within the general framework of embodied cognition and the gestural theory of language origins. Specifically, our argument is that the origin of human verbal skills must be interpreted in reference to the ability of language to be grounded (anchored) in context and that the capacities that ensure the grounding of language in context are associated with a specific aspect of the motor foundation of human communication tied to the embeddedness of the organism in the surrounding environment.

2. Beyond Universal Grammar

A naturalistic perspective of language is a theoretical approach that considers human verbal skills (as well as any other ability) as a specific feature of an animal as other animals. To adopt a point of view of this type means denying that human beings are exceptional in nature: all abilities that characterize individuals of our species (even those that make us unique among other animals) must be interpreted in reference to the “specificity” that characterizes human beings, never to their supposed status of “specialty” in nature (Ferretti 2007). As it is easy to realize, an approach of this kind is a tribute to the tenet of Darwin (1871), according to which the differences, however great, between the most intelligent animal and the most foolish human being are always a matter of quantity and never of quality. From this point of view,

---

This paper is the outcome of a collaborative effort between the two authors. For the specific concerns of the Italian Academy, we specify that I. Adornetti wrote sections 4 and 5 and F. Ferretti wrote sections 2 and 3 for the final draft. Both authors wrote the Introduction and the Conclusion.
a naturalistic model of language is a model compatible with the Darwinian perspective from the outset.

Following these considerations, we propose that to be properly naturalistic, a model of language must be compatible with the general framework of the theory of evolution. An assumption of this kind, which might appear to be simply a common sense intuition for those who analyze language from a naturalistic perspective, is a source of many disputes in contemporary debate. Chomsky – and with him the group of the neo-Cartesian proponents of the standard position of cognitive science (e.g., Fodor 2001, 2008) – while showing great attention to the issue of cognitive plausibility, has always revealed a strong skepticism regarding the evolutionary explanations of language and mind (e.g., Bolhuis et al. 2014; Chomsky 2005: 58–59; Hauser et al. 2014).

The Chomskyan model of language is still the dominant paradigm within cognitive sciences. Beyond its cognitive plausibility – which is, in any event an open question (see Evans and Levinson 2009; Evans 2014; Tomasello 2009) – UG does not acquire evolutionary plausibility despite the attempts made by some authors to “darwinize” it (Pinker and Bloom 1990). Chomsky’s skepticism regarding language as a biological adaptation depends on UG’s incompatibility with natural selection. In spite of the broad debate created by this position (e.g., Botha 2003; Calvin and Bickerton 2000; Christiansen and Chater 2008; Corballis 2011, 2013), Chomsky continues to state that “the human language faculty emerged suddenly in evolutionary time and has not evolved since” (Berwick et al. 2013: 89, our emphasis). Specifically, he maintains that because language has no external reference (this is a very important point for our purpose to which we will return later), it cannot have evolved through natural selection and, therefore, must have emerged in a single step, perhaps as a result of a fortuitous mutation (Chomsky 2010: 59). As Corballis (2013) highlighted, this position smacks of the miraculous because “from an evolutionary point of view the notion that a faculty as complex as language could have emerged in a single step is deeply implausible” (Corballis 2013: 35). These considerations (strongly inspired from dualistic conceptions) are fully consistent with the Cartesian intent to demonstrate mental superiority and human uniqueness, but constitute a serious obstacle to any real attempt of naturalization of human language (for a discussion, see Ferretti and Adornetti 2014). But that’s not all.

Beyond this general problem, UG remains incompatible with the theory of evolution because of two additional specific difficulties. The first critical point is that UG requires a common code (the languages used by communities of speakers) to function appropriately. In fact, the device at the basis
of language operates only assuming the right environmental input: without a common code, UG is like a system for viewing a room without light because it is a device innately specialized for language. However, the existence of a code strongly depends on the existence of UG: in fact, in Chomsky’s opinion, natural languages are just superficial products of UG functioning. If we exclude the reference to the preformism hypothesis, the relationship between linguistic code and UG is highly problematic. On the one hand, analyzing the origin of language assuming the existence of a linguistic code (a code is exactly what it is missing in the early stages of language) is not possible; on the other hand, assuming that brains are predisposed to language before humans are able to exploit the symbolic code used by the community of speakers in which they live makes little sense. What follows is that, from an evolutionary point of view, UG rests on a vicious circle (Ferretti 2013). The solution to this problem, as we see in the next section, is offered by the adherence to the model of communication based on clues proposed by Sperber and Wilson (1986/1995).

The critical point at the basis of the incompatibility of Chomsky’s model of language with the theory of evolution is the notion that UG is a device inside the mind that is completely detached from the surrounding environment (cf. Ferretti and Adornetti 2014). Even though Chomsky has radically changed his conception of UG over time, from UG’s first formulation in the fifties to the recent Minimalist Program (Chomsky 1995), a consistent element in his thought is that language is the device that makes possible the combination of symbols whose functioning is completely independent of the relationship they establish with the reality they represent: according to the generative perspective, what is important is how symbols combine with one another, not how they are related to the external world about which they talk.

Chomsky represents the more orthodox tradition within cognitive science: the tradition in which the mind is considered in reference to the metaphor of a computer and which conceives thought as a form of mental sentences the main character of which is the respect of the principle of formality (Fodor 1975; Fodor and Pylyshyn 1988; Pylyshyn 1985). Indeed, the notion that language competence is a device that analyzes the shape of symbols regardless of their content and the relationship between the uttered expression and its context is part of a broader conception of how to analyze the study of the mind in classical cognitive science (Fodor 1975). To be properly scientific, the notion that analysis of the mind should be driven by the principle of formality and by methodological solipsism (Fodor 1980) continues to be very strong in some interpretative models based on
the classical perspective, despite the fact that the computer metaphor that inspired them is now in sharp decline. In our opinion, proposing a model of language (and mind) completely detached from reality is equivalent to talking about the nature of an organism without referring to its relationship with its external environment. The disembodied and detached nature of language that Chomsky introduced makes it a useful tool for disembodied angels, but hardly for human beings in the flesh. To overcome this problem, we move toward an embodied perspective, and specifically to those interpretative models that within the embodied approach consider the embedded (or situated) nature of cognition to be an essential and fundamental characteristic.

3. Against the code model of communication

The notion of language as a disembodied tool is tied to a specific notion of communication: the code model (Shannon and Weaver 1949). According to the code model, the speaker encodes thought (the message) in a succession of sounds that the listener decodes to be able to share the thought (the message) that the speaker intended to communicate. Adherence to the code model, that Fodor (1975: 106) considered “not just natural but inevitable”, means adherence to a form of parallelism between thought and language. Fodor’s notion is that language can express thought because language and thought share the same logical form. From this point of view, the device that processes linguistic information must be capable of processing the logical form of the utterances: the logical form of statements is a necessary and sufficient condition for language comprehension. Adherence to this position is clearly at odds with the pragmatic theories of communication that focus on context and the speaker’s intention. Indeed, in the classical symbolic framework everything necessary for the comprehension of what the speaker said (in the statement actually uttered) is present—the information content is entirely encoded in the utterance (Fodor 1983). For this reason, the code model works perfectly with the theory of literal meaning and is closely linked to the logical form.

How well does a language model of this type actually work? To communicate efficiently, is it sufficient to correctly combine symbols into sentences that reflect the structure and constituents of thought? A good starting point to answer this question is an interesting reflection made by Uta Frith (1989: 115) concerning the communication skills of Ruth, an autistic girl “stuck on the literal meaning.”
Faithful conveying of information is not a trivial accomplishment. It calls for accurate encoding and decoding of speech at input and output stages. Ruth does this. Echolalic children do it too. Nevertheless, in everyday communication one rarely expects that a listener will have to receive and then transmit a bare message as an exact copy. On the contrary, on expects listeners to know that messages are not bare, but usually contain something more. What really matters in everyday communication is the point of the message rather than the message itself. In other words, as listeners, we need to know why the speaker conveys this thought (rather than another), and as speakers we need to be sure we are understood in the way we want to be understood. We have elaborate verbal and nonverbal signals for getting across these intentions.

What emerges in this passage is that the girl is not able to communicate effectively, even if she respects the requirements of the code model. Specifically, Ruth’s problem is not understanding what a speaker says, but understanding why the speaker is saying what she says. Autism is frequently studied in cognitive pragmatics (e.g., Happé 1993; Pexman et al. 2011) as evidence in favor of the role of the speaker’s intentions in the processes of language production and comprehension, assuming that a speaker’s thought is not explicitly coded in linguistic expressions (Grice 1968). Frith’s quote encourages us to think that, in addition to the assessment of the role of intention, the evaluation of why the speaker is saying what she says also involves the evaluation of the consonance of what the speaker says with respect to the contextual situation. From a perspective on communication in which what really matters is why someone is saying something and not what someone says, the expressive code assumes a role of secondary importance. A view that fits very well with this perspective is the model of communication based on clues—Relevance Theory—proposed by Sperber and Wilson (1986/1995). According to the authors, in communication the speaker simply offers the listener evidence (a clue) of what she intends to communicate (Sperber and Wilson 1986, 2002; Origgi and Sperber 2000).

Inferential communication is a matter of reconstructing the communicator’s informative intention on the basis of the evidence she provides by her utterance. Successful communication does not depend, then, on the communicator and addressee having exactly the same representation of the utterance, but on having the utterance, however represented, seen as evidence for the same intended conclusion (Origgi and Sperber 2000: 160).
For the purpose of our argument, what is important to note is that the question of why the speaker is saying what she says is closely related to the question of how language is grounded in situation. Indeed, showing that language production and comprehension rely on the evaluation of why a speaker says what she says means referring to one of the cornerstones of a pragmatic perspective on language: evaluation of how expressions are linked to their context. Such a consideration has profound consequences for the origins of language. In our opinion, the evaluation of why the speaker is saying what she says precedes (logically and temporally) the question of what she is saying. Indeed, in the early stages of human communication (i.e., in the absence of a code through which to express the content to be said), what regulates the processes of production and comprehension is the analysis of why someone is saying something at a given time. As Gärdenfors (2004: 244, our emphasis) claimed “when communication first appears, it is the communicative act in itself and the context in which it occurs that are most important, not the expressive form of the act” (see also Gärdenfors 2014).

The assumption underlying our argument is that anchoring language to context represents a specific case of the more general grounding of the organism to the environment. Starting from this assumption, and in order to explain the question of language origins, in what follows we propose a two-step argument. The first step shows that the analysis of the grounding problem has to be considered in reference to the embedded (or situated) model of the embodied theory of mind (cf. Cantwell Smith 1999). The second step shows that the theories of language elaborated within the embodied perspective are in accord with the gestural models of language origins. As a result of this two-step argument, we maintain that the embodied model represents the conceptual bridge of the grounding and the origins problems, without which no naturalistic perspective of language is possible.

4. Embedded cognition and action-based language

The conceptual foundation of our perspective can be tracked to the ecologic approach elaborated by Gibson (1979). Criticizing the idealistic conception of vision, Gibson proposed that perception is strongly connected to the movement of the organism in the environment and that, as a consequence, perception is in itself a form of acting. Gibson’s approach represents one of the main theoretical points of reference for the perspectives of embodied cognition (EC) (e.g., Barsalou 2008; Engel et al. 2013; Lakoff and Johnson
The main assumption of EC is that cognition is not a representation of the world but is strongly tied to action in the world (Clark 1997; Varela et al. 1991). Indeed, as Engel and colleagues (2013: 206) maintain, “cognition is fundamentally action-bound, subserving the planning, selection, anticipation, and performance of actions. Thus, cognition and action are not only closely interrelated – cognition seems fundamentally grounded in action”. From such a perspective, for example, knowing what an object is does not automatically imply having an internal abstract and amodal representation of the object but having sensorimotor skills and possible actions to use the object.

We do not intend to discuss the validity of EC compared to classical computational models and the various formulations of embodied theories (see Wilson 2002). For our argument, what is important to note is that the central assumption of EC is the rejection of the symbolic framework, according to which the cognitive processing requires a small set of rules that operate on a vast set of arbitrary, amodal and abstract symbols: the mind is conceived as a computational system of abstract symbols detached from the reality they represent (cf. Fodor 1975; Pylyshyn 1985; Turing 1950). As we noted by discussing Chomsky’s and Fodor’s positions, within the computational framework the relevant aspects of linguistic (and in general mental) processes concern how symbols combine with one another, not how they relate to external reality. As Meteyard and colleagues (2012: 789) stressed:

> the thorny problem of how symbolic representations refer to things in the world was explicitly recognized (Fodor 1987; Pylyshyn 1985; Newell 1980) but never explained within the symbolic framework. Determining the organization of cognitive processes was more important than establishing its content.

Putting aside the question of whether such criticism is valid for all the authors who Meteyard and colleagues cite, it must be stressed that, contrary to the symbolic paradigm supporters’, the authors moving within the theoretical framework of EC consider the symbol grounding problem as the central issue to account for the nature of human language and cognition. Originally formulated by Harnad (1990), the symbol grounding problem is posed to answer to the following question: how can the meaning of the meaningless symbol tokens (as those postulated by the symbolic models of the mind), manipulated solely on the basis of their arbitrary shapes, be grounded in anything but other meaningless symbols?

Regarding this issue there are two considerations. The first is that even though in Harnad’s original formulation, and in the debate that followed,
the symbol grounding problem posed a typical semantic problem, in this paper we propose a cognitive-pragmatic notion of grounding. The issue of using language in a manner contextually appropriate, namely in a way consonant to the situation, in effect, is all we need for a model of functioning (see section 3) and origin of language (see section 5) elaborated within EC. The second consideration is that, in line with the importance of the problem of the grounding of language to context as a specific case of the relationship between organism and environment, our adhesion to EC is in the first place an adhesion to the action-oriented models which exalt the embedded nature of cognition (Shapiro 2010). Indeed, such models, as Pouw and colleagues (2014) emphasize, could be considered “more embodied” because of the fact that they conceive cognition “as being on-line, that is, being tightly coupled with, embedded in, if not extended over, the body and the environment” (Pouw et al. 2014: 1). Indeed, the aspect more relevant to the embedded perspective is situatedness: the organism’s immediate environment has a central role in its behavior; such environment is not only a rich source of constraints and opportunities for the organism, but also a context that gives meaning to its actions (Beer 2014). So much for the grounding problem. Now we have to take into account the second point of our argument: the ability of the action-oriented models to explain the origin of language with reference to gesture.

Obviously, the embedded perspectives (insofar as they are more embodied), first and foremost, are action-oriented models of cognition. It is exactly the attention to action that illuminates the link existing between such embodied-embedded perspectives and the gestural theories. The starting point is the link between action and language. The problem of creating contextually-appropriate behaviour is not only a linguistic issue, but it is also an issue endemic to action systems (Gallese and Lakoff 2005; Glenberg and Gallese 2012). The motor system has solved the problem of producing contextually-appropriate behaviour by being functionally organized in terms of goal-directed motor acts, and not in terms of movement (cf. next section) (Rizzolatti et al. 2000). According to Glenberg and Gallese (2012: 911) “the brain takes advantage of the solution of one difficult problem, namely contextually-appropriate action, to solve another difficult problem, namely contextually-appropriate language”. This bond between aspects of language processing and the organization and activity of the human motor cortex has been demonstrated by numerous neuroscientific and behavioral studies (for a review, see Meteyard et al. 2012). For example, Pulvermüller (2005) found that when subjects simply read the word signifying an action, the motor system activates and represents its meaning: verbs for head, arm, and leg actions produce head, arm, and leg simulation in the respective areas of the motor
system. Glenberg and Kaschak (2002) found that participants more quickly responded to sentences if the response was compatible with the direction of the action implied in the sentence (sentence: “open the drawer”; action: hand moves toward the body). Similar results also came from patients with brain injury. For example, it has been shown that subjects with motor neuron disease or lesions in the left inferior frontal cortex have deficits in action-verb comprehension and in understanding pictures depicting actions (Bak et al. 2001, 2006). The connections between action systems and language processing characterize not only the processing of single words or single sentences, but also the processing of discourse and narratives (e.g., Chow et al. 2014; Kurby and Zacks 2013). The comprehension of stories might lead to the formation of modality-specific grounded representations: readers activated sensorimotor regions relevant to the perceptual information described in the text (i.e., secondary somatosensory and premotor cortex were associated with the reading of clauses that imply motor information [Kurby and Zacks 2013]). Studies of this type highlight that “our understanding of linguistic expressions is not solely an epistemic attitude; it is first and foremost a pragmatic attitude toward action” (Glenberg and Gallese 2012: 96, our emphasis).

5. Action-based communication and gestural origin of language

In the previous section we considered a generic relationship between language and action. Now it is time to deal with evolutionary issues showing the relationship between action-oriented models of cognition and gestural models of language origin. Indeed, the recognition that the motor system has a crucial role (beyond, of course, that tied to the motor control) in higher cognitive function and, specifically, in language comprehension and production, has also provided new views on the involvement of motor system in language evolution, supporting the hypothesis that human language first originated in a gestural-based system of communication (for a review, see Fogassi and Ferrari 2012).

Particularly relevant in this scenario is the discovery of mirror neurons, a specific class of sensorimotor neurons found for the first time in the F5 area of the premotor cortex of macaque’s brain (di Pellegrino et al. 1992; Gallese et al. 1996), and then in area PFG in the rostral part of the inferior parietal lobule (e.g. Gallese et al. 2002). The presence of a mirror system has also been established in the human brain (Grafton et al., 1996; Mukamel et al., 2010). These neurons are defined as mirror because they allow a kind
The pragmatic foundations of communication...

of mirroring between perception and action. Specifically, they discharge when a monkey performs an intentional, goal-oriented act with the hands or the mouth (such as attempting to grasp, bite and tear an object) and when it observes another individual (human or monkey) accomplish a similar intentional act. This activity of the mirror neurons is unlike the so-called canonical neurons, which respond only to the presentation of the object. What is important to stress is that according to this perspective, the motor system is not involved in movement, but in action: “Unlike movement, action is defined by a goal and by expectancy. Movements are the final outcome of action and are programmed and controlled as such only when action is set” (Rizzolatti et al. 2000). In this sense, through its links between action and perception, the mirror system provides a mapping of external reality onto our own internal representations (Aziz-Zadeh and Ivry 2009).

The functional role of mirror neurons is relevant to the origin of language. Several authors noted that the primary function of mirror neurons is related to an implicit, pragmatic, and non-reflective understanding of manual actions (Rizzolatti and Sinigaglia 2008). The idea is that when an individual observes an action performed by another, there is an activation of the neurons that represent (through simulation) that action in the premotor cortex. This (motor) stimulation of the observed action induces the recovery of the internal motor circuit that represents that action. The consequence of this recovery is that the observer acquires the knowledge of the objective associated with the action. The mirror system, in this way, transforms visual information into knowledge (Rizzolatti et al. 2001). This process may underlie some aspects of communication. Indeed, a communicative gesture made by an individual (the sender) retrieves in the observer (the receiver) the neural circuit encoding the motor representation of the same gesture and, in this way, allows the receivers to grasp the message (gesture) of the sender (Rizzolatti and Arbib 1998). Because of the mirror system’s functional role, it was proposed that it may have played a key role in the evolution of a communication system based on hand gestures that paved the way to human language (Arbib 2005; Corballis 2010; Gentilucci and Corballis 2006; Rizzolatti and Arbib 1998; Zlatev 2014). One of the key elements at the basis of this hypothesis is the fact that area F5 of the ventral premotor cortex of the macaque brain is homologous to Broca’s area in humans (specifically to the dysgranular area 44, the posterior part of Broca’s area) (Rizzolatti and Arbib 1998), traditionally considered as a ‘speech area’ (Broca 1861; Embick et al. 2000), but also involved in motor function unrelated to speech (e.g., complex hand movements and sensorimotor learning and integration, Binkofski and Buccino 2004). As Broca’s area developed from
an area originally involved in the processing of action, from this perspective it is assumed that the ability to recognize and perform actions related to praxis, such as those involving the manipulation of objects, provided the basis for the development of the ability to perform and recognize communicative hand gestures that, in turn, provided the evolutionary basis for the development of the brain mechanisms that support spoken language. The extension of the brain systems at the basis of spoken language has been a consequence of the fact that the precursor of Broca’s area was equipped, before speech, with a mechanism for recognizing actions made by others. This mechanism was the neural prerequisite for the development of the inter-individual communication, first based on hand-gesture, and finally on speech (Rizzolatti and Arbib 1998: 190).

This gesture-first account is confirmed by several studies on monkeys and apes that have shown the existence of significant differences between vocal and gestural communication (and, more generally, bodily communication) among these animals. Specifically, it is widely attested that, in nonhuman primates, gestural communication systems are more flexible than vocal ones. The vocalizations of nonhuman primates are, for the most part, genetically determined. Each species has a relatively limited repertoire of calls whose acoustic characteristics are mainly fixed at birth and show minimal change during development (Cheney and Seyfarth 2010; Hammerschmidt and Fischer 2008). In addition, neurological evidence suggests that the vocal productions of monkeys and apes are mainly related to their emotional states (e.g., fear, excitement, etc) and this fact limits the communicative power of vocal calls (although see e.g., Clay and Zuberbühler 2014). In contrast, hand gestures in nonhuman primates can be produced voluntarily by the animals and because of this they can be used in a more flexible way than vocalizations. Great apes use gestures in different contexts to communicate different things (cf. Pollick and de Waal 2007). Apes’ gesture production takes into account the attentional state of the recipient: visual gestures (gestures that are not accompanied by any sound) are frequently used when the receiver is paying attention to the indicator (Tomasello and Call 2007), while auditory and tactile gestures are produced to attract the attention of an individual who are not looking at the signaller (Tomasello et al. 1994). So, although vocalizations are an important mode of communication for most primates and despite that, the vocal mode of communication intuitively is often considered a precursor of speech, the vocalizations of nonhuman primates have little in common with human language (Eberl 2010; Hammerschmidt and Fischer 2008; Ploog 2002; Tomasello 2008).
The experimental evidence and the theoretical arguments discussed in this section allow us to maintain that the embodied mind (specifically, the attention to action that characterizes this approach to cognition) represents a plausible conceptual framework to account for the gestural foundation of language in a phylogenetic perspective. In accord with this result, we can say that even the second step of our argument becomes clearly plausible.

6. Conclusion

In this paper we showed that the interplay between the grounding problem and the theories of gestural origins of language can be used to elaborate a naturalistic model of human communication. In a model of language that conforms with the principle of evolutionary plausibility, in fact, the relationship of language with external environment is a fundamental characteristic (a specific characteristic of the more general relationship between organism and environment). From this point of view, the embedded model of cognition offers a fruitful conceptual framework for a naturalistic approach to the study of language. Given the close relationship between the embedded model and the gestural theories, it is possible to argue that the embodied cognition represents the conceptual bridge to put together the gestural origins of language with the grounding problem. Such a standpoint opens the way to a truly naturalistic perspective of human language.

Acknowledgments

We want to thank Sławomir Wacewicz and one anonymous reviewer for helpful comments and suggestions.

References


Ferretti, F. (2007). *Perché non siamo speciali* [Why we are not special]. Roma and Bari: Laterza.


Abstract. Linguistic politeness (LP) refers to the set of “linguistic features mediating norms of social behaviour, in relation to such notions as courtesy, rapport, deference and distance” (Crystal 2008). Although researchers (e.g. Eelen 2001, Watts 2003) agree that it is intimately connected to normativity, group hierarchy and cooperation – the core questions of human ethology and human behavioural ecology – linguistic politeness has largely been neglected from those perspectives (Eibl-Eibesfeldt 1989 being a notable exception). In this paper we spell out the significance of a naturalistically oriented study of LP, outline a research agenda, and identify a number of methodological problems whose resolution is a prerequisite for such an approach.

Keywords: linguistic politeness, politeness, impoliteness, ethology, language evolution, signaling, normativity, cooperation, group hierarchy
1. Introduction

Recent interest in the topic of language evolution has created a research space for the active pursuit of questions related to the evolutionary origin and development of the language faculty – the uniquely human ability to acquire and use natural language. As a result, there now exists a considerable body of literature dealing with origin scenarios for many of the central aspects of the language faculty, such as phonology (MacNeilage 2008), morphology (Carstairs-MacCarthy 2010), and syntax (e.g. phrase structure out of computational brain patterns involved in motion, Calvin and Bickerton 2000; or re appropriation of the conceptual “who did what to whom” structure for the processing of semantic role, Bickerton 1997). Despite this, relatively little attention has been devoted to the origins of pragmatic competencies1; the capacity for linguistic politeness included.

Consider, for example, the influential taxonomy of the language faculty proposed by Hauser, Chomsky and Fitch (2002). Their account overtly prioritises syntax (or more precisely, the combinatorial capacity for creating recursively organised strings out of discrete elements) by terming it the “faculty of language, narrow sense” (FLN), but also posits two additional components of the “faculty of language, broad sense” (FLB): the conceptual-intentional system and the sensorimotor system. Pragmatic competence does not fit comfortably into any of these proposed classes, but it forms an inalienable part of the human capacity for language. An evolutionary view of language solely as a tool for the transmission of information is both limited and limiting; given what is known about human communication, non-human primate communication, and communication in general, it is abundantly clear that the development of language must have been driven by multiple social constraints, including factors related to diffusing aggression (Lorenz [1963] 2002), promoting social bonds (Dunbar 1996), maintenance of hierarchy and social cohesion (Eibl-Eibesfeldt 1989).

Linguistic politeness is a highly heterogeneous phenomenon, subsuming several distinct subtypes, each manifesting themselves via a variety of culture-specific politeness devices. Despite this, the presence of some form of polite linguistic behavior is considered a linguistic universal (Brown & Levinson 1978, 1987). Researchers agree that it is in essence a linguistic phenomenon, rather than a purely sociological or ethnological one (Lakoff 1975, Leech 1983, Watts 2003, Eelen 2001). We will make the case that

---

1 Sensu Robin Lakoff (1975).
by its very nature linguistic politeness needs to be considered in research on language evolution.

2. Linguistic Politeness

Linguistic politeness is notoriously hard to define, but we begin with Crystal’s (2008) encompassing definition of LP as those “linguistic features mediating norms of social behaviour, in relation to such notions as courtesy, rapport, deference and distance”. In support of this definition we offer the following additional characterizations:

- “[l]inguistic means ... to reduce social friction” (Lakoff 1975);
- “strategic conflict-avoidance [directed at] interpersonal comity” (Leech 1983);
- “a set of linguistic means for softening face threatening acts” (Brown & Levinson 1987);
- “interpersonal supportiveness” (Ardnt & Janney 1985);
- “cooperative social interaction and displaying mutual consideration for others” (Watts 2003);
- “joint venture directed at social harmony and equilibrium” (Fraser & Nolen 1981).

These excerpts are representative of the literature in defining LP functionally in relation to its role in maintaining social harmony and promoting conflict-free interactions. Defining politeness in ways other than functional is difficult due to cultural variation and lack of an established structural description. The specific linguistic devices through which LP is expressed are taken from broad set of possibilities, including:

- explicitness and context-independence; for example, the preferential use of nominals to pronouns, as in: “Give me a nice apple” in contrast to “Give me a nice one”;
- the use of complex clause patterns involving, for example, embedding (such as relative clauses) and subordinations (such as hypotactic clauses);

---

2 “Despite several decades of sustained scholarly interest in the field of politeness studies, a consensual definition of the meaning of the term ‘politeness,’ as well as a consensus on the very nature of the phenomenon, are still top issues in the current research agenda.” (Pizziconi 2006: 679)

“A main problem, we suggest, is the lack of agreement about how politeness should be defined as object of study.” (Janney & Arndt 2005: 22)
– the use of “high” register, the use of explicit politeness markers (such as courtesy subjuncts: “thank you”, “please”, “excuse me”);
– the use of passive and circumstantial voices (e.g. “It is regretted that...
– replacement of personal pronouns with indefinites (e.g. “Somebody might think that...
– point-of-view distancing, such as the present-past tense switch (e.g. “I was wondering whether...
– invocation of general rules (e.g. “I’m sorry, but late-comers cannot be seated till the next interval”);
– nominalisation (e.g. “I am surprised at your failure” instead of “I am surprised that you failed”).

3. Relevance

Research on linguistic politeness has its origins in cultural studies and related disciplines, and has generally been conducted following qualitative methodologies (e.g. see Radcliffe-Brown’s foundational study of the concept of respect; 1952: 107ff.). Recently, there has been growing interest in exploring LP phenomena from a more Cognitive Science perspective (e.g. see Escandell-Vidal 1996).

Nevertheless, investigations from an ethological/evolutionary perspective are rare (Eibl-Eibesfeldt [1989] being a notable exception). We find this somewhat surprising given that politeness researchers (e.g. Eelen 2001, Watts 2003) largely agree that LP is closely connected to the core questions of human ethology and human behavioural ecology, cf. e.g. Watts et al. (1992: xlv):

We believe that [politeness] is tied up with the most basic principles of human, socio-cultural organization involving conceptualisations of appropriate individual behavior, in particular linguistic behavior, the structuring of interpersonal relationships within social groups and, above all, the nature and distribution of power.

In this section we illustrate this intimate link by discussing three areas of special interest: normativity, group hierarchy and cooperation.
3.1. Normativity and evaluativity

Politeness phenomena are treated as essentially evaluative and normative. This is how Gino Eelen (2001: 35) characterises *politeness* and *impoliteness*:

Generally speaking, politeness comprises a positive, and impoliteness a negative evaluation. The noun ‘politeness’ is associated with the ‘civilized’ forms of behaviour (Blum-Kulka 1992: 258), as well as with ‘tolerance’, ‘good manners’, and ‘being nice to people’ (ibid.: 257), while the adjective ‘polite’ strongly correlates with adjectives such as ‘appropriate’, ‘friendly’, or ‘respectful’, and generally leads to ‘pleasant’ interactions (Ide *et al.* 1992: 290). As Ehlich (1992: 75) puts it, ‘polite’ actions are positively marked.

The presentation of the history of the concept of politeness and its cultural variants is well beyond the scope of this work. We shall only indicate that the origins of the modern English concept of “politeness” dates back to the sixteenth century, when it signified “... socially cooperative behaviour to be displayed by male members of the nobility at court” (Watts 2003: 36). One of the principal elements of politeness so construed was the “ability to carry out civil or familiar conversation” (Watts 2003: 36). Konrad Ehlich ([1992] 2005) posits that the idea of politeness is based on evaluation against a standard. He notes that the standard of politeness first emerged in late Middle Ages and was initially related to the rules of *courtoisie*/courtesy, which regulated the life at court. Later, the *courtoisie* standard was challenged by the competing standard derived from the notion of civilité/civility, promoted by the rising urban culture. The two standards finally settled into the European, and specifically English, notion of politeness as codified rules of ideal interaction.

Certainly, the notion of politeness characterised in this way is culture specific, and as has been demonstrated by numerous studies, conceptualisations of politeness may greatly differ according to variables such as ethnicity, social class, or even gender. However, the same studies testify to the existence of the universal politeness element, which comprises a set of core interactional tactics to show “the considerateness for the other person” (Watts 2003: 14). The universal norm of considerateness together with its culture specific applications constitute the yardstick against which the behaviour of interactants is evaluated (also self-evaluated). Therefore, as Eelen argues, the key problem of research on politeness is to account

In view of the above, politeness researchers are directly concerned with the normative aspects of interaction: inasmuch as politeness standards form the basis of evaluating interactants’ behaviour, they constitute interactional norms. Such an assumption lies at the heart of the “social norm view of politeness” formulated by Fraser (1990: 220) and widely accepted in current politeness research (Eelen 2001: 42–43). Even approaches which play down normative sentiments tacitly or openly acknowledge the normative character of politeness. In particular, this is true of Robin Lakoff’s and Geoffrey Leech’s theories, based on the idea that politeness should be described in terms of conversational maxims derivable from Grice’s Cooperative Principle.³ Lakoff and Leech postulate that politeness, understood as a set of conversational maxims, is an essentially linguistic principle, which regulates the relationship between the semantic meaning of an utterance and its pragmatic sense (Leech 1983: 4, Eelen 2001: 122). Leech is probably the fiercest opponent of including normative considerations in the description of linguistic politeness. He argues that conversational maxims, including politeness related maxims, acquire their normative character only when set in a social context. From this perspective, norms are not taken to belong to the language system as such, and therefore normative accounts (including LP) should be relegated from the discipline of linguistics. However, even Leech is forced to acknowledge that conversational maxims are able to perform their linguistic function only by virtue of being prescriptive rather than descriptive: maxims stipulate what speakers are required to do when engaged in communicative acts. Since it is impossible to get rid of normative accounts altogether, Leech emphasises the need to clearly differentiate between the social-normative function of maxims and their linguistic nature (1983: 84).

To sum up, normativity appears to be an indispensable element of linguistic politeness theories. This is the effect of the evaluative character of politeness phenomena, as rules of linguistic politeness are taken by interactants to constitute social norms regulating discursive communication. This has led many researchers to postulate that norms are central to the study of politeness (the social norm view of politeness). Even those who deny normativity a place in a theory of linguistic politeness are

³ For the criticism of the use of Grice’s CP by Lakoff and other politeness theories see Dynel 2009.
nevertheless forced to acknowledge that social norms constitute an essential component of the commonsensical understanding of politeness.

3.2. Group hierarchy

Watts et al. (1992) specifically highlight “the nature and distribution of power” in their description of politeness quoted above. Hierarchical structure within social groups is typically mirrored (and to a considerable extent, reinforced) by group-specific LP norms and devices. Being intuitively attuned to those norms constitutes an integral part of a speaker’s linguistic (pragmatic) competence. Breaches of those principles – such as inappropriate use of honorifics or the T/V distinction – are no less noticeable than breaches of grammatical rules, and appear to be more intensively policed by the conversational group.

The correlation between the levels of social hierarchy and the degree of elaboration of politeness devices is well documented both between and within social groups. Cultures whose ethos is based on egalitarian sentiments, on the values of friendliness and empathy, are designated as positive-politeness cultures – “the friendly back-slapping cultures, as in the western U.S.A., some New Guinea cultures, and the Mbuti pygmies” (Brown & Levinson 1987: 245). By contrast, cultures which emphasise social stratification, deference, and ceremoniousness are defined as negative-politeness cultures – “the negative-politeness cultures are those lands of stand-offish creatures like the British (in the eyes of the Americans), the Japanese (in the eyes of the British), the Malagasy ... and the Brahmans of India” (Brown & Levinson 1987: 245). Within particular cultures, the relative social standing of an individual predicts the degree of reliance on various LP devices: “elites, whose status rests on social conventions and who resent impositions, tend to keep more social distance around them, and use and expect more politeness” (Pinker 2007: 387–388).

It is interesting to note that in many languages the earliest addresative forms coding status differences are derived from familial addresses, such “father”, “mother”, “son”, “brother”, or “sister”, which with the growth of social diversification started to be employed outside the family context as indicative of either egalitarian relations (most importantly “brother”) or elitists ones (such as the “father-son” duo) (Ehlich 1992: 82ff).
3.3. Cooperation

The use of politeness can be seen as facilitating cooperation. Conversational corpora indicate that the accumulation of politeness strategies is typically present in two interactional contexts⁴:

- when conversants introduce a goal they intended to pursue by means of conversational interaction, as is the case in the following invitation: “Uh: would it be alright if we came in a little early”⁵ (Davidson 1984: 115);

- when conversants signal difficulty in accomplishing a goal that has already been introduced into interaction, as in Turn 2 of the following sequence, where B rejects A’s invitation:

  Turn1 A: Uh if you’d care to come and visit me a little while this morning I’ll give you a cup of coffee
  Turn 2 B: > hehh well that’s awfully sweet of you, I don’t think I can make it this morning. I’m running an ad in the paper and I have to stay near the phone⁶ (Atkinson and Drew 1979: 58).

In the first instance, politeness can be seen as serving to recruit a fellow interactant’s help in accomplishing a specific goal (e.g. accepting an offer, fulfilling a request, or answering a question). In the other context, it is best explained as a compensation for the inability to forward a previously introduced goal – here, the dominant function of politeness is to ensure that the blockage to the goal accomplishment does not eliminate the possibility of interactants’ cooperating in the future.

---

⁴ For a longer discussion see Żywiczyński 2010, and specifically for cooperation about intra- and extradiscursive goals in conversation from an evolutionary standpoint, Żywiczyński & Wacewicz 2012.

⁵ The politeness effect of the utterance depends on the use of conventional indirectness, whereby the question format is used to formulate the speech act “invitation”. The other politeness strategies include the present-past tense switch (“would” and “came” instead of “will” and “come”), complex clause structure (the use of the embedded conditional instead of a simple assertion such as “I come in early”), and the adjectival “a little”, which minimises the imposition.

⁶ The politeness effect of the utterance is created by the use of the delay signal (“hehh”) and the announcer (“well”), signaling the speaker’s unwillingness to decline the prior invitation, expression of appreciation (“that’s awfully sweet of you”), indirectness in formulating the declination component (“I don’t think I can make it this morning”) and specification of the reason for declining the invitation (“I’m running an ad in the paper and I have to stay near the phone”).
When considering politeness in evolutionary terms, a reference can be made to Axelrod’s concept of the “tit-for-tat” strategy, thought to stabilise cooperation by means of continual reciprocation of costs. Explaining how cooperation can be promoted in real-life situations, Axelrod makes two statements: (i) stable cooperative strategies tend to “enlarge the shadow of the future,” meaning that initial cooperation incurring small cost promotes subsequent cooperation incurring bigger cost; and (ii) “frequent interactions help promote stable cooperation” (Axelrod 1984: 126–133, cf. Mithen 2005: 213–214). Linguistic politeness can easily be conceptualised along these lines – it entails token costs, is easily repeatable and furthermore can be used by conversants to diagnose their mutual commitment to engage in future cooperation involving higher cost (such as forwarding each other’s goals).

The evolutionary significance of politeness can also be approached from the standpoint adopted by Henrich & Henrich (2007). In their micro-ecological study carried out on the Chaldean community in Detroit, they observed (i) that people tend to cooperate with individuals who share their norms and expectations (typically members of one’s ethnic group) and (ii) that interacting with such individuals actually accrues greater benefits than interacting with individuals who do not share their norms and expectations (2007). Supported by other works (Gil-White 1999, 2001; McElreath, Boyd, Richerson 2003), this led to the formulation of the dual inheritance hypothesis, which postulates a dual inheritance of culture and genes, and is based on the following premises: cultural capacities represent adaptations, and cultural learning mechanisms give rise to a robust second system of inheritance (cultural evolution) that operates by different transmission rules than genetic inheritance.

### 4. A candidate research agenda of evolutionarily-oriented LP studies

Having made the case that LP falls within the scope of the field of human ethology and human behavioural ecology, we will now outline a number of open research questions related to LP that we think are likely to be of particular interest to naturalistically oriented researchers into human behaviour.
4.1. In search of politeness universals

One of the key theoretical problems in politeness studies concerns the universality of politeness phenomena across cultures and languages. The foundational works in this research area were steeped in the tradition of Western linguistics and sociology. In fact, apart from Brown and Levinson’s *Politeness* (1987), all major contributions to the subject from 1970s and 80s were based on the analysis of British or American English (e.g. Lakoff 1973, 1975, Leech 1983, Fraser & Nolen 1981). Nowadays, the question about the universality of politeness is posed in a context of much richer cross-cultural and cross-linguistic research, and is posed much more tentatively. However, the majority of leading authorities in the field agree that members of *each* speech community have a set of linguistic means to express politeness, including *im*politeness (see Eelen 2001, Watts 2003, Culpeper 2011, Locher & Bousefiled 2008). Following Robin Lakoff, it could thus be argued that the knowledge of politeness strategies forms an integral part of the pragmatic competence of a particular language user (Lakoff 1973). Such a claim is often accompanied by a stronger postulate that politeness is expressed by a definite set of interactional strategies; however, the distribution in the use of these strategies varies from culture to culture. This gives rise to comparisons to Chomsky’s universal grammar in syntax, and the search for the universal grammar of politeness has proven to be as frustrating as the one undertaken by syntacticians inspired by Chomsky.

Many of the recently proposed models of politeness – such as Eelen’s “Politeness1/Politeness2” (2001) model or Watts’s habitus model (2003) – tend to emphasise the culturally constructed and context dependent nature of politeness rather than the universalistic sentiments of Brown & Levinson, Lakoff or Leech. However, even they seem to affirm the existence of core linguistic means to express politeness. This “nature-nurture-type” controversy parallels a long-running debate in linguistics about the ontogenetic status of syntax: given that the combinatorial mechanism of syntax is universal, but its details are culturally specified, does it result from a genetically hardwired „universal grammar”? Accordingly, determining which of these explanations is more likely for LP would have significance for the ongoing debates on the nature of cultural and linguistic universals (e.g. Evans and Levinson 2009).
4.2. Costs, benefits and stability of linguistic politeness

There is universal agreement among politeness researchers that politeness has a positive, and lack of politeness a negative social evaluation; this would imply the existence of benefits to being polite. However, being polite does not come at any (obvious) cost, i.e. LP is what in signaling theory is known as a “cheap signal” (Van Rooy 2003). A classic insight from signaling theory is that under normal circumstances, signaling strategies devoid of any inherent costs can easily be used for manipulation, and therefore potential receivers usually disregard them as unreliable (Krebs & Dawkins 1984). Accordingly, we could ask questions such as “why isn’t everyone polite all the time?” or “why do people (apparently) react positively to polite individuals”? The stability of politeness as an apparently cheap but beneficial strategy is hard to explain unless we assume some kind of inbuilt safeguard against its manipulative use – some cost that is incurred to the polite individual.

One way to see linguistically polite utterances as more costly than nonpolite or impolite ones is by positing that the former require more articulatory (e.g. increased length) and cognitive effort of on the part of the speaker. For example, Jan-Ola Östman (1989) (originally intending to verify John Haiman’s [1983: 783] famous proposal that “the social distance between interlocutors corresponds to the length of the message, referential content being equal”) reached a conclusion about the iconic correlation between the length of an utterance and the degree of politeness communicated by it. Östman’s experiments on an artificial language demonstrated that longer messages, either due to the length of a whole sentence or the length of individual words, are consistently interpreted as more formal, and hence more polite, than their shorter equivalence with the same semantic content. A similar point is made by Leech (1983: 108), who explains the increase in length of polite utterances as resulting from growing indirectness of the associated Speech Act.

The relation “more cost = more politeness” was also observed by Brown and Levinson (1987). In their theory, the weight of a Face Threatening Act (FTA) motivates the choice of politeness strategies – the weightier an FTA is, the stronger the politeness strategies which must be used. This view is related to the hierarchisation of politeness strategies in Brown and Levinson’s system: it is assumed that negative politeness strategies, which require more articulatory and cognitive effort on the part of the speaker, are able to produce stronger politeness effects than are positive politeness strategies, and hence are used in contexts that necessitate increased politeness, that is, where the relative values of personal distance (D), power difference (P), and
the rate of imposition (R) are high. Thus, the weightiness of an FTA can be predicted from the type of politeness strategies that accompany it.

Potential alternative construals of the cost of politeness involve the degree of personal attentiveness (Clark & Schunk 1980) or the analogy to monetary exchanges. The latter is developed by Werkhoffer (2005), who – drawing on Georg Simmel’s (1990) *Philosophie des Geldes* – states that despite being a social construct, politeness, like money, can itself motivate and structure courses of action. Like the value of currencies, the value of polite expressions can fluctuate according to the change in the social order, and like with money one can pay with linguistic resources what is due in a given socio-communicative context. Finally, Van Rooy (2003) has proposed to view LP as a *handicap*, speculating about factors such as “incurring social debt”, or greater complexity of polite utterances. To date, however, no account of the costs of polite linguistic behaviour has been offered that could be considered definitive.

It is also possible that LP – or a tendency for greater adherence to the particular politeness standards of the speech community – may be a manifestation of a generalised *norm-abiding phenotype*. On this explanation, being linguistically polite would be a special case of a general phenomenon of obeying the norms of one’s social group. As a manifestation of the common underlying trait, the two would co-vary, and so the more linguistically polite individuals are also predicted to more closely observe the non-linguistic social norms. This explanation is compatible with cultural-evolutionary or memetic scenarios, on which LP devices can themselves be considered independently replicating cultural memes – non-functional but viral norms that exploit or “parasitise” our minds adapted for following norms and conventions.

### 5. Empirical issues

The study of LP, as with most other complex human social behaviours, is complicated by a number of empirical issues related to the collection and analysis of data. In this section we identify a number of the more prominent issues faced by LP researchers hoping to adopt an ethological perspective and suggest a possible first step in undertaking such research.

As was discussed earlier in this text, the usage of linguistic politeness devices is universal across cultures, but the specific strategies adopted by particular groups is culturally determined. Many studies put emphasis on the heterogeneous nature of politeness, underlying the uniqueness of what
it means to polite across cultures. For example, in the Anglo-Saxon cultures
politeness is primarily understood as keeping appropriate personal distance;
in the Russian culture, politeness, or вежливость (Sifianou 1992), largely
boils down to the use of overt politeness markers and avoidance of vulgarisms;
the Chinese limao, corresponding to politeness, refers to upholding social
harmony by means of etiquette (Gu 1990); while politeness among Israelis
is subsumed under two concepts – numis, which regulates the public sphere
and requires the use of formal register, and privately oriented adivut, which
consists in displaying care and attention (Blum-Kulka 1992).

What is more, speech communities are not monolithic in their
perception and use of politeness. A growing number of works indicates that
gender is an important factor in the choice of politeness strategies, whereby
males preferentially employ distance-closing strategies (i.e. Lakoff’s
camaraderie strategies [1975] or Brown & Levinson’s positive politeness
strategies [1987]), while women tend to choose strategies that underline
status differences, promote indirectness and conflict avoidance (i.e. Lakoff’s
strategies of distance [1975] and niceness [2005b] or Brown & Levinson’s
Brown & Levinson take this point further, insisting that any major distinction
in the social fabric of a community is accompanied by a difference in the
usage of politeness strategies. Hence, they insist that any informative
description of a group’s politeness behaviours should be focused on how
differential distribution of politeness strategies is related to social parameters
and variables which exist in this group (1987: 242ff).

Another problem endemic in the study of LP phenomena is their
operationalisation. The intuition of any language user unequivocally suggests
that politeness is gradable – that it is possible to produce utterances varying
in the degree of politeness. This intuition has been used by many researchers
who employed surveys to elicit politeness judgements, i.e. respondents were
asked to arrange utterances according to the perceived intensity of politeness
effects, rendering lists such as the following one:

May I borrow your car please?
I’d like to borrow your car, if you wouldn’t mind.
Would you have any objections to my borrowing your car for
a while?

---

7 There have been many attempts to explain the logic of this gradability, with researchers appealing to utterance’s length (Östman 1989), speech act indirectness (Leech 1983), or speaker’s and hearer’s needs (Brown and Levinson 1987), to mention just a few proposals.
Could you possibly by any chance lend me your car for just a few minutes. There wouldn’t I suppose be any chance of your being able to lend me your car for just a few minutes, would there? (Brown & Levinson 1987: 143–142)

These endeavors have been accompanied by a growing realization that there are no absolute politeness measures because in real life situations, very subtle changes in the textual, physical or pragmatic context can lead to substantial differences in the perceived politeness of verbally identical utterances (such is e.g. the case of elaborate politeness crossing into irony; cf. Leech 1983, Brown & Levinson 1987). This is related to the hotly discussed problem of linguistic markers of politeness and impoliteness, reflected in the universal lexical-conceptual distinction between polite-impolite, and their pragmatic exponents which politeness researchers typically discuss in terms of speech-act indirectness or implicitness as defined by Grice. For this reason, in the current research instead of characterising them as inherently polite, utterances and expressions are considered in terms of their potentiality to be interpreted as polite and conditions under which such interpretations are produced (see e.g. Watts 2003). However, in narrowly defined contexts it does seems possible to build a corpus of utterances that – within those contexts – could be rated ordinally for their relative politeness level with some reliability. Trimmed and annotated, such corpora could then be used to test for correlations with other variables of interest, such as personality traits or the degree of cooperativeness. To give ground to the speculation about LP universals, results obtained in this way should then be compared with data coming from large-scale linguistic corpora.

Finally, the naturalistic account of politeness must not overlook the problem of impoliteness. A growing body of linguistic research on impoliteness suggests that impoliteness should be viewed as distinct form LP (e.g. Culpeper 1996, 2005, 2011; Culpeper et al. 2003; Bousfield 2008, Bousfield and Locher 2008); accordingly ethology and human behavioural ecology could provide an important insight into the nature of impoliteness-related phenomena, e.g. relating it to the problem of aggression as envisaged by Eibl-Eibesfeldt (1989).
6. Conclusion

Linguistic politeness (LP) is a refractory object of study, variable in its manifestations, lacking a single authoritative definition and difficult to operationalise. At the same time, researchers into LP agree about its cultural universality and intimate links to the foundations of human social organisation. In the sections above, through highlighting the relevance of LP to the key areas of human ethology and behavioural ecology, we have made a case for a naturalistically oriented agenda of research into LP. In particular, since LP is widely considered to be an essentially linguistic phenomenon, we point to the need for incorporation of this facet of human linguistic behavior into language evolution research.

Acknowledgements

This research was supported by grant 3704/B/H03/2011/40 from the Polish National Science Centre.

References


Sounds of protolanguages: 
Some preliminary insights 
from developmental psychology

Abstract. In present paper I examine the transition from crying to word in human infants and propose that it might be a treasure acoustic box helping us to conceive how our ancestor’s sounds might have been, and to outline breakpoints that might help us in making hypotheses about possible subsequent phases of language evolution.

Keywords: protolanguage, onomatopoeia, animal communication, speech, protophones, vocalization, language origins

1. Introduction

We know that language in Homo sapiens is unique among mammals, due to: a) peculiar vocal apparatus allowing the existence of an articulatory space more similar to singing birds’ than to mammals’ (Marler 1970; Bolhuis et al. 2010), b) specialized brain areas allowing acoustic analysis of input speech sounds and nervous motor control of output speech sounds (Broca and Wernicke), c) a specialized cognitive equipment allowing the mapping of any acoustic label to objects in the world (Lenti Boero and Bottoni 2006). Although language includes syntax and grammar, I believe that words (and their referents) are the ground of linguistic communication, thus in present paper I will adopt an ontogenetic view addressing the features described in points a), b), c). Those features develop in the first year of human life,
when the human infant proceeds “from cry to words”, i.e. from an early species-specific communication (see below), analogue to other mammalian infants (Peters 1980; Lingle 2012), towards a semantic holophrastic communication introducing syntactic discourse organization (Kagan 1984). It is acknowledged that evolutionary studies may profit from developmental ones and vice versa (Piaget 1972; Butterworth et al. 1985): even if we should be cautious in not adopting a recapitulationistic point of view, language evolution theories might very well profit from ontogenetic studies, especially from those aimed at identifying constraints common to both development and evolution (Costall 1985; MacNeilage 2008). Phylogeny is not a sequence of adults but a sequence of entire ontogenies (Striedter 2005), thus including the immature forms. For this reason, the path of the infant child toward language might be a treasure box because in their first year of life human infants go through a sequence of subsequent anatomical and neurophysiological changes ending in words. Those changes should be considered as steps travelled along the hominid phylogenetic line from the utterance of relatively simple sounds, as our nearest living species (great apes) do, to the complex articulate sounds that compose language (Mithen 2005; Tattersall 1995).

The aim of present paper is to propose that the analysis of the human infant acoustic outputs might help us in defining a logical sequence of sound making that hominids underwent from one species to another. By logical I mean the fact that more simple anatomical features and nervous motor controls might have themselves possessed simple communication systems, and generated the more complex ones. I will try to sketch the communicative boundaries of those systems, and to sketch the possible selective pressures that acted upon those “improvements” from one stage to the next. My models will be animal and early human communication. In our ancestors’ path toward language natural, parental, and/or sexual selection might have acted in order to maintain traits relative to point 1 and 2. In the paragraphs to follow, I will try to hypothesize which selective forces might have been at work.

2. Acoustic output of the human infant: crying, laughter and low intensity sounds

Crying is the first sound uttered by a human being. It is an alarm signal conveying the information that homeostatic imbalance is endangered (Lenti Boero et al. 1998). Crying has striking characteristics: loudness that can be
perceived at least at about 100 m around the emitter – but can also be intensity modulated, and long lasting duration: up to ten minutes (Lenti Boero, personal observation). According to Newmann (1985), it has no homologous in infant primates. Crying can communicate individuality (Cismaresco et al. 1990; Rocca and Lenti Boero 2005), and possibly sex of the caller (Rocca and Lenti Boero 2005). In addition, the infant cry communicates urgency to a recipient (Lenti Boero et al. 2008) and we know that contemporary mothers perceive their own 4/8 months old infant’s cry as “intentional” according to background variables, maternal emotional adjustment, and maternal interactive style (Feldman and Reznick 1996).

Some acoustic content of the cry can be extracted and paralleled to speech sounds, for instance, Italian vocalic sounds “a” and “e” and nasal sounds were heard by Italian musical trained listeners of cries uttered by less than one week old children (Lenti Boero et al. 2008). Interestingly, in middle age, the same prevalent vowels “a” and “ɛ/e” in newborns’ cry were reported by the Florentine poet Antonio Pucci (1310/1388), though we might not agree with his interpretation: “infant males cry emission include the “a” in memory of Father Adam, while infant females include “ɛ/e” in memory of Mother Eve, thus carrying on those cry all world’s troubles” (Giallongo 1997). Nervous motor controls for cry emission are the same as the ones reported in monkey models (Jurgens 1990, 1992; Jurgens and Ploog 1988; Lenti Boero 2009; Lester and Boukydis 1992), and the infant cry must be considered thoroughly an “animal” signal.

Beyond cry, infants utter vegetative sounds, such as cough and sobs that might have some non-intentional communication value poorly explored (Papoushek and Papoushek 1981), but possibly informing the care-giver about the infant’s physiological state. In addition, in positive situational contexts (Papoushek and Papoushek 1981), infants produce low intensity sounds with distinctive timbre, intonation and F₀ contour from three days of age; and, though we have a poor knowledge about the ontogeny and development of laughter, preliminary findings show that a fully fledged laughter may be observed at three and half month (Decurti and Lenti Boero, personal observation). A key characteristic of early laughter is that it is a low intensity sound that can reach few meters around the emitter. Thus a human infant is endowed since her/his early months of life with a basic communication system signaling alarm, well being and basic physiological yet not very alarming states, the only long distant signal being the cry.

We might thus think of a simple communication system based on the above described early infant sounds put under volitional control, as is the case for some animal species (Seyfarth and Cheney 1980; Lenti Boero 1992).
In a hominid social group endowed with the above described communication competencies, an individual could communicate alarm and urgency to group mates from a distance, or from a vantage point not reachable to predators, such as a tree, this communication might have included basic referentiality, as is the case in many living mammalian species: alpine marmots, (Lenti Boero 1992), dwarf mongoose (Rasa, 1986), vervet and Diana monkeys (Seyfarth and Cheney 1980; Zuberbühler 2005) among others. Eventually, low intensity sounds (Oller 2000) and laughter might have been used as a kind of acoustic grooming, sensu Dunbar (1993).

3. The human infant as sound analyzer

We know that during early development infant perception of surrounding sounds, including language, is much more advanced than motor competence (Lenti Boero in press): infants discriminate language phonemes, sharing this capacity with many animal species: rhesus macaques, dogs, chinchilla, quails, and parrots (Adams et al. 1987; Bottoni et al. 2003, 2009; Dewson 1964; Kluender et al. 1987; Kuhl and Miller 1975; Miller 1977; Morse and Snowdon 1975; Pepperberg 2007). This is a key point: why both human infants and many animal species are competent in phonemic discrimination? I believe that this ability is a subcomponent of a more general competence in acoustic spectra analysis, and that it must be widespread at least among vertebrates. Animals, similarly to us, must be updated with the auditory scene, they must distinguish between abiotic sounds from the environment, and biotic sounds uttered by the prey or dangerous predators (according to the viewpoint), between familiar and unfamiliar sounds, and this competence is only achieved by analyzing acoustic spectra (Bottoni et al. 2003).

Infants recognize the melodic contour of maternal language, and detect substitution of musical notes (Cooper and Aslin 1989; Mehler et al. 1988; Tervaniemi and Huotilainen 2003), they are also able to connect input and output sounds: infants shape their cries’ melodic contours on their native language (Mampe et al. 2009). In addition, at three months of age, infants are able to imitate musical pitch (Kessen et al. 1979).

What does this means for language evolution? I believe that the competences outlined above, and common to other vertebrates, are fundamental in language evolution, that indeed should be considered as the product of coevolution between receiver and emitter, analogously as other communication devices (Shannon and Weaver 1949). This is evident
4. The ontogeny of musical babbling and of articulated sounds

A key breakpoint in the path towards language is the shift from relatively simple sounds as described above to the articulated ones. This phase starts approximately by the time of 2 months when infants start producing very low intensity quasi-musical sounds or protophones, cooing or gooing, those sounds are mostly, but not only, vowel-like, having more complex melodic contours paralleling vocal emotional expression as the infant grows (Hsu and Fogel 2001). In addition, the infant improves her/his control of the intensity of the acoustic output, with high volume correlating with high excitement both positive or negative (Decurti and Lenti Boero, personal observation) (Papoushek and Papoushek 1981; Oller 2000; Ruzza et al. 2003). I defined those protophones as musibabbling (Lenti Boero 2009).

Then, from the third month on, sequences of vowels (voiced sounds) are progressively accompanied by protoconsonant sounds, progressively articulated in the retropharynx, in the dental/alveolar, velar and glottal space and bilabial (Oller and Eilers 1992; Roug et al. 1989). This phase might have an internal positive reinforcement, because we know that infants spend much time in exploring their new possibilities, and that this practice gives great pleasure to them (de Boysson-Bardie 2001; Papoushek and Papoushek 1981). Those steps are allowed by a better control of the respiratory cycle, which lengthens the exhaling phase, in favour of vocal emission (Papoushek and Papoushek 1981); and implies a more developed nervous motor control of the entire vocal apparatus (mouth, lips, nose, throat) than the one needed for crying: the acoustic energy can be canalized through the nose producing nasal sounds, or kept lower in the throat, thus producing harsh sounds, and lips can be used as sound makers independently from the mouth (Lenti Boero personal observation).

According to MacNeilage (1998, 2008), this early articulatory component derives from an original precursor in early mammals related to lower jaw oscillation in the purpose of ingestive movements (chewing, sucking, licking). Davis and MacNeilage (2002) notice that the utterance of consonant and vowels requires incompatible mouth movements: depression of the mandible (mouth opening) is required for vowels, while lower jaw elevation is required for consonants. The novelty in this aspect is the total contrast of the two functions: ingestion requires the closing of the air flow and consequently
of the vocal apparatus (if food enters the respiratory apparatus death might occur), oppositely sound making requires an open vocal apparatus enabling air to enter in order to fuel the acoustic energy necessary for vocalize.

We know that those sounds are very much appreciated by parents and are preferred in front of cry (Lenti Boero and Bottoni 2009), probably because they lack the aversive component of cry (Barr 2004; Frodi 1985; Frodi and Senchack 1990; Levisky and Cooper 2000; Zeskind and Lester 1978; Seifritz et al. 2003).

4.1. Selective pressures

The above could point to the selective pressures that might have been at work in favoring articulated sounds in front of the fixed relatively simple sounds described above. In fact, evolution is not directive, and natural selection “should operate at all stages of development” (Hogan 1988): this is particularly true for language (Locke and Bogin 2006). Indeed, those selective pressures favored the exaptation of the ingestion apparatus toward the emission of articulated sounds. Locke (2006) was the first to consider parental selection as an agent for preference for early articulated sounds over cry due to the stressfulness of the infant cry. The “Intrinsic Musicality Hypothesis” (Lenti Boero and Bottoni 2008) could confirm Locke’s proposal. This concept refers to the predisposition, shaped along hominid evolution, of our auditory system to generate conscious, “aesthetic” and/or emotional responses to all heard sounds (abiotic, biotic including human such as cry). Apparently, the aperiodicity of some segments of the infant cry, and its loudness, are aversive to listeners. In a pilot experiment Lenti Boero et al. (2009) showed that the aesthetic quality of musical protophones is rated higher than cry, and produces less stress and anxiety.

In adults, the musical protophone competence might have conferred higher sexual attraction to the carriers. Think of an individual endowed with a vocal apparatus allowing musical protophones, cooing and gooing sounds under volitional control. Because those sounds are nice, and have the possibility to be modulated in their fundamental frequency, they might be used in courtship and preferred over other less modulated signals. Darwin (1871) was the first to propose that musical ability might have been selected by sexual selection: “it is probable that the progenitors of man, either the males or females, or both sexes, before acquiring the power of expressing mutual love in articulate language, endeavoured to charm each other with musical notes and rhythm”. Mithen (2006, 2009) proposes that until the appearance of Homo sapiens, hominids utilized a musical, modulated, holophrastic way
of communication. Beyond musicality, more complex melodic contours might have been chosen as advertisers of good motor control, an important trait conferring advantage to individuals in the challenges of everyday savannah life and thus indicator of “good genes”.

4.2. Animal models

Vallet et al. (1997) propose an interesting model in canaries, where the ability to control singing is sexually selected as an advertiser of overall motor ability. As regards sound imitation, we know that mammals are not able to imitate surrounding sounds; however, many birds are able to imitate surrounding sounds, including the Grey Parrot (*Psittacus erithacus*) that was proposed as model for both music and language evolution (Bottoni et al. 2003, 2009; Pepperberg 2007).

To what extent do protophone articulated sounds allow a communication system? Mithen (2009) suggests that early communication might have included both emotional as well as mimetic (imitation from environmental sounds) signals. I agree with this position: sound imitation is possible with a good control of articulated sound emission, as early sound making shows, and a competence for acoustic spectra analysis (that the human infant demonstrate very early). By means of modulation of the vocal portion of sound emission our ancestors might have been able to introduce a subjective component by adding emotional information to the imitation of surrounding sounds.

The ability for sound mimicry might have been a great help in hunting: this might be confirmed by the hunting strategies of present-day hunters-gatherers and modern western hunters, who still use sound mimics – artificial or natural – in order to lure small prey in their surrounding for kills. Hunting success might have been an additional selective advantage because it enhances individual fitness and attracts mates.

But sound mimicry might have added another important advantage to individuals in groups: for group hunters, as in hominids (Tattersal 1995), it might be useful to share information with group mates about presence of prey or of an incoming danger. Indeed, sound mimicry has this double aspect: it lures the prey, but can also communicate its presence to fellow humans. This property could have raised the communication “entropy” in the social group by adding acoustical referential signs coupled with vocal emotional information to the gestural ones, thus freeing the emitter from being in the visual field of the audience, and might have been primarily shared within the kin hunting group (kin selection).
Is sound mimicry a full communication system? Emphatically not. Abiotic entities, such as stones and places in the environment are silent, a forest sounds only when moved by the wind, a stream has a sound, but a river or a lake might not, animals produce different sounds, and most of all, different actions, which are difficult or impossible to mimic by sounds. However, this might have been the heritage of earlier hominids to *Homo sapiens*.

### 5. Further development in infant communication

From 6 months of age infants enter the so-called **canonical stage** (de Boysson Bardie 2001; Oller 2000). Canonical syllables are composed by a “nucleus” of acoustic energy (a vowel), and at least a “margin” (a consonant), together those sounds last 100–550 milliseconds, the syllabic period guaranteeing the distinct perception of the syllable (Oller and Eilers 1992; Oller 2000). For the canonical stage to appear a crucial maturation is needed: the time for nucleus and margin emission are under nervous motor control for time length, in addition canonical stage is tied with acoustic automonitoring: deaf infants do not enter the canonical stage until the tenth month of life with rare exceptions (Oller and Eilers 1988). The canonical syllables phase is paralleled by many neuroanatomical changes that imaging studies have recently revealed (Matzusawa *et al.* 2001; de Graaf-Peters and Hadders-Algra 2005). In particular MEG results show that at sixth month of age there is an activation of the left inferior frontal portion of the infant brain (Imada *et al.* 2006).

#### 5.1. Social scaffolding in language learning

Most important aspect of language acquisition is the role of social influences on language learning (Kuhl 2004).

After six months, infants show a perceptual magnet effect from their native language category and reflect sensitivity to the distributional properties of sounds in their language. This effect is uniquely human and requires linguistic experience, in fact they focalize on linguistic sounds from their native language (Kuhl 2004). Also, it is of interest that by this time (5/6 months) human infants lose their encyclopedic hearing (de Boysson-Bardie 2001) that is the ability to discriminate the phonetic contrasts of all languages (Kuhl 2004).
5.2. The role of phylogenetic inheritance

In modern humans canonical babbling is scaffolded by motherese, a specific melodic contour that activates the right part of the infant brain (Homae et al. 2006). By means of motherese infants are naturally guided towards the right pronunciation of syllables of the native language, to put different syllables in a sequence of different vowel/consonant or consonant/vowel string, thus overcoming the first phases of babbling when sounds are often repeated (mama, dada etc.), and subsequently conducted towards the acquisition of the meaning of syllables strings i.e. words, (socially shared acoustic icons for items (objects or subjects) or meaningful motor sequences (actions) (Kuhl 2007). Motherese has a peculiar prosody: higher mean, maximum and minimum frequency, higher Fo frequency range, shorter sentences and longer pauses (Fernald and Simons 1989; Fernald 1992; Fernald and Kuhl 1987). Motherese is transcultural: ascending intonation contours for encouragement and descending ones for sedation were similarly used by Chinese, German, US mothers; another study showed that French, English, US and Italian mothers used the same intonation contours for attention seeking, approval, prohibition and comfort (Papoushek 1992; Fernald 1992).

This specific intonation contour was already noted by Darwin while observing his own infants, and Mithen (2009) believes that Homo sapiens’ motherese was inherited by earlier hominid melodic communication forms. Indeed, the prosodic aspects of linguistic communication are processed in the right part of the adult human brain (Zatorre et al. 1992) and it was demonstrated that infants perceive prosody even when sleeping (Homae et al. 2006).

It is universally acknowledged that language is a communicative tool socially shared, and all the studies of language acquisition confirm that (Kuhl 2004) language has been defined as non referent arbitrary acoustic (or gestural) significant for objects or actions. Thus, in a start-up situation of a human (Homo sapiens) social group sharing a complete language apparatus the linguistic code had to be built step by step.

5.3. Onomatopoeia: a transitory phase?

The first idea that earlier forms of language were onomatopoeic was proposed by Darwin in his Descent of Man (1871), and was inspired by his cousin and linguist Hensleigh Wedgwood, who had written a book on language five years before (Richards 1987). Very soon those ideas were discarded by the great Oxford linguist Friedrich Max Müller (1823–1900),
who derisively called them the “pooh-pooh” and “bow-wow” theories of language formation (Richards 1987), and this proposal was discarded. However, a very interesting and up-to-now unique study by Hashimoto and coll. (2006), might shed new light on those old theories: by means of an event related fMRI study those authors found that the brain regions involved in processing onomatopoeic sounds – i.e. sounds imitating the sound or action of objects, animals, and humans (e.g., “buzz”, “whirr”, “bow”, “hoot”, “meow”, “squeak”, “coo”, “hush”, and “boom”) – include the regions that process both nouns and animal sounds. Onomatopoeias and nouns are human speech sounds of similar frequency components, while onomatopoeias and animal sounds are often repetitive and acoustically simple sounds with frequency modulation, and Hashimoto et al. (2006) conclude that onomatopoeic sounds can serve as a bridge between nouns and animal sounds, and postulate that the onomatopoeic sounds contain both verbal and nonverbal sound components. As MacNeilage (2008) asserts, meaning should not arrive out of nowhere: the meaning of words is socially shared; thus, how did early human speakers build their shared vocabulary? I believe that there must have been a transitory phase between sound mimicry (simple repetition of animal sounds) inherited from earlier hominid forms and later, more modern “full human” forms of linguistic communication. And a phase of onomatopoeia (repetition and also re-creation of animal sounds) must have been in the middle, between sound mimicry and full-fledged language. Hashimoto & coll. (2006) indicate that, but it is impossible to say by now if this phase was already in the Homo sapiens domain or in earlier hominid forms. As a matter of fact, when the task of pronouncing the first words is accomplished, children frequently use onomatopoeic sounds in order to indicate surrounding objects. We also know that many, if not all, cultures have a repertoire of onomatopoeic words/sounds used with and by children (for Japanese, Hashimoto 2006; for Italian, Lenti Boero and Habegger, personal observation).

6. Discussion and conclusions

The maturation of the neurophysiological and anatomical apparatus allowing the first words production is a constrained path for human infants and must have been so along the hominid line. In present paper I try to describe that “logical” sequences underwent by hominids in their path towards language. I sketched three benchmarks: 1) a fixed (mammalian like) phase of sound production, allowing both alarm sounds with basic referents, and prosocial
low intensity sounds; 2) a second phase, when low intensity, music-like sounds start to be articulated paralleled by vocal contours. It is noticeable that in this phase infants possess a wide and open range acoustic competence, defined as encyclopedic hearing (de Boysson-Bardie 2001); 3) a third phase when consonant and vowel sounds are uttered and refined in the dyadic context, and are basically helped by motherese (Khul 2004). In this phase infants lose their encyclopedic hearing and focus only on the sounds of their native language (Khul 2007).

This series cannot be reversed in the sense that canonical babbling could not precede cry nor musilanguage. Because natural selection acts on present features and not for future potential benefits (Fagen 1981), I try to figure out the selective pressures that might have been at work along this path. This aspect is seldom considered in in language evolution studies, but see Locke and Bogin (2006) and Lenti Boero (2014).

In addition, I try to examine in more detail the characteristics of protocommunication that might have antedated the appearance of the first words: socially shared sounds indexing objects by means of acoustic arbitrary labels.

Was the transition from one stage to the subsequent the product of chance mutations regarding hominid communication tools? Or was it a byproduct of changes due to other aspects of hominid evolution? Or was the transition from one stage to another driven in an autocatalytic way? Those are among the still open questions in the complex puzzle of language evolution.

**Acknowledgements**

Financial support: The research going into the preparation of this commentary was supported in 1995, from 2001 to 2003, and from 2005 to 2007 by grants from MURST (Ministero dell’Università e della Ricerca Scientifica e Tecnologica), and by funding from the University of Valle d’Aosta from 2009 to 2013. I heartily thank one anonymous reviewer for invaluable (and challenging) comments on a previous draft of this paper.

**References**


The acoustic diversity in the phoneme inventories of the world’s languages

Abstract. A comparative analysis of multi-language speech samples is conducted using acoustic characteristics of phoneme realisations in spoken languages. Different approaches to investigation of phonemic diversity in the context of language evolution are compared and discussed. We introduced our approach (materials and methods) and presented preliminary results of research. We built an online database dedicated to voice acquisition and a storage of good quality speech samples collected across the globe. Software designed for automatic extraction and analysis of phonemes was developed and adapted for languages classification. Research involves both experimental and theoretical works that aim at gaining knowledge about phonetic diversity of languages across the world. Additionally, the expected results may be applied to verify the hypothesis of modern languages expansion from Africa, brought to attention by Atkinson.

Keywords: speech processing, languages comparison, phoneme analysis.

1. Introduction

Every language uses only a small part of possible articulation abilities. Young infants have the ability to acquire a spectrum of sounds broader than those existing in any particular language. Their individual articulation abilities are shaped by the culture that motivates them to master only some phonemes.
Archaeological clues may help us determine when the human kind developed the ability of articulation. However, the existence of anatomical potential for uttering of speech sounds does not imply the fact that speech was used. A significant question is when a sequence of phonemes started to convey complex information (Kirby 2007: 669–681). Besides the archaeological or traditional linguistic approaches, genetic methods, computational linguistic and acoustic analyses or their combinations (Cavalli-Sforza et al. 1988: 6002–6006) have become widely explored in recent decades.

Phonemic diversity (the number of perceptually distinct units of sound: consonants, vowels and tones in a language) was investigated with a view to estimating when spoken language began (Perreault and Matthew 2012: e35289). Using data from the UCLA Phonological Segment Inventory Database (UPSID) (Maddieson 1984: 162–163), they modelled the phoneme accumulation rate and calculated the time it would take a language to acquire the phonemic diversity observed today in African languages. Perreault and Matthew’s study suggests that spoken language appeared early in the history of our species during the Middle Stone Age in Africa, between 150–350 thousand years ago. However, they emphasized that the estimate should be treated as minimal, taking into account the possibility that during evolution the phonemic clock could have pushed back or reset as an effect of the population bottleneck. Information on world languages’ phonemes inventory, including their number, types or acoustic properties can be explored not only as an indicator of dating language origin, but also as a source of knowledge on the language evolution process, extinct spoken language reconstruction (Bouchard-Côté 2013: 4224–4229), population ecology and culture studies.

In this paper we present and discuss the state of undertaken research that is meant to bring a new approach to investigate phonemic inventories of languages. Collected data, proposed methodology of acoustic and statistical analyses together with first experiments are described. As a future result of the research, we hope to formulate new scientific evidence on languages’ similarities, which may contribute to the discussion on the origin of spoken language.

2. Background: a linguistic founder effect

The term founder effect originates from population genetics, and explains the loss of genetic diversity among the migrating parts of a population. In particular, the serial founder effect occurs when populations move over long

Quentin D. Atkinson in his article “Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa” (2011: 346–349) suggested that the founder effect may operate on human languages as well. Expansion should progressively reduce phonemic diversity with increasing distance from the point of origin, paralleling the serial founder effect observed in evolutionary genetics.

Atkinson examined geographic variation in phoneme inventory size using data taken from 504 languages in the World Atlas of Language Structures (WALS). His statistical analysis confirmed that the number of phonemes in a language is positively correlated with the size of population using these phonemes (Hay and Bauer 2007: 388–400) for both, languages and language family levels. Using the available data together with language location information and supplementing them with a series of linear regression predictors, he estimated a map of likely areas of languages origin and evaluated the model fit with the Bayesian information criterion. The model of the language level points to central and southern Africa as the places where language may come from; also in the family-level analysis, the region of origin is expanded to include the whole of Africa.

In another experiment, Atkinson took into account additional parameters that could have an influence on regional phonemic diversity due to increasing contact within and between groups and creating more opportunities to borrow new phonemes (e.g. the number of languages per unit of area), population densities or language area. Including these factors, he came to the same conclusions.

Atkinson’s outcomes are supported by detailed statistical analyses. His results seem to be comparable to those obtained for phenotypic diversity or mitochondrial DNA, but much weaker than for genotypic diversity (Li et al. 2008: 1100–1104).

2.1 Atkinson’s hypothesis: discussion

Atkinson’s article provoked immediate criticism (four technical comments followed by Atkinson’s two responses to the comments were published in the Science Magazine in 2012) and numerous citations. One of the objections was the simplification of the measurement of the phoneme inventories in the data used for the analysis (Wang et al. 2012: 657). Others (Hunley et al. 2012:
argue that most of serial founder effect predictions are violated for the phonemic data. According to their opinion, phonemes cannot provide information about more ancient evolutionary processes because they change rapidly. Van Tuyl et al. (2012: 657) criticized the method of forced linear regression across all continents and suggested that the approach would be proper within a single continent only. They also questioned the comparability of the phonemic diversity data used by Atkinson. They suggest that taking into account historical processes like migrations, conquests, and borrowings would explain language evolution more credibly than the founder effect solely. Concerns about suboptimal data and biased methodology in Atkinson’s work were raised also by Cysuow et al. (2012: 657). In Jaeger et al. (2013: 1042) technical comments are made and other statistical simulations are performed, and the authors suggested that the type I error rate (the incorrect rejection of the null hypothesis) in Atkinson’s analysis is hugely inflated.

Scientific controversy about Atkinson’s results has induced intense research in this field motivating us to develop an independent method to assess the hypothesis.

2.2. Motivation and objectives

Inspired by Atkinson’s hypothesis of modern languages’ expansion from Africa, we want to analyse the acoustic features of phonemes from hundreds of languages from all over the world. We think that an analysis of phonemic diversity can be obtained from the analysis of phonemes acoustic features and that this analysis along with the methods used in evolutionary genetics will allow us to create a taxonomy of natural languages.

A set of sounds (allophones) which represent a phoneme in one language can be regarded as more than one phoneme in another language. Thus, analysing a number of phonemes gives us information about languages from a perceptual point of view and an acoustic analysis provides information about languages from a phoneme articulation point of view. The acoustic analysis of phonemes allows us to calculate spectral distances in order to obtain acoustic similarities between a pair of languages.

3. Database

To be able to analyse the acoustic features of languages we need to collect speech samples of hundreds of languages. For that purpose we created a speech sample database (Mąsior et al. 2013: 79–87; see Fig. 1).
Gathering speech recordings of appropriate quality and length is not an easy task. The results of all of the acoustic analyses can be relevant only if the recording time for each language is sufficiently long (at least tens of hours). The webpage of Global Recording Network organisation (http://globalrecordings.net, DOA: 10 Jan. 2014) is the main source of the recordings collected. The GRN is a provider of Christian evangelistic and discipleship audio-visual materials to the least explored language groups of the world. Moreover, the recordings are collected by the project’s webpage (http://speechsamples.agh.edu.pl, DOA: 10 Jan. 2014) according to the Web 2.0 paradigm (Fig. 1). We have also obtained data from other speech corpora and extracted speech signals from recordings available on the Internet (online radio and audio books).

Currently, the speech recording database contains recordings for 3563 spoken languages (a total of about 7500 hours). More precise statistics are shown in Tables 1, 2 and 3.

Figure 1. The research website (http://speechsamples.agh.edu.pl) interface
Table 1. *Statistics of speech samples database*

<table>
<thead>
<tr>
<th>Area</th>
<th>Languages</th>
<th>Time of recordings [h]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>1241</td>
<td>2618</td>
</tr>
<tr>
<td>Asia</td>
<td>1074</td>
<td>2437</td>
</tr>
<tr>
<td>Europe</td>
<td>77</td>
<td>452</td>
</tr>
<tr>
<td>North America</td>
<td>357</td>
<td>432</td>
</tr>
<tr>
<td>Pacific</td>
<td>553</td>
<td>1100</td>
</tr>
<tr>
<td>South America</td>
<td>239</td>
<td>423</td>
</tr>
</tbody>
</table>

Table 2. *Length, in hours, of collected speech samples*

<table>
<thead>
<tr>
<th>Area</th>
<th>Languages recordings length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt; 1 h</td>
</tr>
<tr>
<td>Africa</td>
<td>560</td>
</tr>
<tr>
<td>Asia</td>
<td>558</td>
</tr>
<tr>
<td>Europe</td>
<td>48</td>
</tr>
<tr>
<td>North America</td>
<td>135</td>
</tr>
<tr>
<td>Pacific</td>
<td>288</td>
</tr>
<tr>
<td>South America</td>
<td>114</td>
</tr>
</tbody>
</table>

Table 3. *Coverage of languages in the database, existing languages number according to Lewis (2009)*

<table>
<thead>
<tr>
<th>Area</th>
<th>Languages with population &gt; 20 K</th>
<th>Languages with population &gt; 500 K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Existing languages</td>
<td>In database</td>
</tr>
<tr>
<td>Africa</td>
<td>1120</td>
<td>810</td>
</tr>
<tr>
<td>Asia</td>
<td>912</td>
<td>638</td>
</tr>
<tr>
<td>Europe</td>
<td>143</td>
<td>70</td>
</tr>
<tr>
<td>North America</td>
<td>92</td>
<td>69</td>
</tr>
<tr>
<td>Pacific</td>
<td>52</td>
<td>46</td>
</tr>
<tr>
<td>South America</td>
<td>52</td>
<td>35</td>
</tr>
</tbody>
</table>
4. Methods

Many years of linguistic research have generated precise phoneme classifications for hundreds of languages. The first stage of language comparison analysis was based on linguistic data. Phonemic data (Wang et al. 2012) for 574 languages from 6 continents were analysed (Kacprzak et al. 2013). Although roughly 9% of world languages were considered, they covered approximately 50% of the world population. The expected numbers of phonemes in the languages on each continent and their standard deviations were calculated. Concerning the languages on each continent, the distribution of vowels and the number of consonants were analysed. Sound diversity among languages differs depending on the continent. All world languages can be divided into four groups. The greatest phonemic diversity was observed (Kacprzak et al. 2013) in Africa and Europe, next Asia, followed by North America. Oceania and South America which belong to the last group were marked by relatively low phonemic diversity.

Nowadays, all these results and conclusions can be verified by a computer analysis of speech recordings. We should expect different dependencies between speech signals for world languages. Such a computer analysis uses the signal processing method and is based on the acoustic properties of speech.

To provide the acoustic analysis we need to segment speech signals into acoustically uniform pieces that correspond to phonemes. Speech segmentation is a core process in the automatic analysis of phonemes.

The collected recordings are segmented into elementary units (Ziółko et al. 2010: 2234–2237) using spectral methods based on the discrete wavelet transform (Daubechies 1992). The methodology for speech recognition developed during the past years provides accurate segmentation and usable implementations. Each fraction of the speech signal is split by digital filters, low-pass and high-pass. Low frequencies have narrow bandwidths and are investigated with a finer resolution, while high frequencies have wide bandwidths, which results in a lower resolution. It is possible to analyse speech in frequency ranges corresponding to a perceptual scale by applying wavelet packet transform (a generalization of wavelet transform), in which low-pass and high-pass bands are split (see Fig. 2).
Figure 2. Frequency sub-bands obtained from wavelet packet transform with a perceptual scale (Ziółko et al. 2010: 2234–2237)

Figure 3. Speech signal segmentation algorithm
The phoneme segmentation (Fig. 3.) and analysis algorithm is based on the analysis of fractions of speech energy in each frequency band. Phonemes can be described as quasi-stationary processes. The role of the segmentation algorithm is to detect significant transitions of energy among frequency sub-bands. Then, qualitative and quantitative analyses of the extracted phoneme features are performed.

Parameterization of phonemes is obtained by calculation of mean energy of acoustic signal in each of twelve frequency bands, weighted by Hamming window. In that way each phoneme is represented as a twelve dimensional vector.

5. Evaluation and initial results

The first verification of inference possibilities from the results of the fully automatic analysis of speech signals was based on the generation of distances matrix from phoneme clouds, calculated for several dozens of languages. Languages were modelled by a single Gaussian distribution, obtained from parameterised phonemes realisations from the gathered recordings.

The top 85 languages (by recording length) were analysed. The Bhattacharyya distance (Brian and Barnard 1996: 2005–2008) was used for calculating similarities between the Gaussian distributions. The UPGMA (Murtagh 1984: 101–113) method resulted in a dendrogram, which is presented in Fig. 4.

Although all of the clustered results are not very encouraging, a separate branch with mostly Arabic languages (ISO 639-1 codes) can clearly be seen:
- aec – Arabic, Sa’idi Spoken,
- arb – Arabic, Standard,
- arq – Arabic, Algerian Spoken,
- azb – Azerbaijani, South,
- ary – Arabic, Moroccan Spoken.

The preliminary results of such approach are consistent with linguistic knowledge (Lewis 2009). Most of the other languages fall into one big cluster because the distances between them are similar. We expect that it will be possible to obtain a much better language division and hierarchisation using a more precise method of language modelling (Gaussian Mixture Models of language phonemes).
6. Conclusion

In this article we present the motivation for our ongoing research on phonemes of world languages and we explain the core methodology of signal processing used in the research. The precise acoustic analysis of multi-linguistic speech may provide answers to the following question: which phonemes are used in different parts of the world and what are their individual features. The first results of the acoustic comparison between languages show that the adopted approach is justified and might result in interesting conclusions.

Acknowledgements

The project was funded by the National Science Centre allocated on the basis of a decision DEC-2011/03/B/ST7/00442.

References


How language evolved as a backchannel between two feedback loops

Abstract. Language is what makes us human. It is the basis of human knowledge, culture, and society. Despite its importance, how language evolved is still a mystery. Various recent studies suggest that humans developed through a “super-fast” evolutionary process found nowhere else within the animal kingdom. This suggests a discontinuity in the evolutionary process itself. We propose the following model: Humans evolved in a unique evolutionary system consisting of two feedback loops, there being a backchannel between them; the lower loop producing the variations needed for selection in the upper loop to take place. What is meant by the “backchannel” here is a structure enabling the selection of the lower loop to “anticipates” the selection of the upper one. The content of this backchannel is displaced action encoded in narration. We show that not only the human brain and language but also most of the unique human faculties (including theory of mind, episodic memory and the unique human altruism) are adapted almost exclusively to developing the functioning of the backchannel (narration) at a super-fast evolutionary pace.

Keywords: biolinguistics; evolutionary tipping point; extended founder effect; human altruism; human evolution; language evolution; Pullo-Vorenus-Hypothesis.

1. Introduction

“The candidates that we have offered so far have not explained the acquisition of language […] so start afresh and find new models different from what we had so far.” (Piattelli-Palmarini 2012)
Language is what makes us human forming the basis of human knowledge, culture, and society. Despite its importance, however, the evolution of language remains largely a mystery, and has been described as being “the hardest problem in science today” (Christiansen and Kirby 2003). Ground-breaking studies in neuroscience, primatology, anthropology, and cognitive science – based to a large part on new technology (including electronic data processing, DNA analysis, neuroimaging technology) – have revealed data that were never accessible before. These have yet, however, to be brought together and synthesized into a larger picture. Perhaps the time is ripe now to develop from existing evidence an entirely new hypothesis, regarding how language evolved.

The starting point for our hypothesis outlined in this paper is the idea that systems construct their own elements. We will claim that there is something special about human evolution (cf. Lahn et al. 2004) and that a fundamental system change occurred before humans evolved. What is proposed is a discontinuity at the system level that came before the origin of our species and we think of humans and language being not the cause but rather the result of it. Our model thus fuses the question of the evolution of language with that of the origin of our species. In terms of the hypothesis presented here – both of these are the effects of a system change–one that can be referred to as a major transition in evolution (Maynard Smith and Szathmáry 1995).

The system that evolved basically consists of two feedback loops and a backchannel between them. The lower circle produces the variations needed for selection in the upper loop to take place. By “backchannel” we mean a structure whereby the selection of the lower circle “anticipates” the selection of the upper circle by implementing a medium that converts the differences of the upper selection into differences in the selection of the lower circle. The backchannel develops into an efficient medium in which language plays an essential role. We assume that most unique human faculties evolved for – and in favour of – this backchannel and were selected through competition for effective transmission. This needs to be clarified.
2. Background

2.1. Super-Fast Evolution: A sign that there is something special about human evolution

The development of the walking foot, the changes in the anatomy of the hands, the S-curved spine, the development of the human brain, the speech apparatus and cognitive abilities, and the like require so many fundamental changes that within the current established framework it seems close to miraculous that all these developments have occurred in such a short evolutionary period of time. Also, there is genetic evidence revealing gaps in our current understanding: our ancestors in fact having undergone fundamental genetic changes within the last 6 million years (Britten 2010; Hughes et al. 2010; Lahn et al. 2004). The changes found are very specific, indicating there to have been a strong and sustained selective pressure and the development of fundamentally new traits. Genomic regions have also been identified that are conserved in vertebrates in general but in the human lineage have accumulated substitutions at a markedly accelerated rate (Bird et al. 2007). The increased rate of substitutions found for the human lineage suggests that their function may have changed entirely (Burbano et al. 2012; Bush and Lahn 2008; Pollard et al. 2006; Prabhakar et al. 2006). It thus becomes clear that the notion that there is only little genetic difference between humans and chimpanzees (Diamond 1991) is nothing more than a popular myth. The large phenotypic divergence between humans and chimpanzees has been driven mainly by changes in gene regulation rather than by altered protein-coding gene sequences (single base substitution), by duplication (gene amplification), deletion, exchange between intron and exon changes concerning the “frames” (Sibley and Ahlquist 1987). Genetic changes of all of these types are exponentially more powerful than single-base mutations are. Such fundamental genetic changes can only be positively selected if there are possibilities for evolutionary change suggesting the development of fundamental new traits. As Bruce Lahn has declared: “To accomplish so much in so little evolutionary time – a few millions of years – requires a selective process that is perhaps categorically different from the typical processes of acquiring new biological traits.” Since Lahn found the pace of evolution here to have been about 16 times as fast as the development found in New World monkeys, he concluded, “Our study offers the first genetic evidence that humans occupy a unique position in the tree of life” (Lahn et al. 2004). This could be an indication for a discontinuity in the evolutionary process itself. To understand how discontinuities occur
we introduce two concepts: the concept of *major transitions in evolution* (Maynard Smith and Szathmáry 1995) and the concept of the *evolutionary tipping point*.

### 2.2 Major transition in evolution

A discontinuity in the evolutionary process is required for the shift from one working evolutionary system to another. These *major transitions in evolution* (Maynard Smith and Szathmáry 1995) are highly improbable events of strong impact that change how the evolutionary process works; their often involving such matters as the unit of selection, how information is transmitted, how variations emerge (heritability), and the frame of selection (the selection circumstances and the selector). Examples of such transformation are the emergence of the cell nucleus and the development of sexual reproduction. The large-scale acceleration of an evolutionary process is often an indication of a discontinuity of this sort. In taking account of this, one can hypothesize that the super-fast evolution of *hominini* – starting with the complete redesign of the locomotion system – could be an effect of a previously overlooked *major transition in evolution*.

### 2.3 The Evolutionary Tipping Point

The *evolutionary tipping point* (ETP) is the point at which a function becomes positively selected for the first time. Once this point has been reached, it may be relatively easy to explain how a certain trait developed further. Since the ETP is the point at which a function is positively selected for the first time, it is evident that a trait which was selected at the *tipping point* was not shaped by the selection for that particular function before. The ETP has in general the following prerequisites: a) the existing organism (constructed by evolution); b) the environment in question, and c) often though not always, a certain temporary context. At the *tipping point* a completely new function appears, one that generates reproductive advantages for the organism. The first positive selection of a new function that appears at the ETP is not necessarily based on a mutation alone, but can be based on the whole context in which the new function is positively selected. An ETP is triggered, therefore, by a fortuitous mixture of a genetic foundation, of behaviour, of circumstances, and of chance. This is especially crucial for the problem of the evolution of language, because language use is always situated in a specific social context.
3. Language – its structure and its unique function

3.1. Linguistic genius and footprints of evolution

It is widely agreed that language is unique to humans (making “human language” a pleonasm) and that it has a genetic component. To illustrate this point, Chomsky often gives an example of his granddaughter and her pet. Although both grew up in an English-speaking environment, only one of them learned to speak English. This point is also echoed by one of Chomsky’s most ardent opponents, Tomasello (2003): “Everyone agrees that human beings can acquire a natural language only because they are biologically prepared to do so.” The process of language acquisition (at the level of grammar) is quite distinct from learning by imitation, its instead being “rather like theory construction” (Chomsky 2010): The infant identifies certain sounds as language and then uses these data as evidence for a syntactic structure that allows it to generate an infinite variety of expressions. As a result, humans (including young children) can utter sentences they never heard before.

If a trait is complex and has specific features related to some particular function connected to its complex design, it is likely that the trait was the target of a selective process (Pinker 2010). Many features of language ability are so specialized that it seems highly implausible that they evolved for reasons other than for furthering linguistic abilities. This is especially true for powerful innate language-learning mechanisms, such as the ability of neonates to identify patterns of tokens (finding word boundaries) on the basis of statistical probability of the phonetic structure in fluent ongoing speech (Pelucchi et al. 2009). Neonates prefer language to all other acoustic stimuli (Shultz and Vouloumanos 2010) and are fascinated by language more than by anything else (Mehler et al. 2006); even foetuses recognize speech with sufficient clarity for this to influence the melody of their first cry (Mampe et al. 2009).

Most of these abilities seem to be so closely related to language that it would be difficult to imagine that they could have developed for non-linguistic reasons. In addition, it has been possible with the help of certain innovative techniques to detect those “footprints of selection” in the human genome that appear to most likely be connected with language (Przeworski et al. 2000; Bustamante 2010; Enard et al. 2002). In contrast to apes, which – if they are laboriously trained – use language for instrumental purposes (such as fulfilling their needs) humans “just love to talk” (Corballis 2011: 163). This is especially true of the innate language faculty suggesting a strong and lasting selective pressure on linguistic abilities.
3.2. Language did not evolve from animal communication

The common preconception that animal communication is the precursor of language is one of the major obstacles to understanding the evolution of language. Language and animal communication are separate and distinct phenomena that evolved separately and for different evolutionary functions (von Heiseler 2014). The parts of the brain in which the two are coded also differ. Instead, there is evidence that the mirror neuron system found in primates is the precursor of certain parts of the language faculty (Arbib 2005). This makes it likely that action reasoning is an important prerequisite to language. But what is the new exclusive function language evolved for, a function that animal communication systems do not fulfil, and why was it put under such strong evolutionary pressure?

3.3. The unique structure of language and its exclusive function

A trait always only evolves so far that it can fulfil its most challenging function, and in turn to construct a convincing evolutionary story of a trait we need to ask why and how it’s most complex feature evolved. The most complex feature of language is syntax. In its simplest form a verb produces – depending on its valency – slots for different thematic roles (agent, patient, instrument etc.). The verb displays the relationship between the diverse thematic roles (marked by inflection, by word order, by a pre- or a postposition; in sign language: by role-taking, by line of sight, direction of signing the verb, by a generalized preposition etc.). Without the verb there would be no syntax. Yet what does the verb refer to? It generally represents an action. Syntax, and consequently language, is thus adapted to describe action. In turn there is no way to describe a displaced action or an absent event without use of syntax: A single utterance will always be interpreted as information about the present: if someone cries out “Fire!” she does not mean that the Bibliotheca of Alexandria burned down more than two thousand years ago, but that there is a fire here and now. Thus, the displacement of an action is possible only if an utterance is given a context through other words within a sentence.

From a cognitive point of view, propositions seem to be more fundamental than words. We do not simply put words together to build a sentence, but first we think of an event that we express in words. Words can thus be seen as fission products of propositions. The sentence “An ape grips the grape”, for example, symbolizes a holistic sensorial experience, the perception of an action. The distinctions used are not “in the world” but
are the result of *categorization* that implies a distinction between apes and non-apes, gripping and actions of all other possible sorts and between grapes and all other types of objects. Thus we could say: the perception of an action is *decomposed* by means of syntactic structure. Yet why should an individual decompose reality? We can think of at least two reasons: (a) for the *purpose of reasoning* (to understand “what is going on”), and/or (b) to tell someone about something that happened earlier and out of sight—to *narrate* displaced actions or events. The marvel of language is that it can describe an infinite number of events through use of a limited number of discrete elements (lexemes).

Goodall argues for the most important unique function of language being that of the displacement of action, that is, the communication of events that are not present, pointing out at the same time that “Chimps […] are unable to communicate about things that aren’t present” (2010). Corballis (2011: 113–114), in turn, writes that “grammatical language evolved primarily to enable us to share episodes. […] Language is exquisitely designed to communicate who did what to whom, where, when and why.” This suggests that language is adapted to describe action. We can further infer that the exclusive function is the communication of an absent action, because the existence of a present action could be communicated simply by pointing. “Absent” would mean something that happened in the past and out of sight: *storytelling*. Consequently there are good reasons to believe that language evolved for narration (for more evidence also see: von Heiseler 2014).

There is evidence that primates use a social reasoning system to understand the actions of others (Rizzolatti and Craighero 2004) and that they use this knowledge for social strategies, to build alliances, to manipulate conspecifics, and so on (Humphrey 1976; de Waal 1982; Byrne and Whiten 1988; Dunbar 2003). Yet why should an individual tell anyone about anything? Understanding *why* it could be a benefit to refer to something that happened earlier and out of sight could be the key to understanding the origin of language. Why did storytelling play such a prominent role in our evolution? All that would be needed is that a narrative sentence is uttered and would give the sender a reproductive advantage. The key to language evolution is therefore to find a scenario in which a narrative sentence would be strongly positively selected. This would mean that the first utterance (no matter how primitive) that referred to a past action conferring an evolutionary advantage on the sender would probably start an escalating evolutionary process, as a result of which the narrative ability would evolve even further. Since the understanding of action implies imitation and internalization (suppression of the physical imitation reflex) (Rizzolatti *et al.*
the main challenge of signing a verb is therefore not the signing itself, but remembering a past action at the particular moment of language use. The challenge for the receiver would be to understand a simulation within the framework of a narration (understanding that signs refer to past actions). At the same time, the first narrative utterance that occurs should be a tipping point in evolution and would need to be connected somehow to a major transition in evolution, a point at which the agent (selector) or the object (the unit selected), and/or the transmission of information, changes. Yet why was such an immense selective pressure placed on the narrative abilities? What could be a context in which symbolizing an absent action would confer a reproductive advantage?

4. The model

4.1. Agents and objects of selection

We can classify different forms of evolutionary process according to the agent of selection (selector): natural selection (here defined as a selection by the environment!) and selection by conspecifics. Furthermore, the object of selection (Mayr 1997) can be either the individual (or, from a more abstract perspective, the allele of a gene pool), or – and let’s propose (as a working hypothesis) if the gene pool is closed and group hostility exists – groups. We can now cross cut these four categories with the three basic functions: nutrition, defence, and reproduction, and therefore construct twelve different categories of selection and thus concepts of the evolutionary processes (Table 1).

Without our consideration of a positive feedback loop and from an intuitive, evolutionarily naïve perspective, some aspects described in the second column (groups selected by the environment) – such as cultural skills to utilize the environment: weapons for hunting – appears promising, because the success of humans is closely related to cultural transmission and social achievements including cooperation within groups. This is why it is not surprising that many researchers use unwittingly concepts of this category, while on the other hand they official deny – for good reason – naïve group selection. All this does not suggest that what is denoted by the table’s second column does not exist (that the cultural development of hominini groups and the collaboration of individuals do provide an advantage to the group), but that it can only explain why humans maintain their predominance over other vertebrates (showed e.g. by the fact that humans build and visit zoos,
while other animals are kept in it), but not how their cognitive abilities and their unique altruism evolved in the first place. To suggest anything as the cause of an evolutionary process without explaining how the gene pool of a population changes its allele frequency is a teleological illusion, that is, it confuses cause and effect.

Table 1. The 12 selection scenarios: This table categorizes the concepts of selection that could be used to explain the human evolution: the first two columns refer to natural selection (in our definition: the selector is nature or the environment) and the last two columns recursive selection (selection of hominini by hominini). In the first and the third columns the individual (or its genes) is the object of selection, while in the second and fourth columns the group is the object of selection. All existing theories of language evolution can be categorized using this table.

<table>
<thead>
<tr>
<th>Individuals selected by the environment</th>
<th>Groups selected by the environment</th>
<th>Competition of individuals within a group</th>
<th>Group competition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nutrition</strong></td>
<td></td>
<td>Competition of individuals in a population for food for example for prey and resources</td>
<td>Groups competing for the resources of a habitat</td>
</tr>
<tr>
<td>The ability to find certain food, break it down and digest it; to hunt and to defend the kill against food competitors</td>
<td>Cultural skills to utilize the environment: weapons for hunting and tools to reach or break up food</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Defence</strong></td>
<td></td>
<td>Pecking order conflicts, building tactical alliances within a group.</td>
<td>Bellicose interactions</td>
</tr>
<tr>
<td>Individual defence mechanism against predators, parasites and against environmental influences</td>
<td>Collective defence against predators; cultural transmission: clothing, medicine, construction etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reproduction</strong></td>
<td></td>
<td>Competition of individuals for sexual partners: a) the intrasexual selection: males compete conflictual for the access to females, b) intersexual selection: female choice. Males compete to be chosen.</td>
<td>Integration of the females of one group into another. Fertilization of a female from another group in the course of bellicose interactions.</td>
</tr>
<tr>
<td>To find (to perceive and approach) a mating partner in the environment</td>
<td>To find another group to prevent inbreeding.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Most monkeys and apes live in groups. They developed most of their cognitive abilities for social reasons to compete within groups (original introduced by Humphrey (1976)), thereby triggering a cognitive arms race (de Waal 1982). An action reasoning system and a prototype of episodic memory – necessary for the capacity for narration – are both good candidates for being just the types of abilities that could have developed within the framework of this so-called Machiavellian Intelligence (Byrne and Whiten 1988). As we can see in our table: this already is a recursive
adaptation process, because the selector is not the natural environment, but the other individuals of the group. However, the difference in the speed of this development between apes and hominini suggests that the competition for social intelligence could be the exclusive cause of the development.

Faced with the fact of super-fast evolution and the development of “capabilities that have no parallels in the animal kingdom” (Saxe 2013), it seems plausible to assume that an entirely new function emerged. The pace of the evolutionary development especially suggests a more powerful positive feedback loop. A positive feedback loop could be a recursive adaptation process. This would occur if the agent and the object of selection – the selector and the selected – were the same: if the selective environment of a species were themselves. Under these circumstances, super-fast evolution seems possible. Because they produce their own selective pressure, with every development whatsoever the challenges grow, and the selective pressure tends to stay on the same level – depending on the asymmetry of reproduction of the individual within a group. Our table shows: there are more recursive adaptations possible beside the rank competitions within groups. The two most promising recursive selective scenarios seem to be: group conflicts (groups selected by groups) and sexual selection (an individual selected by another individual). Are they detectable in human evolution?

4.2. Recursive Adaptations

(1) War before Civilisation. There is evidence for war before civilization: most tribe societies engage in warfare, also our closest relatives, the chimpanzees, wage wars (Goodall 1986). The assumption that our ancestors engaged in conflict is also well supported by a broad literature (among others: Darwin 1871; Keith 1948; Dart 1953; Bigelow 1969; Wilson 1975; Hamilton 1975; Van der Dennen 1995; Keeley 1996 and Pinker 2011). The most convincing evidence, however, shows that the recombination of male and female lines (the haplogroups of the Y-chromosome and the haplogroups of mtDNA) can be explained only by bellicose interactions and by the integration of females of the inferior group into the superior group (see 4.4).

(2) Sexual selection. The basis of (inter)sexual selection is that males and females tend to differ in their level of parental investment. The sex

---

1 There are two kinds of sexual selection: intrasexual and intersexual. In intrasexual selection individuals of the same sex compete conflictual for access to members of the opposite sex. In intersexual selection, one individual is chosen by a sexual partner. We will use the term “sexual selection” only to refer to intersexual selection.
with higher parental investment will be choosier, while the sex with lower investment is courting and advertising itself (Trivers 1972). Since usually the parental investment of the female is higher – beginning with the larger gamete and the gestation of the offspring – females often choose their mating partners more carefully than males do (Bateman 1948). Fisher (1930) showed that sexual selection can start by chance and then escalate.\(^2\) In this context traits can evolve that could not be selected “by nature” (by the natural environment). A strong dimorphism (differences between the male and female phenotype) is not a general sign of female choice – as is sometimes misunderstood,\(^3\) but only of a particular type of sexual selection sometimes referred to as handicap principle (Zahavi 1975). The handicap works as a costly and therefore true sign of fitness: The male shows with a costly signal (e.g. with a big train that male blue peafowl develop), that he is fit enough to survive despite the handicap. The key to the success of this mating system is that males will inherit the handicap along with the fitness to survive with it, thereby ensuring them good mating chances, while the females will generally not inherit the handicap but will receive only the genes proven to be fit (by the handicap) from their (handicapped) fathers. Since the handicap will grow only to the point with the optimal cost-benefit-ratio it will find its equilibrium depending on the costs of the handicap, the nature of the female preference and the asymmetry of the male reproductive success. The costs are paid to the environment and the benefits depend on the female choice. The selective pressure through female choice is much more effective and will lead to a faster evolutionary process than

\(^2\) If in a given population a female preference for a certain quality of a male emerges through mutation, then the male with this attribute will gain a reproductive advantage with that female, while not reducing his attractiveness to other females. This male’s average matings with all females plus his exclusive matings with the females who found his appearance desirable will sometimes lead to the favoured quality spreading through the population, because the offspring of the male with the favoured quality will not only inherit the trait from its father but also the female preference for it from its mother. Thus the genetic basis for the female preference is spread piggyback with the desired quality in the evolutionary process.

\(^3\) For example: Deacon (2010) writes: “sexual selection inevitably produces complementary divergence of male and female traits” and takes this as a reason to doubt the relevance of female choice for the development of the most distinctive of human traits as language and other cognitive abilities: “Therefore, accounting for the extravagant complexity of language in terms of sexual selection requires explaining why it lacks this otherwise-ubiquitous mark of extreme sexual dimorphism.” Deacon confuses sexual selection in general with a special type of sexual selection, namely the handicap principle. Only the latter will produce a strong dimorphism. This dimorphism develops in two steps: first the trait (handicap) evolved due to the sexual selection and then this trait gets suppressed in females by natural selection on the level of gene regulation.
any selection by the natural environment and even for social intelligence. The crucial point concerning human evolution is that theoretically anything can be the target of sexual selection. The only prerequisite is the perceptibility of the trait for the choosing female.

Neither group conflicts nor sexual selection alone can explain human evolution. Sexual selection is indeed a very powerful mechanism, but as Fisher (1930) explained: anything could happen, and its outcome is entirely unpredictable (Miller 2000). In contrast the selection between groups could guide the development in a certain direction (towards bellicose competence), but would not be very powerful, because here an evolutionary mechanism is missing on the gene level. Furthermore there would be no motive to engage in the bellicose interactions if they were risky. Likewise the unique cognitive ability, especially language, could not be selected by group conflicts. But what would happen if both mechanisms would intertwine?

4.3. The ideal(ized) scenario

If the two recursive processes – intersexual selection and the selection between groups – interlock, then a dynamic system emerges in which sexual selection creates random variations on a group level in a runaway process, and these variations then get selected through bellicose interaction. Both sexual selection and bellicose interactions can select traits that the environment cannot. To kick off, this doubly recursive process requires the closing of the gene pool (no casual intermingling between groups). The transformation of the gene pool proceeds now at the speed of sexual selection, a speed more similar to that of artificial selection than of natural selection. But since the groups (only if their gene pools are closed) as a whole are objects of selection, cultural elements and group structures can also be selected by and for bellicose interaction. Both the speed of the development and the kind of traits that can be developed change fundamentally. All this allows us to categorize the emergence of this unique evolutionary scenario as a major transition in evolution (Maynard Smith and Szathmáry 1995), alongside, for example, the emergence of sexual reproduction or even the genesis of eukaryotes.

4.3.1. The Extended Founder Effect

The evolutionary process depends on the production of variations between the selected entities. If groups are the target of selection, then the speed of the process depends on the ability of the system to produce differences
on the group level. The speed of development depends on two factors: the sexual selection as a runaway process and the separation of new founder groups from a source group. The latter is based on what we call the extended founder effect (introduced in this paper). The extended founder effect is the hypothesis that in the hominin line the splitting-off of new founder groups from a source group includes – beside what is known as the founder effect\(^4\) – also non-random elements: related individuals of the same sex could join the same group, based on social bonding or kinship in the source group, assortative mating – the choice of mating partners with similar phenotypes (MacDougall and Montgomerie 2003) – could also play a crucial role. The strongest impact, however, would occur when the formation of new founder groups of hominin is influenced by sexual attraction. In other words, the female preference and male traits must fit. If a subgroup of males with strong social bonds were to separate from a bigger group, then the females who were attracted to them would follow them – or vice versa. And since the reproductive success of any trait depends on the female choice for that trait, the most important factor of group formation is the nature of female preference. Most of the features facilitating the formation of non-random founder groups will increase in line with growing cognitive abilities.

4.3.2. Bravery as a Handicap

If different female preferences prevail in different groups, the object of selection in intergroup conflicts is – indirectly – the predominant female choice. This is a two-step-process: first a preference develops by chance changing the allele frequency of the gene pool of a group and then the groups compete for territories. The selection on the group level is a selection between blind and contingent variants. In this process those female choices will survive that will produce groups that replace others. For this system (variations produced by sexual selection and the extended founder effect that get selected by group conflicts) to function effectively, the groups must regularly engage in group conflicts. Since participation in such conflicts requires the risk of self-sacrifice for the good of the group, this leads us to the problem of altruistic behaviour. If there were an imaginary group of the

---

\(^4\) The founder effect occurs when a small number of migrants split off from a larger population. The new founder group always has less variation than the source group – because it carries only a cut out (a random section) of the original gene pool of the source group (James 1970) – and can, by chance, be distinctively different from the parent population from which it derived. In smaller groups, genetic drift can also play a significant role (Mayr 1942).
brave, the individuals of the group would not use an evolutionarily stable strategy (ESS), because an individual member of it who was a coward would manage to benefit from the replacement of other groups without risking his own life.

The super-fast evolutionary process of hominin legitimates the construction of an ideal scenario that would make this extraordinary development possible. The ideal scenario would be that the differences in males’ victory related behaviour in the bellicose interaction are translated into differences in reproduction. Here we face a classic mapping problem: we have differences on one side (difference in the war-like behaviour) and differences on the other (difference in reproduction influenced by female choice), and we need a medium for the transmission of these differences over time and space. This is to say: Behavior in war-like interactions needs to be translated into mating frequency: the females would need to choose their mating partner according to their perception of absent actions. If this system works, bravery would be a “handicap” (an expansive and therefore true sign of fitness (Zahavi 1975)), and would play the similar role as the train of a peacock. One requirement of a handicap is its perceptibility. In this case the “perception” of the male behavior in the war-like interactions is not observed directly by females but only through a channel that transmits differences of the male behaviour in the bellicose interactions to the female consciousness influencing their choice. Also the channel needs to overcome space and time. Space: The bellicose interactions will be in most cases not in a visible distance to the breading area. Time: The bellicose interaction will not be followed by mating instantly. A good medium for the transmission of male behaviour in bellicose interaction would be narration. In other words, females would need to love war heroes as they appear in narrations.

On the system level this would equate with a backchannel. Before the back channel is implemented the sexual selection and the extended founder effect would produce variants on the group level that are blind for the selection in bellicose interactions. Now by the emergence of the back channel the selection of the lower level (female choice) can somehow orientate itself towards the selection of the higher level (the bellicose interactions). This makes an altruistic behavior in form of bravery evolutionary stable. In other words: A behavior that appears to be altruistic becomes the best strategy to spread the genes of the individual. The size of the handicap (the optimal degree of braveness) would depend on the variation in male reproductive success and the value placed on the handicap of bravery in the female choice. The groups therefore compete in the variation of male reproductive success and in the “clearness” of the backchannel.
Since one’s reputation is about one’s exploits of war, the social structure is indirectly (through the production of bravery as a handicap as transmitted by the medium of narration) oriented towards the outer challenges of the group. This is to say: Because the female choice is the most important factor of the development of a group the changing of the gene frequency can orientate indirectly the on the outer challenge.

In the ideal situation, the individual providing the greatest benefits for the superiority of the group would reproduce most often. This ideal scenario obviously never can be achieved because the contribution to a victory of an individual will not always be perceptible and the medium of narration always include interference (noise). The perfect channel, however, would seamlessly connect the behaviour relevant to ensuring victory and reproductive success. In reality even the orientation to heroic deeds could be quite “noisy” (corrupt), because the victory-relevant behaviour is not always visible for anyone – and even if it is, certain crucial qualities can hardly be successfully transmitted by narration. However, the major noise source – decoupling successful war behaviour and reproduction – would lie in the narration itself: If males could make females believe that they are braver than they actually are, it would give them a considerable advantage, because they would not need to put themselves in danger in order to reproduce. Every male will try to find a way to “cheat” and thereby corrupt the system: not in a conscious way, but through the evolutionary logic of variation and selection on the individual level. If a male could propagate his genes without constantly risking his life, this strategy would spread through the population; therefore, self-propaganda would be one of the biggest interference sources in the backchannel (relating bravery in war and reproduction), as it could minimize the costs of bellicose bravery and increase the benefits from female choice, thereby decoupling reproduction from victory-critical behaviour. This form of corruption could escalate because females that would choose linguistic genius would probably give birth to great talkers, which will be disproportionately successful.

All this leads to a dialectical development: The language ability would be on one hand a show-off behaviour that would disconnect reproduction from bravery – and therefore be the main interference in the backchannel – while on the other hand, in the long run, the selection of the great talker would improve the resolution capacity of the channel (how precise a narration about the bellicose interactions can give a picture of what really happened). Paradoxically, the channel in this autopoietic system improves itself by its own noise.

The aggressiveness of a group does not reflect the variation in male reproductive success in general, but only the part of the female choice (which
produces the variation of the male reproductive success) that is guided by the attraction level to bravery. Because the optimal quantity of bravery – as a real risk taken in the bellicose interactions – becomes minimized by the female choice for the great braggadocio, the level of aggressiveness of groups in which this female preference rules will decline, and the groups in which the females care more about the great talkers will avoid other more aggressive groups and may settle in less attractive habitats or even immigrate to uninhabited regions.

There are two feedback loops both depending on the nature of aspects of the female choice. (A) The outer loop includes two selective mechanisms. (A1) The female choice selects the hero how he appears in the medium narration and thereby selects the bellicose talent and the bravery positively. (A2) In the bellicose interactions the talent is selected positively while the bravery is selected negatively. This complex outer loop results in: (1) a regulation of the degree of bravery and (2) a positive selection of the
bellicose skills. (B) In the inner loop the male narrative performance (based on the narrative competence) is selected by the female choice. This is a positive loop limited by the positive variations on the gene level. The outer loop (A) and inner loop interact in two ways: On one hand both loops compete, because the female choice for heroic behavior (as transmitted by narrations) and narrative performance compete; on the other hand the result of the escalation of the inner loop (narrative competence) is the basis of the media (narration) of the outer loop. Every improvement of narrative competence improves the media of narration and thereby the transformation of victory-relevant behavior in the bellicose interaction into reproduction. Because the inner loop selecting the narrative competence is positive, while the outer loop (positive selection of bellicose skills, regulation of bravery) includes a negative selection (bravery in the conflicts between groups), the system itself shifts towards the inner circle. As a result the females will be less and less impressed by bellicose skills and more and more aroused by narrative performance. However, in bellicose interactions the groups compete in the strength of the outer loop. For this four aspects play a part: (1) The asymmetry of the male reproductive success; (2) the quality of the backchannel primarily premised on the narrative skills; (3) the extent to which the female choice is influenced by heroic deeds as they appear in narrations in contrast to other preferences (including narrative performance) and (4) the danger of bellicose interactions depending on the bellicose skills, the culture of a group and relative strength of the competing groups. The strength of the outer loop conforms to the averaging (and evolutionary stable) level of bravery and therefore corresponds with the aggressiveness of the group. Time scale: The outer feedback loop is a regulator that can adapt to the optimal level of bravery and aggressiveness of the group in a few generations. The fastest change concerns the danger of the bellicose interactions depending on the strength of other groups. An expanding superior group will therefore turn more and more aggressive. The calibration of the bravery level can also include personal experience and epigenetic effects. Another result of this system is the escalation of bellicose culture, including weapons, tactic of ambushes, war paint, techniques of signal transmission, methods of synchronisation, practises of motivation (such as war dance), formation of combat units etc. The bellicose skills adapt to the bellicose culture. Furthermore the narrative competence escalates in this system and thereby many heterogeneous sub-skills. This escalation is limited by the positive mutations concerning the narrative competence.

In isolation the system tends to the inner circle because a good narrator does not need to put his live in jeopardy. Because the great narrators will
be more successful than the great heroes, the female choice for the great narrators will be more successful as well. This will reduce the optimal level of bravery. In contrast: in bellicose interactions the groups compete in the strength of the outer circle. Groups governed by female choices that prefer great talkers would be replaced by groups ruled by female choices that prefer the great war hero. This makes it likely that victorious groups by chance develop a culture that blocks a fast decrease of bravery through female preference for good and well performed narration.

4.4. Problem of the reduction of variability and the integration of females

If a dominant group were to displace all other groups and spread throughout all habitats, genetic variability would be extremely reduced: the variability of all populations of all habitats would be reduced to the original variability within the dominant group. Since evolution always depend on genetic variability, the victory of a small group over all others would limit the possibility of further development. Moreover, it is likely that in all groups some beneficial mutations could emerge; all of which would be destroyed by total replacement. In fact, there seems to be a mechanism that not only has a distinct advantage for overall development (selected in competition with developments in other regions) but which also pays off at the level of the individual (or its genes). This is to integrate healthy, young, attractive females from the losing group into the victorious group. The advantage on the genetic level is on the side of the males of the victorious group. For the overall process, however, it is important that all mutations that can be the object of later sexual selection be retained, even if they develop in a losing group. As the female preference may simultaneously be for different characteristics in different places, it would be possible for different qualities to develop in different groups, which then get recombined by bellicose interactions and the appropriation of females from the losing into the winning group. The only prerequisite for the survival of these new qualities introduced by the females of the losing group is for these attributes to be a target of positive selection by female choice (either this preference could be already found in the predominant group – say for narrative abilities – or it could be introduced by the integrated females). Because the narrative skills do not produce as many costs as the brave bellicose performance, the genes concerning the narrative abilities will spread through the dominant group. This recombination is particularly important in terms of the development of complex traits. This is especially true for those concerning the formation
of narrative ability, since this includes various skills on many different levels that can be selected by female choice (see 4.5). Every aspect that produces perceptual differences can be chosen by sexual selection, including linguistic complexity, and narrative clarity (intelligibility of the narration). In the next generation, the backchannel therefore could be improved and differences between groups concerning the clearness of the backchannel could get selected on the group level.

There is evidence for the recombination of male and female lines that lie in our genes. The different structure of the Haplogroups of the Y-Chromosome and the Haplogroups of mtDNA can only be explained by bellicose interactions and by the integration of females of the inferior group into the superior group.

4.5. The Speed of Evolution and the development of the unique human traits

The speed of evolution depends on the production of variation, on the probability of the variation being an improvement, on the number of fields in which improvements are possible, on evolutionary pressure, and on the number of individuals involved in the evolutionary process. Bigger changes will be based on mutations of the gene expression and include insertions and deletions (Britten 2010). But those bigger changes can be positively selected only if there is something to improve. Most traits are already optimized in a long evolutionary history. Furthermore, the problem
with most traits is that any change can be destructive. Narrative abilities, however, can improve on many levels simultaneously and if a certain quality is improved in the next generation, the demand will also improve, while the selective pressure will be constant, depending on the variation of the male reproductive success. There are different fields that could be positive selected if the narrative performance would confer a reproductive advantage such heterogenic qualities as: vocal characteristics, humor, theory of mind, episodic memory and lexicon size could develop. In other words: Narrative skills comprising many widely different sub skills and all of them could be tested by narrative performances. It is likely that the female choice will be oriented around the qualities that produce significant perceptible differences (based on the female’s own cognitive abilities). First females will only be interested in males that fulfil the minimum standard in all perceptible fields (size, symmetry, signals of health, beauty of the gait, status, smell, etc.), than they focus on their preferred field, but would also appreciate significant improvements in all other fields. In the case of *hominini*, positive mutations can be combined through the back flow through the female germline into the surviving group.

Alongside episodic memory, having a *theory of mind* is an important prerequisite for being able to tell a complex story. First, a good storyteller always needs to keep two things in mind: the whole story (his own knowledge of the narration) and what he has related of it so far (the knowledge of the receiver of the story). Second, to understand a story, both the narrator and the receiver need to understand the motivations and beliefs of the hero (because to understand a story means to understand the motives of the hero). The receiver needs to assume that the hero acts according to his knowledge and not according to objective facts. However, the narrator has to make sure not only that he himself understands the motives of the hero, but also that the receiver understands them at every given point in the story. Thus, to understand whether the receiver grasps the story, the narrator needs a second-order theory of mind (i.e., he needs to understand what the receiver believes the hero believes). From this we can conclude that the *theory of mind*, the ability to understand the beliefs and motivations of others, will be put in the adaptation process to narration under a strong selective pressure.

### 4.6. Evolution and cultural development

In our scenario, language is neither simply adapted to the brain, nor the brain simply to language, instead, *both language and brain adapt to storytelling*. Some of the abilities will be useful mostly for language (e.g.
specialized learning algorithms, phonology, syntax, lexicon) others can be used for other social and cultural aspects (e.g. episodic memory, theory of mind, prosody).

After the evolutionary tipping point, – the first narrative utterance that conferred a reproductive advantage – human brains and language are both selected for the beauty of narrations (beauty is here defined as the attractiveness of a narration: the qualities that make the narration pleased and the narrator attractive). Every adaptation of the brain and its cognitive capacities changes the selective pressure on language (as a cultural entity), and every development of language and its use slightly alters the selective pressure on the cognitive capacities and the brain. The human brain structure can be explained on one hand by older features, which developed for other reasons (in the long evolutionary history of animals, vertebrates and apes) and on the other hand by the adaptation to narration, structures specifically developed in our scenario. Likewise, the structure of narrations is based on cognitive capacities and in particular the language abilities, and the cultural transmission including lexicon, syntax, narrative conventions and so on. This is to say, the springboard of this recursive evolution is the pre-existing structures of both brain and language (as part of the culture in a certain group), but will be subdued under the selection for the beauty of narration.

5. Discussion

5.1. Selection on two levels and the closing of the gene pool

Sometimes it is said, that if there are two levels of selection the lower level (e.g. competition between individuals) would be stronger while the higher level (e.g. competition between groups) would be much weaker. But this is misleading because “strong” and “weak” would presuppose that both would work on the same level. A complex trait can never be the result of the group level and the selections are not comparable on a one-dimensional scale. For example: It is likely that the Homo sapiens displaced the Neanderthals in Europe about 35,000 years ago, because Homo sapiens developed a mimetic culture. But it would be impossible that the Homo sapiens developed mimetic culture in order to displace the Neanderthals. That we populated the planet and not Neanderthals could be therefore caused by a selection between species – what makes it a big difference (at least for us). Such an effect can hardly be called “weak”. But “weak” would be also wrong in another way: no complex trait can ever be formed by any selection of the group level – no matter how slow.
These two levels of selection are generally found on the level of the individuals (or its genes) and on the level of species that compete for the same niche. However, in this competition not only the individual fitness plays a role, but also the social structure, the mating system, the level of collaboration etc. The level of collaboration can never be a direct result of the selection between species, but if the replacement process would happen often, it is likely that the surviving species has a beneficial social structure. In the regular evolution the selection on the higher level (on the level of species) is very rare and could only happen in some millions of years. If the individual of a species – as suggested by our model – would not interbreed arbitrarily with other individual of the same species within a habitat, but would build groups that would interbreed only exceptionally, the pace of the selection on the higher level would increase enormously: Every split off of a new founder group would provide the selection on the higher level with a new variation.

This would implicate that the proof that *hominini* in a habitat are not interbreeding arbitrary (but are building groups to do so), would be indispensable for our model to work. The evidence for this special social structure of our ancestors comes from different fields: (1) The comparison of the structure of the human male germline (Y-Chromosome line) and the female germline (mtDNA-line) shows that groups are not intermixing with other tribes arbitrarily and makes it seem likely that groups displace other groups by killing the males and integrating some of the females (see 4.4). (2) The different tribes on the Andaman Islands do not interbreed for thousands of years and keep their very diverse appearance. The same is true for the tribes in New Guinea (Diamond 1991).

### 5.2. How could the transition in evolution emerge–leading to the development of humans?

The starting point of our scenario would be the closing of the gene pool. There is evidence that this change into a new mating system was triggered by a climatic change. While in western and central Africa the tropical rainforest remained, volcanic activities led to the development of a natural barrier, as a result of which the region east of the Great Rift Valley dried out continuously (Coppens 2004). In an open and clear territory consisting of isolated gallery forests surrounded by open grass fields – such as developed increasingly due to climatic shifts – having intercourse with individuals from another population without being discovered appears practically impossible, especially because the territories in the savannah are much more expansive than in the rainforest (e.g. while a group of chimpanzees in the forest has
How language evolved as a backchannel between two feedback loops

a home range from 5 to 40 square kilometres, their territories in the savannahs are 120 to 560 square kilometres). If now in one population the males guard their territory, kill all males that enter it to mate with their females – and at the same time mate with females of other groups – this behaviour will spread through all groups through the male germline or by the replacement of the groups with other strategies. The system that emerges here has two positive feedback circles but no backchannel between them: the selection on the lower level is blind to the selection on the higher level. The lower level (sexual selection) can only indirectly – by try and error – adapt to the selection of the higher level (group conflict). On this stage genes that help males survive in bellicose interactions will spread through every single surviving group. However, female preference cannot identify the good warrior directly.

On a more abstract level, we could say that the target of selection is the predominant female choice of a group (selected in bellicose interactions). A good female choice would be enhanced bipedalism to free the hands for weapons (e.g. rough stones to strike and to throw, clubs and spears from perishable material) and to pick the good hunter. If the female choice fortuitously favours the upright gait or position as part of the courtship behaviour (e.g. the male showing its genital), the upright gait would evolve with maximal speed. This process could be reinforced by the use of sticks in rang competitions within groups. One group making critical progress would replace all others. With this replacement the female preference for the upright gait would spread, and therefore improvements concerning upright locomotion could happen in different groups simultaneously. The best way of choosing the good hunter would probably be to find males attractive that present hunting trophies. If now the females would orientate on the distinctive signals between males they would choose rare hunting trophies. This would make the animals to hunt bigger, harder to hunt and more danger. This is to say: Because female choice is successful when it detects distinguishing (distinction creating) qualities the hunted animal grows bigger and more dangerous in the evolutionary process until the most dangerous animal would be the hominin himself. This all would be possible without any representation or any understanding that a war or hunting trophy signifies an absent action and can be explained just by the logic of the system itself. This already would be a back channel on the system level, even if it would not include any consciousness of an absent deed. The consciousness of the displaced action is therefore a result (and not the basis) of the back channel. The advantage of understanding a trophy as a sign lies in the improving of the detection of fitness and is more flexible than the instinctive preference for a rare trophy. Now, what would be the most effective technique to present
a trophy? A good way would be, to present it in an engaging way, to make some noise to attract intention and maybe repeat a movement showing the killing. This “showing the killing” would be a mimetic gesture, a sign that would refer to a displaced action. This would make it a narrative proposition. If this narrative proposition were positively selected, this would be the tipping point for the evolution of language and the starting point for a selection for narration. The simplest possible narration (SPN) could be:

> I < kill[ed] < this enemy < with a stone <

This miniature narration (SPN) is, we imagine, only one gesture (likely to be repeated) added to the presentation of the war trophy, but the crucial point is that this gesture would signify a past action. In this first signed proposition there is no chronology, because the signing only contains one movement. The agent would be an implication (the narrator himself); the verb would be a mimetic movement (signifying a past deed); the instrument the real weapon and the patient itself: the dead body (the signifier) stands for the living enemy or animal (the signified). If this simplest narration (SPN) is understood and creates a reproductive advantage, the narrative abilities would get into a positive feedback loop and escalate. Because the index, in the form of the trophy, needs to be always present, there is no danger of lying.

Since the understanding of action implies imitation and internalization – suppression of the physical imitation reflex (Rizzolatti et al. 1996) – the main challenge of signing a verb is not the signing itself, but remembering a past action at the particular moment of narration. Thus the major cognitive challenge would not be the imitation itself, but the reference to the past event. To become narration the action must be stored as a memory, which must then be accessed in a narrative situation. On this stage the remembering of the past event is always triggered by the presenting of the trophy. However, an even bigger cognitive challenge lies on the side of the receiver and requires the understanding that the mimetic gesture is representing a past action. With the first narrative utterance that had a reproductive advantage to the sender the existing abilities are put under a selective pressure for narration. After this ETP, the escalation of the system includes the development of all cognitive features that are necessary for a good narration. Another critical point in this development would be the point on which the narrations would circulate within a population. This can occur only if someone else retells someone’s narration. For this to happen there must be a motive to tell a story about someone else.
The motive is obvious in the case of first person narrations, since they valorise the narrator. But why should anyone tell a story of anyone else? The only reason to tell a story of someone else would be if this behaviour would generate a reproductive advantage. This could happen when females would be fascinated by good narration and the narrative skills itself. Therefore the question would be: Why is it a superior female choice (and therefore an evolutionary stable strategy) to choose the male with impressive narrative abilities? A narrative genius would gain a greater reproductive success with the same bellicose performance. This advantage would be massive because while bravery in the bellicose interactions could be costly, the narration is not (but limited by the narrative competence). When narrations are circulating, the burden of proving truth switches from the indication of the trophy to the controlling of a narration through other narrations. In other words, in the I-narration world, a narration can only be told if the narrator can present a trophy. This makes bellicose interactions a competition to acquire war trophies and also makes later conflicts about the possession of the war trophy probable. The change to a system with circulating narrations, in which every narration is controlled by others narration, suggests a more cooperative strategy. When narrations are the exclusive backchannel, the female choice is attracted to great narrators and to the hero as he appears in the circulating narration (as females today are turned on by heroic deeds and entertaining narrative performances). Furthermore not only does the genetic basis of the language ability evolve, but language itself adapts to its narrative function, which in turn gives rise to new challenges for the narrator. The cultural aspects of language develop due to their adaptation to the narrative conventions of a special culture and thereby modify slightly the selective pressure (also depending on the female preference concerning narrations). Thus the challenges could vary between different groups and could therefore produce different adaptations that can later be combined through the female germline. If the female choice would be orientated on distinctive linguistic qualities, it could trigger an escalation of complexity even beyond communicative efficiency. From this we can assume a super-fast evolutionary pace of both the narrative ability and narrative conventions on different levels including syntax.

5.3. Methodological reflection

For heuristic reasons we first developed an ideal scenario. That does not mean we claim humans to have evolved in a perfect or God-given evolution, but that our scenario is simplified and idealized. The method is first to develop
a scenario that would solve the given problems (such as the development of the unique human abilities in a super-fast evolution) and then to examine how it could possibly be implemented—presupposing it would work much messier in reality than in our concept. The background of this simplification is abstract cybernetic modelling (see Wiener 1948). This is to say that in the modelling of a self-regulating system (Foerster 1981) we would first search for possible feedbacks, analyse the logic of transmissions of signals dictated by the nature of the media, etc. (see von Heiseler 2008). The benefit of this perspective is that it can identify differences that would be unrecognizable in direct observation. Two examples: (1) at a certain state in the development of our ancestors our scenario predicts a male acquire a reproductive success by presenting a trophy. This could be a gazelle, a fang of a hippopotamus or a hand of hominini from another group. For a direct observation there is not a big difference between all three trophies. For a cybernetic modelling there would be a vast difference between a hunting trophy and a body part of a hominini, because the war trophies would create an endless positive feedback (as an arms race) and would implement a backchannel between female choice and intergroup conflicts (2) If there would be a directed observer of our early ancestors she probably would find that a main influence on the reproductive success would be based on inner group competition, alliances, status etc. and that even the female choice would mainly be affected by the status of an individual. But once in a while the female choice is influenced by the presenting of war trophies and narrative abilities accompanying the presentation. Here again we detect a positive feedback and realize that this kind of female choice would foundation of an orientation of the social system of a group as a hole on an outer challenge (the intergroup - competition): The lower feedback circle (the female choice) gets informed by the higher feedback circle (group completion). As with all positive feedback there will be an escalation. Thus the cybernetic modelling can predict the development in this case much better than direct observation.

5.4. The consequences

With our scenario we can explain the development of most unique human abilities, such as theory of mind, episodic memory and language: everything that could be selected by selecting a good narrative performance. Furthermore the circulating narrations could be the basis of the unique human altruism: If reputation is based on the circulating narrations, behaviour that could be observed by any possible narrator could
be influenced by the anticipation of the narration about the behaviour. In other words: Acts will be shaped by possible narrations about them. The consequence: A reputation-economy based on narration also could be the basis of the unique human social order.

For further research it could be productive to investigate the importance of narrations for our lives and the cultural development of narration. The essence of an object often relies on its origin (such as the difference between an original piece of art and forgery) and is therefore constructed by a narration. Myths give people an identity, our self-concept is based on our autobiographic narration, fictive and religious narration can give us role models and our reputation is encoded in gossip and other forms of circulating narrations in different media. Everything that is essentially meaningful to us is connected to narrations.

References


How language evolved as a backchannel between two feedback loops


The case for protoconcepts: Why concepts, language, and protolanguage all need protoconcepts

Abstract. Fodor is infamous for his radical conceptual nativism, McDowell likewise well-known for suggesting that concepts extend “all the way out” into the world and arguing against what he calls (per Sellars) The Myth of the Given: the idea that non-conceptual percepts justify conceptual frameworks. One need not go so far as either researcher, however, in allowing merit to their arguments. It seems we are predisposed, from the beginning of our lives, to look at the world in certain ways and not others. The world need not be “fully conceptual” to be never entirely free, for the conceptually minded agent, of conceptual taint. It seems structured remarkably like our concepts are structured because our concepts present it that way, and our concepts present it that way because of predispositions that are substantively innate. The Protoconcept Hypothesis holds that such protoconcepts are onto- and phylogenetically prior to concepts, themselves onto- and phylogenetically prior to (proto-)language. If that is right, then an account of language genesis and evolution requires a corresponding account for concepts and an explication of protolanguage assumes an explication of protoconcepts.

Keywords: concepts, protoconcepts, innateness, Myth of the Given, Kantian spontaneity, systematicity, productivity, evolution.
1. Introduction

They are the building blocks of structured thought, the set of abilities by which we respond to the world we encounter in a coherent and consistent matter. Concepts structure our experience of the world, delivering the world to us as categories and categorical boundaries. At the same time, they generalize away from the particulars of any given moment, context, or (contra Gallese and Lakoff 2005, Berkeley 1999) application, even as they always apply back to specific contexts. They function as they do because they are both stable enough to be applied systematically across unboundedly many contexts and productively combined and recombined in unboundedly many ways, as well as flexible enough to adjust to each new context – seemingly, by definition, at least subtly different from any context that has come before. They are intentional in the sense that they are always directed at something (a concept is always a concept of); and they are always directed at something for someone: a conceptual agent who is ultimately their master, deriving new concepts as needed and discarding old ones as they become irrelevant.

Although his position is not always clear and appears to have shifted over time, Jerry Fodor is known, and not infrequently ridiculed, for his radical nativism: the idea that all or most of our primitive (unstructured) concepts are innate. John McDowell is likewise known for his conceptualism: the idea that experience just is experience to the extent to which it is conceptually structured (a formulation I owe to Michael Beaton) and that the world arrives for us, as conceptual agents, (fully) conceptually structured: “mind is able to

---

1 For purposes of this paper, I am taking categories to be one of the key defining aspects of concepts: i.e., concepts categorize the world for us; but one can also talk about the nature and structure of concepts at least somewhat independently of how they categorize: e.g., their role in inferential reasoning. That is, ”concept” is the more general term, ”category” the more narrowly focused one.

2 Compare the definition of concepts here with the one that Laurence and Margolis (2002: 28), following Fodor, offer: “…sub-sentential mental representations, that is, representations with sub-propositional contents”. I wish largely to avoid using the term ”representation” here (never mind the even more problematic term ”mental representation”) given the way the term is used ubiquitously and almost casually, without being defined. I also wish to avoid any stipulative (non-empirical) link to language or language-like structure for reasons that should become clear.

3 As Eric Laurence and Stephen Margolis note (1999: 62), such primitive concepts include all our lexical concepts. “Fodor’s innateness thesis amounts to the claim that most of our concepts are innate, a result that virtually everyone finds patently absurd” (Kaye 1993: 198) – a point echoed by Laurence and Margolis (2002: 26).

4 Personal communication.
represent world because world, like mind, turns out to be conceptual: reality is itself conceptual” (Koons 2004: 130). McDowell rejects, in the strongest terms, the notion of a non-conceptual “Given” at the level of our percepts that gives rise to our conceptual structures. In a now famous turn of phrase, he writes, “the idea of the Given offers exculpations when we wanted justifications” (McDowell 1996: 8).

The two issues interrelate. Logically it might seem – as it does to McDowell – that only concepts can beget concepts, and that a truly non-conceptual realm cannot serve as foundation to a conceptual one. Likewise it might seem – as it does to Fodor – that the ways we structure our understanding of the world are, from the beginning of our lives, severely constrained and that the new-born mind does not constitute a tabula rasa. Both intuitions are, I think, at heart correct. Certain concept-like predispositions are with us from the beginning, and they carry us through to the end. Throughout our lives, conceptual and non-conceptual do not pull cleanly apart, either when our thoughts are directed “inward” toward our other thoughts or “outward” toward the world.

In keeping with such philosophers as Albert Newen, Andreas Bartels (Newen and Bartels 2007), and Colin Allen (1999), I hold that concepts – in the sense of systematically, productively, compositionally, and intentionally structured thought – are shared with at least a few and possibly more than a few other species. As such, they are prior to language both onto- and phylogenetically. The “silence” of other species in the face of our questioning should not be mistaken for a lack of conceptual sophistication; as Merlin Donald writes (1998: 185), “humans are undoubtedly unique in their spontaneous invention of language and symbols; but, as I have argued elsewhere … our special advantage is more on the production side than the conceptual side of the ledger. [Non-human] animals know much more than they can express.”

On this view, rather than making conceptually structured thought possible, language transforms and extends conceptual abilities: allowing concepts to be more abstract, more richly compositional than ever before, at the same time as making them far easier to share. If this is right, then any account of protolanguage requires an account of concepts to ground it, and any account of concepts requires – to take the right lesson, I believe, from Fodor and McDowell – an account of protoconcepts to ground that.

Protoconcepts are not concepts because they lack several of the usual desiderata of concepts:5 in particular, they are too few in number to be,

---

5 For a list of such desiderata, see e.g. (Parthemore 2014 in press, Chrisley & Parthemore 2007).
of themselves, productive (cf. Evans’ *Generality Constraint*: Evans 1982: 100–104); they are too general – too applicable to every conceivable context – to be open to revision, as I think true concepts require (though see Woodfield 1994); and, being innate, they are not under the agent’s *endogenous control*: Jesse Prinz’s (2004: 197) preferred phrase for Kantian *spontaneity*, by which the conceptual agent is ultimately the master of the concepts and not the concepts the master of the agent. Given the appropriate environment and the appropriate interactions with that environment, a remarkably small set of protoconcepts can and does give rise to the most richly structured conceptual frameworks.

In Section Two, I attempt to sort out Fodor’s nativism and ultimately reject it in favour of a much weaker form of nativism. In Section Three, I address McDowell’s conceptualism and opt for a non-conceptualist or “protoconceptualist” alternative. Section Four explores the relationship between (proto-)concepts and (proto-)language in the context of Fodor’s *Language of Thought Hypothesis*. In Section Five, I summarize the argument for why we need protoconcepts: why, *contra* Vittorio Gallese and George Lakoff (2005), sensorimotor engagement is not enough; what our protoconcepts probably consist of; and how they give rise, through sensorimotor engagement, to fully-fledged concepts and conceptual frameworks. I introduce the *Protoconcept Hypothesis* whereby a small set of innate protoconcepts give rise to most sophisticated of conceptual frameworks. In Section Six I offer my summary and raise the unaskable question: what would our conceptual frameworks and our world be like if our protoconcepts were different?

2. Responding to Fodor: A more modest nativism

In his reassessment – or reinterpretation – of his original Language of Thought Hypothesis (Fodor 1975), Fodor (2008: 131) takes it as “not seriously controversial” that “minds like ours start out with an innate inventory of concepts, of which there are more than none but not more than finitely many.” On the face of it, this is a considerable retreat from his earlier *radical nativism* with its distinctly Chomsky-like feel to it.

---

6 “If we want to be able to take it that the operations of conceptual capacities in experience impinge rationally on our thinking, as they must if they are able to be recognizable as operations of conceptual capacities at all, we must acknowledge that those rational relations fall within the scope of spontaneity” (McDowell 1996: 52).
One can certainly understand reasons for such a retreat. Despite having had its share of enthusiasts (see e.g. Levin and Pinker 1991, who endorse its logic if not its conclusions), radical nativism runs quickly into conceptual problems. As Jesse Prinz writes (2004: 230):

[...] One might wonder how evolution could have endowed us with some particular mental symbol that is predestined to track spatulas. If evolution set aside some symbol for this purpose, then it would have had to anticipate the invention of spatulas [...]. [However] if concepts are individuated by the properties that nomologically control them [as they are on Fodor’s account], then to say that I have an innate spatula concept is not to say that I have some particular mental symbol in my head at birth [...]. It is only to say that I am disposed to enlist some symbol or other to serve as a spatula indicator. In other words, we are not born with spatula symbols; we are born with spatula detecting abilities [...].

The idea that some level of conceptual abilities is innate is much easier to defend. As Stevan Harnad writes (1990: 2), “clearly, no organism is born a blank slate. Some categories are innate.” What goes for “lower” animals goes, one presumes, more so for language-using humans: “infants do not know the grammar of the particular language community they are born into, but they do have some understanding of the conceptual world that the surrounding language users are expressing” (Levin and Pinker 1991: 4) – an idea that finds much resonance in the writings of Colin Trevarthen (see e.g. 2012), for whom infants are born with a rich intellectual endowment, indeed.

Fodor expresses his more modest nativism like this, emphasizing the intimate link between concepts and experience (2008: 145):

The central issue isn’t which concepts are learned, since, if the (emended) LOT 1 argument is right, none of them are. Nor, however, is it which concepts are innate, if an innate concept is one the acquisition of which is independent of experience. Quite likely, there are none of those, either. Rather, the problem is to explain how a creature’s innate endowment... contributes to the acquisition of its conceptual repertoire; that is, how innate endowments contribute to the processes that start with experience and end in concept possession.
The problem with putting experience first is that it raises the following question: to what extent can there be experience without something to structure it? I agree with the conceptualists this far: entirely non-conceptual “experience” is not experience in the usual sense at all. That is why, in (Parthemore and Morse 2010: 297–298), I suggested a circular causal relationship between concepts and experience, rather than a linear one:

Most if not all concepts require experience to give rise to them. Most if not all experience requires concepts to give structure to it. It’s like the chicken-and-the egg problem: which comes first? Logically something must start things off, but it need not be either concepts or experience as we understand them. Caught within our conceptual perspective, we cannot step outside of it: we cannot simply put our concepts or our conceptually structured experience aside. Concept acquisition and application go hand in hand. Acquiring concepts is a process of applying concepts, which may themselves change in the process of acquiring the new concepts.

The model of causality is not linear but circular.7

Fodor’s “emended” account is problematic in other ways – at least if you believe, as I do (Parthemore 2014 in press, 2014, 2013, 2011a; Parthemore and Whitby 2014; Parthemore and Morse 2010), that concepts as things and concepts as abilities are two sides of one coin. For Fodor with his informational atomism (Fodor 1998), concepts are (physically instantiated) symbols – something that does not seem to have gotten revised in LOT2, though the “emended” account is still, I believe, a step in the right direction.

As Harnad says, we do not start out empty handed. Pace Gallese and Lakoff, with their sensorimotor account of concepts (2005) – according to which even our most abstract of concepts are nothing more than specific sensorimotor engagements, with portions of the motor response suppressed8

---

7 Note the seeming contrast with the following: "the principal motivation for the non-conceptual theorist is simply that the color concepts that we possess are a function of the perceptual discriminations that we are capable of making. It is not the case that the perceptual discriminations we are capable of making are a function of the color concepts that we possess" (Bermudez 2007: 61) – this despite my assumption, in this paper, of a generally non-conceptualist perspective.

8 As I note in (2014), "the account is oddly reminiscent of George Berkeley’s discussion… of triangles and his argument that no one has an abstract concept of triangle that is anything more or other than a specific triangle instance (for Berkeley, a mental picture of a triangle). Like Berkeley, Gallese and Lakoff reject the notion of abstract classes in favor of 'concrete' instances."
something more than a capacity for sensorimotor engagement is needed, not just to keep the process going, but to get it started in the first place. That said, that “something more” must work very intimately with the sensorimotor system – to borrow a phrase, “more closely than a hand fits into a glove”.

At the same time, and in contrast to Fodor’s early radical nativism, one almost certainly wants to start with the absolute minimum necessary built in, so as to afford the greatest possible flexibility. The reason why machine consciousness systems are not – despite their creators’ occasional claims – “minimally conscious” (Parthemore and Whitby 2014: 153–155) and why AI programs are not, as of yet, artificially intelligent, is not just that they violate the usual albeit unstated presupposition that consciousness presupposes cognition and cognition presupposes life (cf. Zlatev’s semiotic hierarchy: Zlatev 2009) – if not necessarily life as traditionally conceived, with a particular biological form and evolutionary history.

The problem is that they build too much in. As the frame problem should have convinced us long ago (McCarthy and Hayes 1969), this cannot be the way to go.

3. Responding to McDowell: Protoconceptualism over conceptualism

McDowell’s conceptualism finds its clearest articulation in Mind and World (1996), based on his 1991 John Locke Lectures at Oxford University. Like Fodor and in contrast to someone like Prinz, McDowell is coming more from the rationalist more than the empiricist tradition in philosophy, and his conceptualism – by which the content of perceptual experience just is conceptual – has, in consequence, a distinctly rationalist-inspired feel to it. Note that, for all that he is otherwise Kantian, McDowell is not making a Kantian-type distinction between the experienced life world and mind-independent reality: for McDowell, these are one and the same. In doing so, he sidesteps Kant’s so-called transcendental idealism – really a form of antirealism – without resolving the issues that it raises. McDowell has

---

9 Fodor (2001: 37) describes this with his characteristic humor: “we still don’t have the fabled machine that can make breakfast without burning down the house; or the one that can translate everyday English into everyday Italian; or the one that can summarize texts; or even the one that can learn anything much except statistical generalizations.”
stood accused of idealism (e.g. Morris 2009, Koons 2004\textsuperscript{10}), and not without reason, for his space of reasons is meant to be complete unto itself – self-supporting – looking like the very “frictionless spinning in a void” (1996: 11, with many subsequent references in the text) that he seeks to avoid.

Concerns aside that McDowell may be over-intellectualizing matters, his conceptualism is a problem for anyone who believes – as I do – that all experience, for the conceptual agent, is a mix of the conceptual and the non-conceptual, or who otherwise sees a continuity between conceptual and non-conceptual.\textsuperscript{11} The usual argument for the non-conceptual content of experience – albeit one that Bermudez (2007: 62) believes to be in many ways a mistake – is that experience is far more fine-grained than finite conceptual abilities would seem to allow: e.g., normally sighted persons can reliably distinguish hundreds of thousands if not millions of colours, and yet it strikes many people as odd, to say the least, to suggest that people possess individuable concepts for each of them.

Bermudez would prefer to give equal space to subpersonal non-conceptual mental content, which I think is a mistake; I don’t believe we should be talking about subpersonal mental content at all – which puts me at least somewhat in sympathy with Daniel Hutto (see e.g. Hutto and Myin 2013), who would eschew mental content altogether.\textsuperscript{12} The non-conceptual content I want to talk about – and I do want to call it content – is very much bound up with conceptual content, rather like two sides of a coin. The basic argument is that, when we experience the world, we encounter it – as McDowell correctly maintains – conceptually structured, albeit conceptually structured perhaps only because our concepts shape it to fit that mould. To make sense of that conceptual structure though, one needs – I believe – to assume additional non-conceptual content (or, if you will, less than fully conceptual content\textsuperscript{13}) to glue it together. We catch sight of such content best

\textsuperscript{10} “[For McDowell] world is assimilated to mind: reality is in the space of reasons” (Koons 2004: 130).

\textsuperscript{11} “Content is nonconceptual just if it can be attributed to a subject without ipso facto attributing to that subject mastery of the concepts required to specify it” (Bermudez 2007: 55–56). The notion of non-conceptual content is usually traced back to (Evans 1982), though most contemporary discussion focuses around non-conceptual content of conscious experience, which is not Evans’ focus, which is rather on the non-conceptual content underlying such experience.

\textsuperscript{12} The personal/subpersonal distinction was first expressed that way by Dennett (1969). Roughly speaking, ”personal” is that which is ”for the entire agent”, ”subpersonal” everything below that: i.e., the brain and nervous system processes that make the agent possible.

\textsuperscript{13} Alva Noë is certainly open to such a continuum between the conceptual and the non-conceptual, with a pragmatic line dividing the two (2004: 31): “the understanding of con-
when we wake from a particularly deep sleep or from anaesthesia and, for a few moments, we “see” without our vision resolving itself into shapes; we “hear” without our hearing resolving into sounds. Such experience is still conceptually structured – we are, after all, aware that we are seeing and hearing – but much less so than is normally the case.

The problem is not – as McDowell seems driven to with his neo-idealism – that mind and world are one. It is rather – as I think Kant correctly understood – that mind and world do not cleanly pull apart. The solution, as I see it, is not to embrace McDowell’s strong conceptualism but to take on board a form of it that even a non-conceptualist could love: call it protoconceptualism. Protoconceptualism accepts McDowell’s assault on the Myth of the Given, acknowledges the problem with the (fully) non-conceptual justifying or otherwise giving rise to the conceptual. Provided that concepts do not fully or cleanly pull apart from non-conceptual referents “in the world”, then some amount of conceptual residue will remain “in the world”, and the conceptually untouched Given will, indeed, be – as Sellars and McDowell correctly diagnose – a myth (Parthemore 2011b: 89–91). It is a myth not only over the general course of our lives but from the very beginning.

4. Responding to Fodor: Why there is no language of thought

There is a picture of the brain, available on the Internet, where the brain is divided, not into the usual areas of e.g. frontal lobe, parietal lobe, temporal lobe, etc., but analogously shaped areas labelled “sex”, “I’m gonna google that”, “must. pick. split. ends”, “is this food still good”, “I’m eating it anyway”, etc. It’s easy to slip into the idea that thought just is linguistically structured – either in everyday language or, as Fodor rather more esoterically would have it (2008, 1975) – in a clearly Chomsky-inspired frame of mind – in a special mental “language” he calls mentalese, distinguishable from everyday language because it is completely unambiguous. Indeed, it is not difficult to find philosophers or psychologists – I have people like Sellars (1956), Daniel Davidson (1987), and Zoltan Torey (2009; see esp. 46, 123) in mind – who think that human-style language is necessary to having a mind

cepts is supposed to be a paradigm of personal-level accomplishment. But just as there is no sharp line between the personal and the sub-personal, so there may be no sharp line between the conceptual and the nonconceptual.”
at all and that non-human animals are, in a very Descartes-like way, “mere”
automata.\textsuperscript{14}

Such a picture poses a number of problems, not least that, in prioritizing
language over thought, it puts communication before thinking and the capacity
to communicate prior to the capacity to think – leaving evolution to do all
the work, onto- and phylogenetically. It might appear that agents are sharing
ideas before they even have them.

Then there are some number of people will stubbornly insist, on
introspection, that they can think the most abstract of thoughts without
consciously entertaining words at all: i.e., they “hear” no running inner
narrative (nor, will they say, are they thinking in a series of conscious
images). That many people do claim to “hear” an inner narrative consistently
is not proof, of itself, that language of whatever kind is constitutive of our
thoughts. Indeed, it is hard to imagine what such proof would look like – at
least if taken as an empirical matter rather than logical necessity.

There are other problems. Though rarely making such commitment
explicit, all but the most boldly eliminativist or hardline reductionist
approaches to consciousness seem to take for granted that consciousness
– as opposed to simple awareness (which even single-celled organisms
possess) – is conceptually structured.\textsuperscript{15} Concepts and consciousness seem
depthly intertwined: where one finds consciousness, one finds concepts
and conceptual frameworks; where one finds concepts and conceptual
frameworks, one finds consciousness; and, as Thomas Nagel has reminded us
(1974), we are willing to attribute consciousness, albeit in varying degrees,
to a range of other species. We may differ on whether frogs have it and be
inclined to say that insects do not, but we are pretty certain that bats and,
indeed, most other mammals do.

\textsuperscript{14} There are, of course, intermediate positions between such a position and the one I am
setting out: so e.g. Wacewicz (2011) sees language as more fundamentally transformative to
cognition than I am allowing while accepting that non-human species do have minds and
are not automata. Although Wacewicz defines concepts \textit{just as} lexical concepts and their de-
rivatives, his point is meant to be more than stipulative/ purely terminological, and he would
likely say that one or more of the desiderata I listed at the start of this article are exclusively
human: notably systematicity, productivity, and compositionality, all of which are implied
by the aforementioned Generality Constraint, which he thinks arrives only with language.
Where we agree is that there is no mentalese, nor is thought otherwise linguistically struc-
tured, even for human beings.

\textsuperscript{15} Wacewicz (2011) implicitly acknowledges this relation between concepts and con-
sciousness in conceding that a position like his could be and has been used to argue against
consciousness in non-human agents, though he himself wishes to remain agnostic.
Indeed, concepts and consciousness are alternatively held up as the hallmark of complex and flexible cognition: what separates us and anyone like us from “true” automata. As the field of cognitive zoology (also known as comparative cognition) is increasingly revealing – and winning converts – several, and possibly more than several, non-human species – in particular among the great apes, delphinids, corvids, and parrots – have complex cognitive abilities that previously were thought to be exclusively human: future-oriented cognition, also known as mental time travel (Osvath and Martin-Ordas 2014) – including the ability to plan for future deception (Osvath and Karvonen 2012) and defer exchange (Osvath and Persson 2013); emotional contagion, including synchronized play (Osvath and Sima 2013); sensitivity to the perspective of others, including non-conspecifics (Bugnyar 2011); the construction of tools from materials at hand (Osvath and Karvonen 2012, Osvath 2009); mirror self-recognition (originally described in Gallup 1977); and so on. What seems to set these species apart physiologically is a much higher-than-average brain-to-body-mass ratio, regardless of absolute body size. What sets them apart behaviourally is an ability to adapt to a remarkable range of terrestrial environments, as can be particularly seen with the corvids, whose range of habitats is nearly as broad as humans’.

Fodor comes tantalizingly close (1987) to suggesting that such species are – in the sense that I am talking about concepts (though not his own, more overtly lexically bound one) – conceptual agents. Of course, what he could say is that these species – all of which lack human-style language – nevertheless possess a form of mentalese; but that, I think, would be a mistake. First, it raises the question of why at least some of them do not, in fact, possess human-style language after some fashion (perhaps not vocal, given limitations on their articulatory abilities, but e.g. gestural). Second, the matter is, as noted before, seemingly impervious to empirical determination, resolvable only by attempted logical fiat. Finally – given that language is paradigmatically symbolic – such a move assumes that conceptual cognition is largely if not entirely symbolic: something that many contemporary researchers, notably connectionists (see e.g. van Gelder 1990), dynamic systems theorists (see e.g. van Gelder and Port 1996), and enactivists (see e.g. Hutto and Myin 2013), are pointedly hostile toward. Peter Gärdenfors (2004), with his Conceptual Spaces Theory – on which I base my own Unified Conceptual Space Theory (2014 in press, 2013) – deliberately casts concepts / conceptual cognition as neither intrinsically association based nor symbolic, sitting somewhere in between.

What I think that Fodor should say, instead – along with other researchers inclined toward or insistent upon an inseparability between concepts and
language – is that, rather than making our conceptual abilities possible in the first place, language transforms and extends them, facilitating their increasing abstraction at the same time as their social dissemination. Language is a tool that becomes so essential to us; it gets incorporated into our core self image, so that it is hard for many of us to imagine thoughts without language. On the one hand, as linguistic agents, we seem compelled to use language to discuss our concepts and conceptual frameworks. On the other, we should not therefore leap to the conclusion that language is foundational to the structure of our thoughts. Instead, I suggest that it throws a veil over our conceptual nature – one that we can only partially begin to lift by looking at pre-linguistic infants and non-linguistic yet concept-using species.

5. The case for protoconcepts

With these preliminaries taken care of, I am now ready to set forth the case for protoconcepts, as they relate to both the (onto- and phylogenetic) evolution of mind and of language / language abilities. If any account of language requires an account of protolanguage to ground it – and there are many who would say that it does, even as the means by which to do so differ (see e.g. Tallerman 2007, 2005; Arbib 2005); and if, as I have argued in the previous section, any account of (proto-)language requires an account of concepts to ground it; then any account of conceptual genesis and evolution – in the individual (ontogenetic) or species (phylogenetic) – requires an account of protoconcepts to ground it. This, I think, is taking the right lesson from Fodor and McDowell! Protoconcepts have the sort of nomic (law-like) relations that Fodor wants all concepts to have: i.e., what makes a protoconcept a protoconcept – all that makes it the protoconcept that it is – is that it reliably tracks all and only what it is meant to in the environment; nothing more can be said about it (unlike our ordinary concepts of dogs, laughter, throwing a ball, etc., of which much more can be said than the bald relation between concept and referent). Meanwhile, conscious experience comes not conceptually structured – not in the beginning, and not entirely, ever – but protoconceptually (and non-conceptually) structured.

---

16 For example, language can make our concepts appear more stable than they actually are: just because the lexical label is stable does not mean that the underlying conceptual framework is, as Thomas Kuhn pointed out with respect to the pre- and post-Copernican notions of "planet" (see e.g. Kuhn 1990: 5).
Protoconcepts reflect an innate predisposition to respond to the world in certain ways and not in others – regardless of the ultimate (mind-independent) nature of that world. Their “proof” lies neither in empirical investigation nor in logical fiat but – if I am right – in their explanatory power. If there is an explanatory gap between the fully non-conceptual and the conceptual, either onto- or phylogenetically, then protoconcepts are meant to bridge it.

Remember that there are three things, in particular, that concepts possess and protoconcepts are meant to lack:

– **productivity**: a finite number of concepts are nevertheless sufficient in number to give rise to unboundedly many complex concepts and, at least in the case of linguistically minded agents, propositions;
– **revisability**: concepts are open to (if not in a continuous state of incremental) revision, and also subject to obsolescence;
– **spontaneity**: concepts are substantially (though far from entirely!) under the control of the agents possessing and employing them.

The **Protoconcept Hypothesis** is that, given the appropriate environment and the appropriate interactions with that environment, a very small set of innate protoconcepts (or, if you will, protoconceptual abilities) can give rise to the most richly structured of conceptual frameworks.

Just how many do we need? It would seem as though we are predisposed to encounter the world in terms of (abstract or concrete) objects or entities (relatively stable or even static), (abstract or concrete) happenings (relatively dynamic), and properties of one or the other. So I propose three basic protoconcepts: **proto-objects, proto-happenings, and proto-properties**.\(^{17}\)

In English, **proto-objects** map conveniently to the grammatical category of nouns, **proto-happenings** to verbs, and **proto-properties** to adjectives and adverbs (along with, of course, prepositional phrases that play the role of adjectives or adverbs). Other languages will map things differently, so there are e.g. languages in which nouns and verbs are said to fall into the same grammatical category. This is not a problem as the road from protoconcepts to concepts to language is meant to be a long and far from straightforward one. The point is that every language can express concepts deriving from

---

\(^{17}\) For further details, see (Parthemore 2014 *in press*). I choose these labels to distinguish these entities from objects, happenings, and properties, which are universally lexicalized. Protoconcepts are not usually lexicalized at all; rather, they are presupposed and normally all but entirely out of sight. They are the far more basic core essence behind the universally lexicalized concepts identified by Natural Semantic Metalanguage theory (Wierzbicka 1972).
these three broad categories and – more controversially – no concepts exist that cannot be traced back to one of these three.

Add to these a few derivative protoconcepts. For example, to make sense of proto-objects or proto-happenings, one needs three abstract proto-objects: \textit{proto-time} and \textit{proto-space}, both of which can be described along \textit{proto-dimensions}. Putting this another way, we are born with a basic sensitivity not just to objects, happenings, and properties, but to space and time, along with their basic configuration.

Two types of proto-space are needed: \textit{proto-physical-space} for locating concrete proto-objects and proto-happenings and an analogous \textit{proto-conceptual-space} for locating proto-properties along with abstract proto-objects and proto-happenings. Two contrastive pairs of proto-properties are needed: \textit{proto-static/proto-dynamic} and \textit{proto-abstract/proto-concrete}. All of these are so basic, I believe, that it really is impossible to imagine conceptual thought without them. Finally, some basic connectors are needed for joining these together, breaking them apart, and comparing/contrasting them: including an addition-like operator \((a + b)\), a successor-like operator \((a \rightarrow b)\), an exclusive-or-like operator \((a \text{ XOR } b)\), and an equals-like operator \((a = b)\). The result is a minimally structured logic both far simpler and at the same time far more expressively powerful than existing formal logics.

Briefly, then, how does one get from protoconcepts to fully fledged concepts, taking the developing individual as the most easily observable case? (We have not been around long enough to have much experience of species evolving.) The story begins with sensorimotor engagements but, \textit{contra} Gallese and Lakoff (2005), sensorimotor engagements are not enough. Why not? Because even our most seemingly concrete of concepts (Gallese and Lakoff’s choice of examples is \textit{grasp}) is, in very important ways, different from any specific instance of its application: among other things, this difference is what allows us to grasp both concrete objects (including ones whose shape and nature is unlike any we have seen before) and abstract ideas.\footnote{For further details, see (Parthemore 2014).}

Gärdenfors (2004: 205) describes the problem as one of how the agent generalizes from discrete observations to general principles. As he implies, the movement from protoconcepts to initial concepts to higher-order concepts (concepts of concepts) is largely an inductive-driven process. It is, at the same time – to borrow one of the better ideas of Daniel Dennett (1991) – an iterated process of pattern recognition driven by a notion of salience grounded, most likely, in the survival of the organism. From
patterns to patterns of patterns to patterns of patterns to patterns of patterns that lose themselves within patterns, one moves from concrete sensorimotor engagements to increasingly abstract thought. In the language of Gärdenfors’ Conceptual Spaces Theory (2004), it is the progressive partitioning and re-partitioning of an initially minimally partitioned conceptual space. With each step, the agent steps further back from the present moment and further away from “life in the moment”. The present moment itself, and the agent’s general notion of “moment” as the minimally individuable unit of time, become increasingly stretched out.

6. Asking the unaskable question

In this paper, I have taken what I consider to be the germ of truth in the heart of Fodor’s conceptual nativism and notion of mentalese, and in McDowell’s conceptualism; and I have used it to make a case for an extremely minimal protoconceptual structure – protoconceptual, because it fails the usual criteria for being conceptual. That is, protoconcepts are effectively hard-wired into us as individuals and, by extension, as a species. We could not be the conceptual agents nor the conceptual species we are if we were not predisposed to look at the world in certain ways and not in others. Those innate dispositions could reflect the nature of the mind-independent world, as the natural-kinds philosophers and the metaphysical realists are inclined to have it; or they could simply reflect the nature and limitations of our conceptual abilities – living in a world that, on many accounts, continually outstrips our capacity to explain it, or at least to explain it completely and consistently.

Pretend, for a moment, that we could set our hard-wired protoconceptual nature aside. What if those dispositions did not reflect the nature of the mind-independent world; in which case, what if they were different? What if we did not perceive the world around us in terms of objects or happenings or properties of either but carved it up in some quite different way? If such a thing were indeed possible, then the resulting conceptual frameworks truly would be completely incommensurable with our own. Where there is complete incommensurability there can be, by definition, no mutual understanding.

The so-called Fermi Paradox asks why, given the apparent size of the visible universe and the number of presumably habitable planets, we have not encountered another “intelligent” species – which is to say, with proper deference to the delphinids, corvids, parrots, and great apes, another species with distinctively human-like intelligence and a similar capacity for an
interest in the technological. The more intriguing question might be, having encountered such a species, whether we would even recognize that we had done so.

**Acknowledgments**

The author gratefully acknowledges the financial support and creative intellectual environment of the Centre for Cognitive Semiotics at Lund University, directed by Prof. Göran Sonesson and assisted by Prof. Jordan Zlatev; along with the helpful discussion and criticism received at a seminar presentation on this topic at the Centre for Cognitive Semiotics on 2 October 2014.

**References**


Abstract. The Evolution of Social Communication in Primates, edited by Marco Pina and Natalie Gontier is an important contribution to the current debates on language evolution. The volume includes texts discussing the emergence of basic social skills connected with language, arguments for vocal and gestural protolanguage, and theories of development of symbolic and compositional language in the history of humanity. In this review, I present the structure and content of the book, but also highlight issues that reflect key controversies in this research area, such as the transition from gestures to speech and from simple vocalizations to modern language. I discuss one specific issue, pointing in nonhuman primates and human children, to offer some remarks on theoretical and empirical criteria for using psychological concepts in debates on social communication in primates.

Keywords: language evolution, evolution of social communication, primates, comparative psychology, pointing.

1. Introduction

How did social communication evolve in humans and other primates? To what extent is there evolutionary continuity in the communicative skills of humans and other primates? Assuming that there is continuity, how did
the simple series of vocalizations and manual gestures present in primates transform into compositional, recursive and symbolic human language? Finally, how can scholars investigate these questions? Which methodologies are available and how do they differ from one another? These questions, and more, are the central focus in the anthology *The Evolution of Social Communication in Primates: A Multidisciplinary Approach*, edited by Pina and Gontier (2014). The book is an outgrowth and expansion of the 2012 International Conference, *From Grooming to Speaking: Recent Trends in Social Primatology and Human Ethology*. In contrast to most post-conference publications, which merely provide an agglomeration of presented papers, the editors of this anthology thoughtfully composed the structure and content of their book.

In my discussion below, I do not aim to present a comprehensive summary of the collection (a daunting task, given its breadth). Instead, I bring to light a selection of topics that define the key points of the current debate on language evolution, like the transition from gestures to speech and from simple vocalizations to modern language. The specific issue of pointing in nonhuman primates and human children, while not a central focus of the book, is particularly interesting for theoretical reasons: being illustrative of the problems with research on social skills in non-human primates, it serves as an inspiration for my discussion in the second part of this paper.

*The Evolution of Social Communication in Primates* presents an analysis of language evolution: from the existence of basic social skills, such as joint attention, intentionality, theory of mind, communicative gestures and social learning, through theories of speech evolution which incorporated already existing physical and social ground, to the emergence of full-blown, symbolic and compositional language. The main goal of the editors was to provide an overview of different methodologies and theories on the evolution of language and social communication. The book is divided into four parts.

Part 1 concerns the history of studies on communication in primates, with texts by Blancke (2014) on Lord Monboddo’s *Ourang-Outnag* and by Swart (2014) on the morally ambiguous cross-fostering experiments.

Part 2 considers the elements of social communication, including texts on chimpanzees learning sign language (Jensvold 2014), original theoretical view on the study of communication in primates (Botero 2014), the universality of basic emotions (Gaspar, Esteves and Arriaga 2014), the evolution of joint attention (Racine *et al.* 2014), and ways of understanding and interpreting mental states in psychology (Nagataki 2014).

Part 3 deliberates the transition from mostly gestural communication to modern language. It contains texts on the bodily mimesis origins of speech
Pointing in the right direction...

(Zlatev 2014), and multimodal accounts on language evolution (See 2014; Leavens, Taglialatela and Hopkins 2014).

Part 4 refers to the emergence of modern linguistic structures. We find there texts on evolution of related symbolic mediums, like thinking, language and art (Nolan 2014; Tattersall), the studies of language evolution based on the computational theory (Benítez-Burraco, Mineiro and Castro-Caldas 2014) and experimental simulation approach (Tamariz 2014) and the challenges on evolutionary biolinguistics (Boeckx 2014).

2. Communication: from social to symbolic

While the book brings together a variety of different perspectives, there are some important ideas that are common to all the authors. Besides understanding language as a medium for abstract thought, most of them understand language as a tool for social communication. They admit that language has to evolve in a social environment and, most interestingly, that modern language is based on the structure of a more primitive form of communication, involving both gestural and verbal modalities. Finally, they mostly agree that the evolution of language is methodologically and theoretically associated with the evolution of abstract thinking and a good theory of language evolution should also refer to the emergence of abstract concepts.

The long-running debate about the nature of proto-language in our hominid ancestors is addressed in the chapters by Zlatev (2014), See (2014) and Leavens et al. (2014), who offer a synthetic solution to the older gestural-versus-vocal origin of language debates by embracing a multimodal approach to the evolution of language. By claiming that communication has, from the very beginning, combined both gestures and vocalizations, they avoid “the greatest unsolved problem for gestural protolanguage theories”, as Fitch (2010: 448) puts it:

the problem of how gestural protolanguage transitioned into vocal language. The multi-modal approach endorsed by the authors enables scholars to solve this problem, because there never was a “mute” gestural phase. Communication has always combined manual gestures and vocalizations, and human language today still makes use of both modalities. A relevant evolutionary question that needs to be raised therefore, is not how did manual gestures transition into vocal speech, but how did vocal language become the predominant channel of human language?

The authors point out several reasons for this transition. According to Zlatev (2014: 176), speech is a more economic modality to use in bigger
groups (for counterarguments see Fitch 2010: 445) and speech is considered to be more symbolic and arbitrary, therefore better suited to “differentiate more clearly between an extensive set of concepts, even when their referents are visually similar.” To explain how speech co-evolved with manual signals and eventually became the predominant modality, Leavens et al. (2014), interestingly merge Dunbar’s from grooming to gossip approach (1996) and Corballis’s from hand to mouth approach (2002). According to Corballis, the brains of hominids were prepared for speech production through the long period of controlling intentional manual gestures by the left hemisphere (grooming is usually performed by the right hand). Dunbar (1996) has developed a theory where grooming, which functions as a tool for social bonding, was gradually replaced with vocalizations when group size increased up to the point when it became too time-consuming to socially bond with all group members individually through grooming. To reinforce the connection between producing vocal and manual intentional signals, the authors argue that the attention-getting calls used by apes to share attention with others are an amplified version of grooming.

Moreover, theories of speech evolution have usually struggled with the assumption that vocalizations are inflexible because they are involuntarily produced. But Leavens and coworkers (2014), as well as See (2014), provide convincing examples that many primates, and especially chimpanzees, have more volitional control over their vocal apparatuses than originally assumed, and they can put them to use to acquire attention from an audience. The debate on whether they are actual examples of intentionality remains open. Leavens says yes, while See, using the criteria developed by Tomasello (2008) to understand gestural signals as intentional, says that the same criteria can be used to understand attention-getting calls as being of an intentional nature. Vocal signals can also be learned and are thus not merely “instinctive”. For example, pant hoot calls of wild chimpanzees are group-specific (Crockford et al. 2004). Finally, because there is voluntary control over both intentionally-produced gestures as well as attention-getting calls, there is a factor of choice. Indeed, under experimental conditions, chimpanzees sometimes alternate between vocal and gestural signals to obtain a certain response when it is delayed by experimenters (Leavens et al., 2014).

The explanation of how symbolic and compositional language emerged remains problematic. In the anthology, several authors put forward distinct theories by approaching the problem from within different methodological frameworks. Nolan (2014: 241) states that symbolic art such as cave paintings or symbolic ornaments “preceded, facilitated, and triggered the emergence of language and with it the growth of abstract thought”. She based on the fact
that art as symbolic activity shares much in common with language and thinking. Art, thanks to its semantic-syntactic structures, could be the first stable in time and intersubjective medium for symbolism. Another approach comes from Tamariz (2014), who treats language as a complex adaptive system that evolved via cultural transmission over many generations of speakers. She understands language as a culturally-evolved system that is shaped by its own selective pressures and where different linguistic structures compete to be learned and transmitted among new individuals. This view is interesting because it assumes that language owes its present form to the fact that it is the most convenient form for humans to communicate with each other. Experiments on the pragmatic features of conventional systems (Galantucci 2005; Scott-Phillips et al. 2009) are very informative in extracting the relevant aspects of effective communication. Nonetheless, such research is also risky and tough to interpret, because preferential favoring of linguistic constructs is examined in participants who already have the biological equipment for language and are immersed in modern types of conventional communication. Benítez-Burraco et al. (2014) proposed a different hypothesis. They point out that the uniqueness of modern language is based on its computational features and focus on the evolution of brain structures that are responsible for language processing.

A third major topic in the book concerns an investigation into what the common components of social communication in humans and other primates are. In the field of language origins and evolution as well as social primatology, scholars remain divided on the issue. The selected authors in this anthology mostly agree that we share our ability for social communication with other apes, but they differ in determining demarcation points. Comparative studies (in this book presented by Gaspar et al. (2014) and Jensvold (2014), among others) look for traces of abilities that constitute pre-linguistic social interactions in our nearest cousins. The theories and operationalisation of these skills are mostly borrowed from developmental psychology. For example, Baron-Cohen (1995) distinguished four features that enable humans to participate fully in a social interaction: (1) intentionality detector; (2) eye-direction detector; (3) shared-attention mechanism, and (4) theory of mind mechanism. The question emerged whether these features are universal for humans and are unique to the species.

The great apes are most likely able to understand intentional actions (Call, Hare, Carpenter, and Tomasello, 2004). On the other hand, there is no evidence that they understand beliefs or false beliefs (Kaminski, Call, and Tomasello 2008), which would be the base for ascribing them a theory of mind. There is still a debate as to whether apes can follow eye-
gaze (Okamoto, Tomonaga, Ishii, Kawai, Tanaka, and Matsuzawa 2002) or if they are directed only by head movement and body posture (Tomasello, Hare, B., Lehmann, H., and Call, J. 2007). However, what may be more important, is that they do understand the perception of others; they do know if a conspecific or a human sees something or not (Call and Tomasello 2008). The ability to engage in shared-attention activities is also in dispute.

3. Pointing: a closer look

In the second part of the paper, I would like to take a closer look at the case of pointing in chimpanzees because it seems symptomatic of discussions in the fields of comparative psychology and evolutionary psychology. Pointing is considered a very significant developmental capacity, because it is an indicator of joint attention, a necessary element of referential communication. Canonical pointing is claimed to be a uniquely human state where both parties are aware that they mutually share attention toward a common goal. It seems to be an initial and necessary step to successfully refer to objects in the triadic interaction. Developmentally, these kinds of behaviors emerge early, in the first year of life, and it is claimed to be cross-culturally universal (Butterworth 2003). This has made some nativists argue that children are born with the understanding of intersubjectivity (Trevarthen 2011, discussed in Racine et al. 2014). Many socio-pragmatic theories of word-learning use the fact that children can understand communicative intentions of adults before they start to learn meanings of words (Bloom 2000). If there was convincing evidence of apes actually directing and sharing attention with others, they would have perspective-taking abilities, the basis of communication acts. Overall, gestural theories of language evolution are often based on the assumption that gestures made by modern apes are similar to the practices in hominids, and are prerequisites for learning language (Fitch 2010: 442). These theories have to prove that pointing in apes is used for communication or directing attention and not only ontogenetic ritualization.

What is interesting is that there is disagreement among authors of the book about the ability to understand and produce pointing in apes. It is not surprising because they represent a variety of theoretical and methodological approaches, so they interpret empirical findings differently. For example, for Racine et al. (2014: 133), it is clear that, “great apes use manual gestures, including pointing gestures”. Similarly, Leavens et al. (2014: 182) and Botero (2014: 88) claim that chimpanzees in a captive environment are able
to redirect the attention of others by pointing. On the other hand, Tamariz (2014) convinces us that apes do not point without training and Nagataki (2014: 154) that chimpanzees cannot be involved in joint attentional interactions. In fact, in the field, there is a fierce and complicated debate as to whether chimpanzees, gorillas and macaques are able to truly engage in joint actions.

It is quite undeniable that, in captive environments, apes do make moves similar to pointing when they want food that is unreachable (Hopkins, Russell, McIntyre, and Leavens 2013). It might still be considered ritualized reaching, but apes behave differently adjusting to the attention of a human in the room (Hopkins, Taglialatela, and Leavens 2007), indicating that those moves are audience-oriented. However, theorists such as Tomasello and his group (e.g., Carpenter and Call 2013) claim that pointing in apes is very different from pointing in human children. They refer to the differentiation between imperative pointing, which is strictly instrumental and a goal-oriented activity, mostly to use people as a social tool for providing the chimpanzee with a piece of food, and declarative pointing, which is performed in the desire to show something to somebody, without instrumental cause, just to share attention. While imperative pointing says “I want that! Go there!”, declarative pointing means, “Look at that! Something interesting is going to happen soon!”

There is even more disagreement in the case of pointing recognition in chimpanzees. Hare and Tomasello (2004) showed that chimpanzees are very poor at understanding informative pointing in the collaborative context (where conspecifics or humans showed them the place with hidden food) but they are quite good at recognizing goal-directed movements, when a conspecific or a human ineffectively reaches for the object with hidden food in a competitive context. On the other hand, Hopkins et al. (2013) reports that chimpanzees understand intentions behind the pointing when a human requests a chimpanzee for an object from his/her cage.

Skeptics agree that captive apes might point in an imperative way. There is a report that wild chimpanzees also point to the part of the body where they want to be groomed by conspecifics (Pika and Mitani, 2006). However, there is no convincing evidence that apes point declaratively (or proto-declaratively), only to direct another’s attention to the object or event. It is claimed that imperative pointing is not an indicator of intersubjectivity because it does not require an understanding of mental states and could be successful without it. Imperative pointing is often viewed as an effect of ritualization (Brinck 2003), even when it can be extended to new cases (as in Hopkins et al. 2013). In contrast, declarative pointing has been
seen as a triadic relation, which manifests in pure motivation to share attention. In other words, only this elaborate kind of pointing may be an important indicator of joint attention, which enables us to truly engage in mutual exchanges (Carpenter and Call 2013). Moreover, successful social communication, even non-verbal, needed more than just sharing attention, for instance, the existence of common ground between interlocutors, to solve the problem of ambiguous references as well as the epistemic perspective-taking during interactions (Clark 1996).

Comparative and evolutionary psychologists who try to describe the basis of social communication in primates meet similar problems to developmental psychologists. They have to study very primitive forms of very sophisticated and language-dependent capacities in individuals who practically do not have any language. It raises a methodological problem: which group of behaviors will give us certainty that primates (and newborns) actually understand others’ minds? Scientists want to explain the basis of social communication by ascribing a set of abilities that usually manifest themselves in language (such as referential acts and attention-sharing) to individuals without language. This paradox might be a reason why “most authors are not comfortable attributing these mental faculties to non-human animals” (Botero 2014: 91). However, serious problems are theoretical in nature. How can we define minimal prerequisites of understanding intentions, goal-directed behaviors, communicative acts and beliefs? Is it enough to ascribe a theory of mind to an animal that understands the perception and goals of a conspecific, however, does not understand false beliefs? Is it more in sharing attention than coordinating other’s focus toward a desired goal? Or does the second-order representation about mutual knowledge have to be involved in that process?

An interesting methodological proposition for researchers was given by Tomasello, Carpenter and Liszkowski (2007). As we cannot rely on the simple observation of a behavior to judge what its nature is in terms of social motivation and mental content, we have to indirectly compare different skills as a complementary group. For example, there is no sense to ascribe to a child or an ape the ability to intentionally share attention towards a mutual object if there is no pure motivation for cooperation or knowledge about a mutual cognitive perspective.

The discussion about pointing shows that there is a need to develop at least minimal empirical criteria for basic concepts in the theories of communication. The first step in formulating these criteria is to compare the most influential theories and methodologies in the field. The Evolution of Social Communication is an important contribution to this discussion.
However, the explanation concerning psychological and cultural capacities underlying communication is only a beginning for theories of language evolution. They have to be consistent with the fact that nowadays we use a symbolic and partially arbitrary system of signs with a complicated grammar, which is easily learned by children all over the world. What is the bridge between communication based on iconic and intuitively recognizable gestures that are used to direct the attention of others, and the abstract language that is mostly conventionally connected with the world, and is far more artificial than pre-linguistic communication? What were the selective pressures which contributed to the emergence of that system? These questions are still waiting for answers.

References


The shades of social.
A discussion of The social origins of language,
ed. Daniel Dor, Chris Knight and Jerome Lewis

Abstract. Turning to the social dimension has been an influential trend in recent language evolution literature, as documented by e.g. Dunbar et al. (ed. 2014), Scott-Phillips (2014), or Pina and Gontier (ed. 2014). The social origins of language, edited by Daniel Dor, Chris Knight and Jerome Lewis, is of special interest, because rather than just being part of this trend, it aims to redefine the current discourse in language origins research, making it inclusive and “society first”. Collectively, the twenty-four chapters of this volume make a powerful statement for a broad, incorporative, “everything counts” approach to language evolution. By demonstrating the relevance to language evolution research of a wide variety of social, cultural and cognitive factors, The social origins of language is potentially – and hopefully – a game changing contribution to this field of study.

Keywords: language evolution, evolution of language, language origins, social communication, social signaling, cultural evolution

1. Introduction

Contemporary evolutionary linguistics, or more broadly research into language evolution, can be seen as growing out of the philosophical tradition of big “origins” questions and into pursuing these questions with more down-to-earth, empirical methodologies of cognitive science. For quite some time, however, the weight of interest has been shifting from the cognitive to
the social. The achievement of the cognitive turn was to take the spotlight away from, for example, the anatomical prerequisites for speech to questions regarding the “language faculty” – symbolization, mimesis, theory of mind and other cognitive underpinnings of language. But now, the “social turn” is taking it one step further: not only by putting the ontogeny and phylogeny of those cognitive capacities in a social context, but also by pointing to mechanisms and phenomena irreducible to individual cognition.

Language evolution literature has been addressing and even directly referencing “the social” for some time now\(^1\), but recent – and very recent – publications representing this trend are much more numerous. Examples from 2014 include Dunbar *et al.* (ed 2014), Scott-Phillips (2014), as well as *The evolution of social communication in primates* (ed. Pina and Gontier 2014), which is reviewed in the present volume (Dębska 2014). However, *The social origins of language*, edited by Daniel Dor, Chris Knight and Jerome Lewis, holds a special place in this mix, aspiring – rightly, in my view – to the role of an important programmatic statement. Dor *et al.*’s aim is to redefine the current discourse in language evolution: from “genes first” to “society first”.

### 2. What’s “the social”?

A good place to start is the classic Wittgensteinian insight that all language is necessarily social – at least in a certain minimal way. The famous “beetle-in-the-box” example (Wittgenstein 1987 [1953]: 100) serves as a reminder that language by nature is always public. Since linguistic labels are arbitrary and conventional, their meanings are not fixed by individual fiat but must instead arise dynamically through repeated communicative interaction. Private language is an illusion, and a solipsistic agent cannot be linguistic – the social group comes first.

But what does “social” mean in “The social origins of language”? This question is tackled head on in a couple of chapters, such as those by Sverker Johansson (2014), and N. J. Enfield and Jack Sidnell (2014), but the answer turns out to be tricky and elusive. The twenty four chapters present, in fact, a very diverse range of perspectives that are not easily united by any obvious single common ground. Here, Wittgenstein is helpful again. The reason

---

\(^1\) An early example is *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, ed. Knight *et al.* (2000), a volume growing out of the second edition of the field’s main conference, EVOLANG.
The shades of social...

the volume does not lose its integrity – just like with the Wittgensteinian “games” – is that it contains several distinct but converging senses of the “social”, forming threads that interweave throughout the publication to keep it together.

One way to construe the social origins is as the kind of selection pressures leading to the evolutionary emergence and development of language. This could mean the thesis that language evolved for communication, in the footsteps of Steven Pinker and Paul Bloom (1990) – and pace e.g. Noam Chomsky and his camp, who suggest at times that language could have evolved for cognition, or even for no reason in particular\(^2\). While still influential, the latter remains a minority view in the language evolution circles, and the former is shared by a broad majority – a stance acknowledged (e.g. in the chapter by Johansson [2014]), but not actively defended in the volume.

In a related sense, the selection pressures in question need not be involved in shaping language directly, but rather in forming the necessary preconditions, or preadaptations, for language (c.f. e.g. Donald 1999; Hurford 1999). In this sense, the social thesis would mean that the “prime movers” of the hominin cognitive development were social rather than, for example, ecological. This point is much more nuanced, as it depends on what cognitive capacities one takes as prerequisites for language emergence, and it enjoys less universal support. Notably, Derek Bickerton (e.g. 1998; Calvin and Bickerton 2000) observes that complex sociality is characteristic of many primate species who live in large groups and exercise Machiavellian intelligence to navigate intricate political webs. Even monkeys can possess an impressively advanced social calculus allowing them e.g. to process transitive rank (Seyfarth and Cheney 2001), and a correlation between the group size and neocortex volume is well attested in primates (Dunbar 1993), yet none of this seems to turn their communicative systems into anything resembling language. To Bickerton (1998), the selective “prime movers” behind the evolution of human-like cognition must have come from ecological factors, especially the challenges of extractive foraging on the savannah.

\(^2\) “We know very little about what happens when 1010 neurons are crammed into something the size of a basketball, with further conditions imposed by the specific manner in which this system developed over time. It would be a serious error to suppose that all properties, or the interesting properties of the structures that evolved, can be ‘explained’ in terms of ‘natural selection’.” (Chomsky 1975:59), quoted by Jackendoff (2002: 234).
Bickerton’s work is referenced surprisingly sparingly throughout the volume, and his position is not confronted directly. Rather, the book implies a different conceptualization of cognitive evolution and of the nature of language. The human difference does not necessarily begin with greater generalized “processing power” that would form a foundation for language superimposed on top of it. Rather, the chapters in Dor et al.’s collection stress how the phylogeny of the internal cognitive capacities for language is, from the very beginning, immersed in a rich external social context and dependent on an intricate external social-cultural scaffolding.

3. Social signals and cooperation

A unifying theme central to the entire volume is the question of the honesty of linguistic signals – and its cooperative underpinnings. Research on cooperation has now become a major industry, spanning several disciplines, from intercultural developmental psychology to neuroeconomics. Cooperation is approached from various directions, from phenomenological to behaviouristic, and using diverse methodologies, from theoretical work (e.g. West et al. 2011), to experimentation (usually in the social dilemma games paradigm, e.g. Ledyard 1995), to computational and mathematical modelling (e.g. Nowak 2006).

In language evolution, the cooperative nature of human communication is of absolutely critical importance. As is often emphasized, humans share valuable information with biologically unrelated individuals and do so without incurring any costs (at least, any obvious costs) to back up the honesty of the message. This cooperative dimension is generally recognized in the language evolution circles as a key theoretical challenge, or a “central puzzle” (Fitch 2010), and so it is in the volume3. “The major transition, all of us agree, must have been the establishment within social groups of unprecedentedly co-operative, trusting relationships”, state the editors (Dor et al. 2014: 12). The different authors approach this issue from different angles – trust, honesty, reputation, reciprocity, morality, pro-sociality, normativity, collaboration, altruism – but almost every chapter at least acknowledges this consensus. Many also work out its implications in more detail.

---

3.1. From vocal grooming to reverse dominance

In two chapters aimed at theoretical syntheses, Chris Knight (2014) and Jordan Zlatev (2014) address the origins of distinctly human, cooperative sociality. Zlatev singles out four “major players” – Dunbar’s (1996) vocal grooming account, Deacon’s (1997) social contract account, Tomasello’s (2008) shared intentionality/prosociality account, and Hrdy’s (2009) alloparenting account – and takes those theories to task by comparing their answers to five questions: “Why us and not others?”, “How?”, “When?”, “What kind of social groups?”, and “Development?” Zlatev’s overall conclusion is that all four are not only possible to reconcile with multilevel selection, but in fact highly compatible with multilevel selection mechanisms, including group selection.

Dunbar’s account grows out of one of the most intriguing and important discoveries in language evolution research – the correlation between primate non-visual neocortex volume and group size (also length of the juvenile period: see Joffe 1997) – but Zlatev rightly notes its shortcomings: vocal grooming by itself does not have to lead to humanlike sociality or communication (see also e.g. Johansson 2005: 213). One should also observe that vocal grooming, unlike real grooming, is a cheap signal, so its communicative functioning presupposes rather than explains the existence of a cooperative context (see Power 1998).

Zlatev is relatively sympathetic to Deacon’s “marriage/male provisioning” account, which envisages the males as “forced” to cooperate with the females to provision unusually altricial hominin infants, this cooperation being governed by symbolically mediated social norms. That said, Zlatev’s highlight on the group-selection elements, though not unfounded, seems to be overstated: “male provisioning” is defensible in terms of individual reproductive success of offspring benefitting from greater parental investment4.

Tomasello’s shared intentionality/prosociality account is very well established in the language evolution literature, but Hrdy’s alloparenting proposal – convergent with Tomasello’s in many respects – has recently received growing recognition. Hrdy’s central insight lies in observing that non-human primates can in fact exhibit highly cooperative social patterns: in providing collective, group-level care to offspring. Where humans are special is in combining this pattern of cooperative breeding specific to (some)

---

4 It should also be remembered that Deacon himself considers “male provisioning” merely as a scenario, additional rather than central to his now classic 1997 book.
monkeys with advanced cognition characteristic of the Great Apes. Zlatev observes that although Hrdy’s proposal fails to provide candidate selection pressures behind this unique combination, it identifies the preexisting platform of trust on which intersubjectivity and prosociality can grow; this makes her account not only compatible, but also complementary with that of Tomasello.

In his review chapter, Chris Knight (2014) agrees that symbols require trust, but he offers an insight that goes deeper. Following the argument developed in Camilla Power’s (2014a) chapter, Knight emphasizes that normally in animal communication a small ‘efficacy’ cost of transmitting the information must be accompanied by a large ‘strategic’ cost of proving that the information is true (see e.g. Maynard Smith and Harper 2003). This cannot work for symbolic communication. Language, Knight argues, is a lie not only in the sense of potentially conveying false propositional content, but much more immediately, in each word being a fake: a fiction not certified for its veracity by a large strategic cost. If each symbol had to be proven true, token by token, it would be next to impossible to transcend one-symbol utterances and achieve combinatorial meaning. Hence, social agreement has a much more fundamental role to play in symbolic communication than purely establishing the sign’s reference: first of all, it mandates believing in the sign. Symbolic communication requires taking symbols as true without demanding proof, that is suspending the disbelief which is the default condition in communication by nonhumans.

Central to Knight’s standpoint is the notion of reproductive levelling. In social groups with appreciable reproductive skew, males have incentives to vie for dominance over the rest of the group, as this ensures greater reproductive success. But increasing size of hominin groups and increasing Machiavellian intelligence of the groups’ members create more opportunities for subordinates to form powerful coalitions that could successfully subvert the position of the leader and undermine his (Darwinian) payoff. With successful counter-dominance strategies in place, competing for status only poorly translates into fitness advantages, which leads to groups being more egalitarian and more cooperative. Thus, symbols depend on cooperation, which in turn depends on egalitarian social patterns (not accidentally so characteristic of present-day hunter-gatherers). Knight provides an overview of several theoretical positions on each of those issues, and as the last – chronologically, first – step, he sees the likely origins of egalitarianism in female coalitions. Here he mentions the accounts by Sarah Hrdy and Kristen Hawkes, but sides with the Female Cosmetic Coalitions (FCC) model.
3.2. Female Cosmetic Coalitions

The author of the FCC model, Camilla Power, has contributed two texts to the volume, Chapters 4 and 15. The first one (Power 2014a) serves to set the stage: working from rigorously Darwinian assumptions about cooperation and signalling, she delves deeper into the special status of language, which functions by combining elements that are individually unsubstantiated “fakes”\(^5\). She also views group size – with resulting increased encephalisation and Machiavellian intelligence – as promoting counter-dominance and reverse dominance strategies, which in turn lead to egalitarian social patterns. What is more, as competition for social intelligence favours greater investment in large-brained but altricial offspring, hominin females face intense pressures for mobilizing male provisioning and blocking male philandering.

Those are the foundations of the Female Cosmetic Coalitions (FCC) model, which Power presents in the later chapter (2014b). Its key component is menstruation – given concealed ovulation of the human female, the only visible sign of her reproductive status and imminent fertility (also one of the most central themes in present-day hunter-gatherer cultures). This may provoke conflict both between males – for sexual access – and females – for maintaining the investment of males, who may be tempted to desert their currently non-cycling partners to pair with the cycling females. On Power’s FCC account, hominin females would have taken counter-measures by forming a coalition dedicated to collectively “jamming” the menstrual signal. They would have applied red cosmetics to their bodies, which would have masked their status and (even more importantly) signalled solidarity, forcing the males to seek sexual favours on the terms of the female coalition rather than their own. Cosmetic adornments, originally proto-ritual displays indexically referring to menstrual blood, would then have evolved into more elaborate symbolic rituals standing for female collective denial of sex; such rituals would be the original source of a group’s repertoire of shared fictions.

The chapter by Ian Watts (2014) fleshes out this scenario with archaeological detail. Most evidence in favour of the FCC model has to do

\(^5\) “Signal evolution theory is the main body of theory applied to animal communication. So it is axiomatic that any scientific study of the evolution of language adopts this theoretical approach as starting point. To argue that the evolution of language is a special case to which signalling theory does not apply, we have to explain why not, within that theory’s terms... We are not systematically monitoring for lies or liars. Indeed, language is in a sense built from components which are ‘lies’–figures of speech whose ‘truth’ emerges only on a higher combinatorial level.” (Power 2014a: 50)
with the use of red ochre, arguably the oldest archaeological trace of human activity that can be linked to symbolism of any kind\textsuperscript{6}. Watts points to several facts, such as the kind of ochre used or correlations with encephalisation: the blood-red types of ochre were preferred to other colours, regular ochre use appears to co-occur with the last brain growth spurt in the Middle Stone Age, and finally, in Europe ochre use seems to follow a pattern of being less intense in the periods where ecological conditions produced birth seasonality (which would have reduced opportunities for male philandering, so that female coalitions would have had less demand for cosmetic rituals).

Watts’ review of the evidence is admirably detailed and informative, but it also remains a sobering reminder of the limits of evidential power of the archaeological and fossil records, which can provide only most general grounding to the more advanced human origin models, such as the FCC. Inferences, when not disappointingly cautious, quickly become far-fetched: there is very little that fossil and archaeological record can legitimately tell us on the exact shape of Middle Stone Age ritual practices or social norms\textsuperscript{7}. A good example of fallible reasoning is when Watts takes the postulated increase in brain size in the MSA to mean that modern language was not yet present at that time (\textit{contra} what Dediu and Levinson [2014] have to say in their chapter). But why would language have put a stop to pressures for greater brainpower? This runs counter to the logic of authors such as Dor and Jablonka (2014), whereby culture, fuelled by increasingly effective symbolic communication, installs a feedback loop working to promote, not suppress, encephalisation.

Overall, the FCC is a complex model with many merits, such as accounting for the deep symbolic meaning of menstruation in hunter-gatherer societies,

\textsuperscript{6} A number of more ancient artefacts exist, including the recently reported engraved fossil shell dating back to about 500 000 years ago (Joordens \textit{et al.} 2014), whose interpretation, however, is much more controversial.

\textsuperscript{7} “There seem to have been significant obstacles to establishing a form of social organization that allowed some individuals to remain at a home base, indefinitely maintaining captured wildfire. Speculatively, these may have concerned an absence of mechanisms of punishment and reward to enforce cooperation between strangers, a prerequisite of ‘institutional facts’.” (Watts 2014: 215)

“Intriguingly, cut-marks on bone are more abundant and randomly oriented than in later Levantine contexts, suggesting that more individuals were involved, with little or no formal apportioning of meat (Stiner \textit{et al.} 2009). This might be read as another indication of an absence or weak development of institutional facts.” (Watts 2014: 215)

“The Wonderwerk pigments strongly suggest collective ritual dating back to 300–500 kya, with fire-lit ‘song-and-dance’ performances, the dancers glowing red and glittering.” (Watts 2014: 217)
and emphasizing the role of costly ritual in establishing social norms. That said, the repeated assertions of testability are overstated. The predictions it generates fit empirical data in ways that are appealing, but post-factum and vulnerable to confirmation bias; also, as mentioned above, archaeological data by their nature allow for only loose fit. The FCC remains an interesting contender, but one must remember that each step of this elaborate and multi-step scenario needs to be treated as somewhat speculative.

3.3. Why talk? The concept of risk-free killing

Yet another perspective on the topic of trust and evolutionary stability of communication comes from Jean-Louis Dessalles (2014). He formulates this problem with rare succinctness: “If information has any value, it is in the interest of no one to give it for free. And if information has no value, why are there ears ready to listen to it?” (Dessalles 2014: 284). Here, Dessalles’ principal interest lies in the “talker” rather than the “listener”. While most theorizing focuses on the receiver’s risks of being manipulated, Dessalles, admirably, asks us to consider the reverse question: what benefits does the sender derive from producing the message, given that consistent deception is not an option if communication is to be evolutionarily stable? In other words, why talk?

Dessalles points to the basic facts of human ethology, citing Geoffrey Miller’s (2000) observation that where we compete is to speak rather than to listen. Humans find talking inherently gratifying, which suggests it must have had adaptive value but does not tell us its exact nature. To Dessalles, the Darwinian answer is that using language is competitive signaling. Speakers want to show off, and they accomplish this by making their conversational turns relevant and imbuing them with highly desirable qualities: unexpected information, emotional content, and logical consistency. Even if the information conveyed is trivial rather than important (as is commonly the case in conversation), the speaker whose contributions meet the above criteria demonstrates cognitive characteristics which would make him a valued ally in the human EEA (Environment of Evolutionary Adaptedness).

To explain this last point, Dessalles turns the reader’s attention to the phenomenon of risk-free killing. The main technological innovation in early hominins were stone tools, and one consequence of being able to

---

8 Fitch (2010) proposes a term derived from German, *Mitteilungsbedürfnis*, to capture the human urge to share information and emotions with others.
use a stone tool effectively is being able to use it as a weapon. Assassination becomes an option. Unarmed, killing a competitor requires having enough strength and social support to overcome not just the opponent but also any individuals that might come to his rescue; even then the encounter may prove risky and costly. Weapons change this dramatically, as almost everyone becomes capable of dealing a lethal blow to a sleeping rival. The only protection, Dessalles notes, is having vigilant allies – and vigilance is best advertised in precisely those conversational contributions that are relevant and logical, and have unexpected content and emotional impact.

In my opinion, the concept of risk-free killing is the stronger part of Dessalles’ scenario, and potentially a key idea in hominin evolution. Weapons, and the decreased costs of killing dominant individuals that came with them, would have had two related consequences, both very important. Firstly, social hierarchies would have depended even less on physical power and even more on social skills, which would have added fuel to the arms race of Machiavellian intelligence (with all the consequences to brain size, neonate altriciality, and group organization for infant provisioning). Secondly, the possibility of (almost) everyone to have in check (almost) everyone else would have constituted a powerful drive for social levelling. Those two points deserve closer attention from scholars whose models prioritise the social brain as well as counter-dominance and reverse dominance. Dessalles’ account, in turn, would be more complete if he factored in the possible pressures from inter-group conflict: killing a member of one’s own to replace him in hierarchy pays off immediately, but has an equally immediate cost of making the group more vulnerable to attacks from rival bands, resulting in its physical as well as genetic obliteration.

4. Culture

The role of culture in language emergence could be seen as another common thread, even though it may be better understood as not one but two distinct insights. Firstly, we have the cumulative potential of the cultural mechanisms of invention and innovation, most successfully captured in the ratchet effect analogy (e.g. Tomasello 1999). Innovation in a group would normally be lost after each generation, but cultural transmission acts as a pawl preventing the reset. By culturally inheriting the innovation, the new generation no longer has to start from scratch, and rather than devoting their lifetimes to reinventing the state of the art, they can “pick it up from there” and devote those lifetimes to advancing it further on.
Secondly, we have the phenomenon of *gene-culture co-evolution*, first introduced into language evolution by Terrence Deacon, whose work is widely acknowledged throughout the volume. Deacon (1997) illustrated his case with the example of the human adaptation for post-weaning lactose tolerance, evolved along the Baldwinian lines. With culturally invented and transmitted technology of milk processing, such as fermenting or making cheese, milk would slowly become available as a food source to human adults in certain populations; this would increase their reliance on milk and dairy and result in a selection pressure for inborn lactose tolerance, leading to an even greater dependence on milk products and thus stimulating further technological progress in a gene-culture positive feedback loop.

Daniel Dor and Eva Jablonka (2014) offer an analogy of their own: the latent skill of *echolocation*. In humans it manifests itself to an appreciable degree only under rare circumstances, yet in a fantastic but at least conceptually possible scenario, in an isolated population of blind people this skill could be useful enough to make a difference. Not only could we expect Darwinian selection for it, but also that the development of this skill in ontogeny would be encouraged, and that technology and culture would be gradually redefined and “built around it”, to assist and promote it on the one hand and exploit it on the other; all this would culminate in increasing natural selection to close the positive feedback loop. Likewise, language evolution is “phenotype-first”, fuelled by developmental *plasticity* that is magnified by cultural-technological progress.

The step further that Dor and Jablonka want us to take, and the novelty of their proposal, is that the phenotype in “phenotype-first” is *cultural* rather than individual. The way language is different from lactose tolerance or echolocation is that it inevitably transcends a single person. It is no use for a lone individual – language necessarily means interaction between several minds, implying not just individual, but *collective* creativity. By its own nature, language involves *group-level* innovation. Each communication act unavoidably taps the advantages of brainstorming, leading to gains that exceed the sum of individual contributions: “group genius” over “lone genius”. Collaborative innovation drives fast change in communicative technology, which can then be recursively applied to itself: as tools may be used to make better tools, emerging language becomes a means of promoting even more trust, collaboration and creativity.

Several other texts in the volume look into the role of cultural transmission, to a various extent. For example, Enfield (2014) looks into *cultural epidemiology*, specifically aiming at a more principled account of *transmission biases* (roughly equivalent to selection pressures
in biological evolution). He puts forward a model of a four-stage cycle of cultural transmission: exposure (of an individual to the cultural “meme”) – representation (encoding the “meme” in the memory) – reproduction (when the individual retrieves and uses the “meme”) – material (the physical properties of the external representation); each stage involves a set of characteristic transmission biases. Enfield emphasizes the role of iterated practice in language, which goes beyond the standard iterated learning approach by closing up on the *enchronic* timescale: “the timescale of moves and countermoves in sequences of human interaction” (2014: 327).

5. What is language?

“Your theory of language evolution depends on your theory of language”, says Ray Jackendoff (2010). The debate between the proponents of the narrow delineation of what can legitimately be called language and the advocates of a broader, integrative view (Hauser *et al.* 2002, Pinker and Jackendoff 2005; see also Wacewicz 2012, and Wacewicz and Żywiczyński, this volume) has been alive for more than a decade, and it continues to shape the controversies in the field. It may well be that the most important contribution of *The social origins of language* is not arguing for anything social *per se*, but rather making a consistent, convincing and powerful statement about the “broad” nature of language, and as a result championing the need for an all-inclusive perspective on its study. The introductory chapter, remarkable for its apt phrasing of many points, puts it eloquently:

> With the rise of a multitude of new sub-disciplines, specialized journals, and conferences, and with the gradual decline of the Chomskian paradigm as a unifying framework, more and more of what we learn about language remains confined to specialized professional circles. However, to understand the origin of language requires a move in the opposite direction—a large-scale, collective interdisciplinary effort at theoretical synthesis. The detective-like analysis of circumstantial evidence knows no disciplinary borders. Everything counts. (Dor *et al.* 2014: 1–2)

…a very wide spectrum of entangled conditions is required—cultural, social, political, cognitive, and emotional. In other words, language is an internal component of a much wider continuum: social intercourse and culture in distinctively human form. This, then, is why the problem is so difficult: to explain language, we
Language is not just a biological adaptation. The conceptualization of the language faculty only in terms of the organism’s genetic make-up is incomplete, and Chris Sinha (2014) explains why by usefully analogizing language to animal artefactual niches, such as the bower in bowerbirds. As products of behavioral instructions largely coded in the organisms’ genes, artefacts are organisms’ creations – but the reverse is also true. Animals such as bowerbirds, beavers or termites are profoundly affected by their artefacts, which literally form the animals’ niches, indispensable for their daily functioning, survival and reproduction, and equally heritable as genes. This analogy argues against taking biological and cultural evolution separately, and for treating the evolving entities as biocultural complexes: “phenogenotypes”.

Another of the volume’s excellent analogies is found in Charles Whitehead’s (2014) chapter, who – building on the insights of Émile Durkheim and Chris Knight – notes that “…any attempt to explain language as an isolated trait is akin to explaining the emergence of the credit card without considering the preconditions on which credit cards depend—including commerce, money, banking, the digital computer, and the means to detect and punish fraud” (Whitehead 2014: 157). Accordingly, approaches that focus selectively on the formal combinatorial properties of the human language faculty can be compared to studying the structural and material properties of the plastic plate. This is not to question their validity or usefulness; problems begin when such a narrow approach is claimed to exhaust the scope of what is interesting or possible to know about language and its origins (cf. Wacewicz and Żywiczyński, this volume).

And what is interesting about language clearly cannot be narrowly limited to grammar or, for example, vocally transmitted information. In his chapter, Adam Kendon (2014) argues compellingly that extra-oral visible bodily action – communicative movements of the hand and arm, but also a variety of other signals such as movements of the trunk or head, or facial action – is an integral, essential and inalienable component of language9.

---

9 “Extra-oral visible bodily action, in short, is deployed by speakers in ways that serve a great diversity of semantic functions but which yet, if done while the person is speaking, are so closely co-ordinated with this activity that it is to be understood as an integrated component of it.” (Kendon 2014: 68)
Extra-oral visible bodily actions, rather than being random “supplements [or] add-ons” to vocal signals, are patterned (rule governed), semantic (individually meaningful or related to the meanings expressed vocally), and very deeply integrated with speech. As a result, we should treat speech and the various visible expression systems as an ensemble, and “[a]ny theory of the evolutionary origins of language needs to take [extra-oral visible bodily action] into account” (2014: 71)\(^{10}\).

Kendon calls our attention to ways in which human linguistic activity is not only multimodal, but also depends extensively on semiotic means other than symbolic signs. Jerome Lewis (2014) does the same in discussing multimodal communication in the Central African BaYaka Pygmy groups, modern hunter-gatherer representatives of the most ancient genetic lineages of humankind. Their daily communicative practices consist in versatile drawing on a variety of means of expression, from words, to reenactment, to different forms of music (themselves forming a continuum with speaking), to sonic mimesis: mimicked sounds that can be used flexibly in conversation to substitute conventional lexical labels. Lewis highlights how different this is from the default assumptions about the nature of language, especially those that present-day Western researchers, with their lifetime immersion in the written word, unwittingly bring to the study of language origins:

Mbendjele and other Pygmy groups’ multi-modal communicative strategies targeting different audiences remind us of the environmentally embedded context of language use likely to have dominated in the past. BaYaka seek to speak as many ‘languages’ (djoki) as they can. Their speech is incorporative, open, encompassing, and inclusive. It is a skilful multi-modal deployment of a range of capacities inherent to human bodies that serve to establish relationships with as many creatures as possible. By contrast, most language users today think of languages as conceptually fixed to a distinctive vocabulary, grammar, and speech style, facilitating interaction between members of a particular human group, and as being political by being selective, exclusive, and oppositional. (Lewis 2014: 85)

\(^{10}\) Some may find it surprising that Kendon does not support any of the “gesture first” theories of language origins (e.g. Corballis 2002); but this is precisely for the reason just described – emphasis on the visual rather than vocal modality would face the same problem with explaining the deep integration between the modalities.
6. The language mosaic

Overall, the twenty four chapters give us a broad, complex, and multifaceted vision of language evolution: what Hurford (2003) calls *mosaic evolution* or what Knight (personal communication) calls *jigsaw evolution*, as opposed to the *magic ingredient X* evolution scenarios which seek to reduce language origins to a single crucial component. Again, this inclusive position is something that Dor, Knight and Lewis take special effort to emphasize: “On one point, we are all agreed: languages began evolving as a consequence not of one social factor but multiple interacting ones. This, then, is our overall thesis” (2014: 4). As the editors say, “everything counts”, including “shared childcare, the control of fire and cooking, projectile weapons, big game hunting, increasingly equal power relations between the sexes, emotional bonding through music, dance and other forms of ritual—and, as a consequence of increased trust within relatively stable coalitions, steadily increasing chances for cultural innovations to be preserved and transmitted to future generations.” (2014: 4)

It is impossible to do justice here to all twenty four texts in the volume. In the chapters not discussed above, Sverker Johansson (2014) looks into accessible sources of evidence in which to ground a language origins theory; Nick Enfield and Jack Sidnell (2014) demonstrate that one such source may be Conversation Analysis; Daniel Dor (2014) develops the view of language as a *communication technology* (introduced in the chapter co-authored with Jablonka); Simeone Pika (2014) reviews a wealth of information on visual communication in apes, with emphasis on chimpanzee gestures; Zanna Clay and Klaus Zuberbühler (2014), in contrast, focus on ape vocalisations, warning against writing them off as uninteresting emotional reactions; Dediu and Levinson (2014) overview a broad range of archaeological, paleoanthropological and linguistic evidence, all consistent with modern language being present as early as 500 000 years ago; Emily Wyman (2014) is interested in language as a tool for doing things, and in the social prerequisites for enabling such performativity; this is related to Ehud Lamm’s (2014) interest in *normativity* as such a precondition, but also something that grows together with language in a co-evolutionary dynamics; in another co-evolutionary account, Simona Ginsburg and Eva Jablonka (2014) look into the power of lexical labels as cognitive enhancers, presenting *word-based episodic recall* as a crucial cognitive novelty; Chris Knight and Jerome Lewis (2014) build on their previous chapters, working towards a scenario of language origins from vocal mimesis; and Luc Steels’ (2014) aim is to pull
down what he sees as a number of false dichotomies: between synchrony and
diachrony, nature and culture, competence and performance, processing and
describing, and formalism and functionalism.

Because of the richness of the collection, there are a number of quite
fundamental recurrent threads which I could not address: for example,
the role of ritual (Knight, Lahm, Lewis, Power, Watts, Whitehead), the role
of play and pretend-play (Dor, Ginsburg and Jablonka, Power, Whitehead,
Wyman), and the idea of language as a medium for creating nonphysical
but binding reality of Durkheimian institutional facts (Knight, Power, Watts,
Wyman).

Collectively, the twenty four chapters deliver a clear message: no single
discipline or research trend “owns” the field of language evolution. There
is room there for conversation analysis, cultural anthropology/ethnography,
gesture studies, robotics, and so on. Just as cooperation in hominid groups
must have been a crucial prerequisite and a central piece in the puzzle
of language emergence, cooperation across disciplinary borders will be
a crucial prerequisite for solving it.

References

protolanguage to full human language. In J.R. Hurford, M. Studdert-Kennedy
and C. Knight (eds.). Approaches to the Evolution of Language. Cambridge:
Cambridge University Press.

Chomsky with the human brain. Cambridge, MA: MIT Press.

in chimpanzees and bonobos. In Dor et al. (eds.), 141–156.

Corballis, M. C. (2002). From hand to mouth: The origins of language. Princeton,

the Brain. New York: Norton.

of Social Communication in Primates, (eds.), M. Pina and N. Gontier. Theoria
et Historia Scientiarum 11.

Dediu, D., Levinson, S. (2014). The time frame of the emergence of modern language


In M. C. Corballis and S. E. Lea (Ed.), The descent of mind: psychological


The shades of social...


