

**HUMANA**  **MENTE**  
JOURNAL OF PHILOSOPHICAL STUDIES

# Origin and Evolution of Language

EDITED BY  
Francesco Ferretti and Ines Adornetti

ISSUE 27 - DECEMBER 2014

EDITORIAL MANAGER: DUCCIO MANETTI - UNIVERSITY OF FLORENCE  
EXECUTIVE DIRECTOR: SILVANO ZIPOLI CAIANI - UNIVERSITY OF MILAN  
VICE DIRECTOR: MARCO FENICI - UNIVERSITY OF FLORENCE

## INTERNATIONAL EDITORIAL BOARD

JOHN BELL - UNIVERSITY OF WESTERN ONTARIO  
GIOVANNI BONIOLO - INSTITUTE OF MOLECULAR ONCOLOGY FOUNDATION  
MARIA LUISA DALLA CHIARA - UNIVERSITY OF FLORENCE  
DIMITRI D'ANDREA - UNIVERSITY OF FLORENCE  
BERNARDINO FANTINI - UNIVERSITÉ DE GENÈVE  
LUCIANO FLORIDI - UNIVERSITY OF OXFORD  
MASSIMO INCUSCIO - EUROPEAN LABORATORY FOR NON-LINEAR SPECTROSCOPY  
GEORGE LAKOFF - UNIVERSITY OF CALIFORNIA, BERKELEY  
PAOLO PARRINI - UNIVERSITY OF FLORENCE  
ALBERTO PERUZZI - UNIVERSITY OF FLORENCE  
JEAN PETITOT - CREA, CENTRE DE RECHERCHE EN ÉPISTÉMOLOGIE APPLIQUÉE  
CORRADO SINIGAGLIA - UNIVERSITY OF MILAN  
BAS C. VAN FRAASSEN - SAN FRANCISCO STATE UNIVERSITY

## CONSULTING EDITORS

CARLO GABBANI - UNIVERSITY OF FLORENCE  
ROBERTA LANFREDINI - UNIVERSITY OF FLORENCE  
MARCO SALUCCI - UNIVERSITY OF FLORENCE  
ELENA ACUTI - UNIVERSITY OF FLORENCE  
MATTEO BORRI - UNIVERSITÉ DE GENÈVE  
ROBERTO CIUNI - UNIVERSITY OF BOCHUM

# TABLE OF CONTENTS

## INTRODUCTION

Francesco Ferretti, Ines Adornetti <i>Origin and Evolution of Language: a Close Look at Human Nature</i>	III
---	-----

## PAPERS

Lluís Barceló-Coblijn <i>On the Relationship between the Air Sacs Loss in the Genus Homo and Duality of Patterning</i>	1
Teresa Bejarano <i>From Holophrase to Syntax: Intonation and the Victory of Voice over Gesture</i>	21
Michael C. Corballis <i>The Gradual Evolution of Language</i>	37
Erica Cosentino <i>Embodied Pragmatics and the Evolution of Language</i>	61
Peter Gärdenfors <i>The Evolution of Sentential Structure</i>	79
José Luis Guijarro <i>The Origin of Languages. A Constrained Set of Hypotheses</i>	99
James R. Hurford <i>What is Wrong, and What is Right, about Current Theories of Language, in the Light of Evolution?</i>	123
Adrien Meguerditchian, Jacques Vauclair <i>Communicative Signaling, Lateralization and Brain Substrate in Nonhuman Primates: Toward a Gestural or a Multimodal Origin of Language?</i>	135

Antonino Pennisi, Alessandra Falzone <i>Residuals of Intelligent Design in Contemporary Theories about Language Nature and Origins</i>	161
Ian Tattersall <i>Language as a Critical Factor in the Emergence of Human Cognition</i>	181
Jordan Zlatev <i>Human Uniqueness, Bodily Mimesis and the Evolution of Language</i>	197
Ines Adornetti <i>Making Tools and Planning Discourse: the Role of Executive Functions in the Origin of Language</i>	221
Francesco Ferretti <i>Travelling in Time and Space at the Origins of Language</i>	243

## COMMENTARIES

Alessandra Chiera <i>Origins of Human Communication</i> by Michael Tomasello	269
--	-----

## BOOK REVIEWS

Silvia Felletti <i>The Recursive Mind. The Origins of Human Language, Thought, and Civilization</i> by Michael C. Corballis	283
Serena Nicchiarelli <i>More than Nature Needs. Language, Mind, and Evolution</i> by Derek Bickerton	293

Introduction  
Origin and Evolution of Language:  
a Close Look at Human Nature

*Francesco Ferretti*<sup>†</sup>  
francesco.ferretti@uniroma3.it

*Ines Adornetti*<sup>‡</sup>  
ines.adornetti@uniroma3.it

Understanding the origin and evolution of language has been defined – rather provocatively – as *the hardest problem in science* (Christiansen & Kirby, 2003). To be sure, the study of the development of language is a subject that consistently generates great interest and controversy. The interest is largely dependent on the fact that a great part of Western theoretical investigation has attributed the “uniqueness” that characterizes our species to our “talking nature”: language constitutes the element that, more than any others, defines what it is to be human. The controversies relate to the methodological difficulties involved in investigating this topic. Language does not leave (literal) fossilized traces (or, at most, it leaves very indirect traces): it is not possible to reconstruct the origin and evolution of language in the same way in which we reconstruct the origin and evolution of other important human characteristics, for example the emergence of bipedalism and standing in an upright position. Moreover, what is considered valid evidence differs from discipline to discipline, also because language itself is rather difficult to define (or, at least, there is not a unanimous understanding of what constitutes language). Because of such difficulties, the Linguistics Society of Paris in 1866 in the Article 2 of its statutes banned any kind of debate among its members about the topic of language origins: “The Society accepts no communications concerning either the origin of language, or the creation of a universal language”. Following the edict of Paris, the Philological Society of London in

<sup>†</sup> Department of Philosophy, Communication and Visual Arts, Roma Tre University, Italy.

<sup>‡</sup> Department of Philosophy, Communication and Visual Arts, Roma Tre University, Italy;  
Department of Human Sciences, University of L’Aquila, Italy.

1872 took a similar decision. The scholars of the time, in fact, intended to avoid intractable arguments and disputes that the speculations on the origin of language, based on fragile empirical evidences, would have undoubtedly generated (cf. Müller, 1861).

The ban on such discussion and investigation along with the attitudes which underlie it had an enduring impact on research on the origin of language in the following 100 years, as linguists focused almost exclusively on how language works in its fully fledged form in modern humans, avoiding consideration about how it may have evolved. After nearly a century of silence, interest in these questions have been revived in modern linguistics, starting in the 1950s, thanks to the abundant collections of language data, comprehensive understandings on the behaviors of humans and other animals, and significant contributions from many other disciplines (cf. Gong, Shuai, & Comrie, 2014). Nevertheless, even if over time the situation has considerably changed, and the emergence of Darwinian evolutionary theory has allowed scholars to address the topic scientifically and systematically (cf. Tallerman & Gibson, eds., 2012), the old ban continues to engender among contemporary researchers the idea that the topic of the origin of language is not worth too much effort. Emblematic in this regard is the position of the most influential contemporary linguist, Noam Chomsky, according to which the study of the origin of language «is a complete waste of time» (Chomsky, 1988, p. 183). Despite these considerations, in our opinion the reasons that continue to fuel the ostracism against investigations into the origin of language are of an ideological kind rather than empirical. First of all, these reasons are linked to the suspicion with which the humanities view the theory of evolution (and the theory of evolution is, from our perspective, the only proper way to take into account the origin of language). But, above all, the motives behind the hesitancy to consider the origins of language as a research to take into account seriously are connected to a specific way to interpret language and its role in the constitution of human nature.

One of the most important ideological impediments to the recovery of studies on the origin of language is the Cartesian tradition that continues to influence some theoretical models within the sciences of mind and language (e.g. Chomsky, 1966; Fodor 1983, 2008). According to Descartes, the rational soul is the foundation of the *qualitative difference* between human beings and other animals: for many contemporary perspectives on human communication, language (which has taken the place of the soul) is a feature

that makes human *special* in nature. From these perspectives, human beings' uniqueness is not akin to how one animal species is different from another animal species; rather, human beings are *wholly* different from animals. Indeed, from Cartesian perspectives, humans are not animals at all. As consequence, the study of the origin of language is a complete waste of time «because language is based on *an entirely different principle* than any animal communication system» (Chomsky, 1988, p. 183, our emphasis). Language is the Rubicon which divides man from beast, and no animal will ever cross it (Müller, 1873).

From the perspective adopted in this special issue of *Humana.Mente* devoted to the origin and evolution of language, the fact that we can be proud of the extraordinary abilities that characterize our species does not contradict the notion that, indeed, these abilities can be attributed to the animal nature of human beings. The study of language in the Darwinian framework, then, starts from very opposite presuppositions than those of the Cartesian tradition. If it is true that human individuals are liable to numerous, slight, and diversified variations, which are induced by the same general causes, are governed and transmitted in accordance with the same general laws, as in the lower animals (Darwin, 1871), then it must be recognized that the differences existing between *Homo sapiens* and other animals, great as they are, are differences of degree and not of kind (Darwin, 1871). Language does not make an exception to this rule. Of course, there are differences between human language and animal communication. Nevertheless, these differences, however great, are interpretable in *quantitative* terms and not qualitative. A consequence of adhering to a framework of this kind is that language is not conceived as a *special* character that places our species apart from the rest of nature, but rather as a *specific* trait, just as the echolocation of bats used to navigate or the trunk of elephant used to manipulate objects are *specific* traits. It is in reference to an operation of this kind that Darwin's teaching—that conceives human beings as animals among other animals—is fully realized. When such a shift in perspective is accomplished, the question of the origin of language becomes congruent with the Darwinian tradition and becomes fully legitimate.

The articles collected in this special issue reflect the inherently interdisciplinary nature of research on the origin and the evolution of language. The volume provides a comprehensive survey of the most recent and advanced studies in language evolution investigation, bringing together the major perspectives on the topic, as shown in the primary fields represented:

archaeology, cognitive sciences, cognitive semiotics, evolutionary biology, linguistics, neuropsychology, neuroscience, paleoanthropology, philosophy, primatology, psycholinguistics, and psychology. In addition to providing an overview of the various ways in which it is possible to analyze the topic of the origin of language, the articles in this collection give a clear sense of the great intellectual strength and the propelling force that currently characterize this area of research. This leads us to think that the famous Paris ban definitely should be considered as something that belongs to the distant past. Despite the bad reputation of those who deal with the topic of the origin of language, the efforts of the participants in this volume cannot certainly be branded as “a waste of time”.

#### ACKNOWLEDGMENTS

We want to thank the editorial board of *Humana.Mente* for accepting to host this special issue on language evolution and for the support given to the publication of the volume. We thank the authors of the articles for kindly participating to the project. A special thanks goes to all the reviewers for their indispensable work.

#### REFERENCES

- Chomsky, N. (1966). *Cartesian Linguistics*. New York: Harper and Row.
- Chomsky, N. (1988). *Language and problems of knowledge*. Cambridge: MIT Press.
- Christiansen, M., & Kirby, S. (2003). Language Evolution: the Hardest Problem in Sciences? In Christiansen, M., & Kirby, S. (Eds.), *Language Evolution*, New York: Oxford University Press, 1-15.
- Darwin, C. (1871). *The Descent of Man; and Selection in Relation to Sex*. London: Murray.
- Gong T., Shuai, L. & Comrie, B. (2014). Evolutionary linguistics: theory of language in an interdisciplinary space. *Language Sciences*, 41, 243–253.
- Fodor, J.A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge: MIT Press.
- Fodor, J.A. (2008). Against Darwinism. *Mind and Language*, 23(1), 1–24.
- Müller, M. (1861). The theoretical stage, and the origin of language. In *Lectures on the sciences of language*. London: Longman, Green, Longman, and Roberts.

Müller, M. (1873). Lectures on Mr Darwin's philosophy of language. *Fraser's Magazine* 7-8, 147-233.

Tallerman, M., & Gibson, K.R. (Eds.). *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press.



# On the Relationship between the Air Sacs Loss in the Genus *Homo* and Duality of Patterning

Lluís Barceló-Coblijn<sup>†</sup>  
lluis.barcelo@uib.cat

## ABSTRACT

In a series of works, different models have shed light on the acoustic properties of *air sacs*, an organ located in the laryngeal region that is present in all great apes with the exception of humans. These works have shown how the loss of air sacs expands the number of possible digits but not the amount of signals per se. The number of signals in human language increased when the codifying property known as *duality of patterning* became characteristic of the codifying system, allowing digits to be combined in order for new signals to be codified. A direct relationship between air sacs and duality is presently being plotted, integrating linguistic theory and data from computational models into an evolutionary and developmental perspective of the evolution of modern speech.

Keywords: air sacs, duality of patterning, codifying system, combinatorial space, genus *Homo*

## 1. Introduction: air sacs and the genus *Homo*

Great apes in general could be argued to be similar in certain areas. Not just at the genomic level, but also regarding general morphology— for example, those traits that separate apes from monkeys – and even in some aspects of their behaviour. Nonetheless, there are some striking differences that provide interesting clues to divergent evolution, separating humans from the rest of great apes. Speech is one of these salient features. Speech is based on a highly specific vocal tract shape and a particular neuronal configuration (Ackermann, Hage & Ziegler, 2014). Thorough observation of the anatomy of great apes has verified one of the remarkable factors that make our vocal tract different

<sup>†</sup> University of the Balearic Islands, Spain.

from theirs: the fact that modern humans lack air sacs, an organ located in the centre of the larynx (connected to the hyoid bone in other great apes).

All great apes with the exception of humans have air sacs and can use them in their vocalizations. It is intriguing, that such an evident physical trait as laryngeal air sacs has not attracted more attention in evolutionary studies of linguistics. It was Fitch, who first demanded more attention for this apomorphy of the genus *Homo* (Fitch 2000; Hauser & Fitch, 2003). Thus, a frequent question is, when – in the lineage *Homo* – did these laryngeal air sacs disappear? Part of the answer has been related to the shape of the hyoid bone: it clearly shows the presence or absence of air sacs in great apes. According to what we have seen until now, the absence of air sacs in the genus *Homo* goes back at least to *H. erectus* (Capasso et al., 2008).

However, it is still uncertain whether *H. habilis* already lacked that organ, because no hyoid bone of this species has been found. What we know for sure is that the species *H. sapiens*, *H. neanderthalensis* and *H. heidelbergensis* share a “modern” hyoid bone (with some differences in comparison to *H. erectus*’ hyoid bone), in the sense that there was no orifice to connect the air sacs. Some authors consider *H. heidelbergensis* as simply an archaic state of *H. neanderthalensis* and hence these two species could be considered as a single species (Stringer, 2012; Cela-Conde & Ayala, 2001). If this theory is finally accepted, then only two species - and not three - would be confirmed to have a derived hyoid bone. The Denisovan hominin could also be included, given its close relationship with *H. neanderthalensis* (Reich et al., 2010).

We do not know whether these sacs appeared before or after the emergence of the genus *Homo* or even the emergence of our own species. According to Ann MacLarnon,

One possibility is that this occurred when the human thorax altered from the funnel-shape of australopithecines, to the barrel-shape of *Homo erectus*, as, in apes, air sacs extend into the thorax. It therefore quite probably occurred prior to the evolution of human speech-breathing control, and it may also have been a necessary prerequisite stage. (MacLarnon, 2011, p. 233)

Thus, it seems that the key lies in the use of the comparative method and the application of the principle of parsimony: many mammals have laryngeal air sacs; among the most representative and closely related to us, we find most primates, but also cetaceans.

Regarding cetaceans, Reidenberg & Laitman (2008) have made a physiological study where research was undertaken into the air sacs of two species: odontocets (for example, dolphins) and mysticets (for example, whales). The former have three kinds of sacs (nasal, parapharyngeal and laryngeal) while the latter show only a single laryngeal sac (considered homologous to the one present in artiodactyls, such as the reindeer or the takin from the Eastern Himalayas). Nevertheless, the function of air sacs seems to be different in odontocets, in comparison to great apes. However, it is important for the present discussion to consider that the presence of that organ in mysticets suggests how old this feature is in mammals. Of course, the function they serve can vary during evolution (following Love 2007, it is worth noting that it is structures – along with their activities – that *evolve*, but not functions).

Recent works have explored the acoustic characteristics of air sacs and have discovered a special influence of the sound produced by the air sacs on the sound generated by the vocal folds (de Boer, 2008; Riede et al., 2009). The results of these works coincide regarding the effect on vocal sounds, whose formants are affected by the sound of the superimposed air sacs. A recent work on acoustic perception by de Boer (2012) has shown that humans have difficulties in discerning the type of vowel sound when there is air sac intervention, pointing to co-evolution of speech and hearing. De Boer considers that, given the benefits of a vocal tract without air sacs, which counts on a richer vocalic space, the human vocal tract has evolved so that communication has been enhanced:

If it is assumed then that the experimental results are due to lower distinctiveness of the stimuli and that communication is more successful when one is able to produce more distinctive signals, it follows that having an air sac attached to the vocal tract is an impediment for successful communication through speech. (de Boer, 2012)

I would like to propose a further analysis that takes into account de Boer's considerations about the acoustic evolution of speech. This analysis focuses on signal design and the factors intervening in human signals. This approach singles out the role of *duality of patterning* – a property of the human codifying system – in the expansion of signals. It is also shown how the eventual loss of air sacs has affected the combinatorial space within which the codifying system (and hence the property of duality property) operates. Hence, to understand

better the relationship between the loss of air sacs and signal expansion, duality of patterning must be taken into account. The second section is devoted to the explanation of certain significant notions of language as a codifying system and the consequences of increasing (1) the number of primitive meaningless digits or (2) the length of the resultant signals. I will then go to show how the human capacity for classifying linguistic signals is not only tied to the typical linguistic sounds but also to other types of external input. These observations lead me to the final section where I will make a reflection on the evolution of air sacs and signalling in mammals, drawing special attention to primates.

## 2. Acoustics, hearing and air sacs

De Boer (2008) has recently created an initial physical model of air sacs in order to study this laryngeal organ. The results suggest that this organ has a relevant effect on vocalization. De Boer's model takes its inspiration from the morphological characteristics of the howler monkey (*Aoulatia guariba*). The author compared both kinds of productions and obtained very similar results: Howler monkey vocalizations have peaks around 300, 750 and 1410 Hz, whereas the artificial model reaches 215, 725 and 1215 Hz. De Boer has observed that, when air sacs are added to the model, the formants rise, reaching higher values and new frequencies, constraining the ability to articulate new sounds. This could be the reason, the author argues, why air sacs were lost during the evolution of genus *Homo*. Notwithstanding, de Boer points out that the remaining functions attributed to this organ could be equally valid and, for the moment, he sees no reason to reject them (on this matter, see section four). Riede and colleagues (2008), in their turn, have created several models and their results point to an increase of the variability of the impedance of the vocal tract. Coinciding with de Boer, the results of Riede's team show that air sacs destabilize the sound source. The authors have recently created a progressive model to which new elements (the larynx, the bulla, the air sac, etc.) are added in order to compare the results of each "phase". The authors have taken into account currently available data on the Siamang (*Symphalangus syndactylus*, from the family *Hylobatidae*), who are known to fill the air sacs during their "boom" call. Riede and collaborators detected two acoustic effects produced by the air sacs: 1) air sacs increase the dynamic range of sound emission, but

only at the higher and lower limits; 2) the vocal variability can be increased in different ways, some of them subject to non-linear and unstable phenomena.

It has been argued that the conception of co-evolution of vocal tract anatomy and hearing is possible (Barceló-Coblijn, 2011a, 2011b) – and certainly appealing for a more embodied and (evo-)developmental approach to organic evolution. In other words, speech goes hand in hand with hearing, since, for a sound signal to be voluntarily emitted with a communicative intention, the sender has to be able to (1) recognize the signal, and (2) categorize it as a linguistic signal. De Boer, (2012) has proven that modern humans find it more difficult to perceive vowels properly if vowels are modified by the superimposed sound of air sacs. The participants recognized them as linguistic sounds, but had problems to determine the identity of a pair of vowels (for instance, [I] was confused with [y], both native sounds in Dutch, the participants' language). Thus, data from both acoustics and perception studies suggest that, on the one hand, the presence of air sacs makes it somehow problematic for humans to perceive speech sounds adequately. On other hand, the data also suggest that for a human-like “speech system” it is better not to have attached air sacs to the laryngeal structure, in order to produce a wider range of speech sounds. How do these findings relate to the modern speech capability of humans?

In what follows, I will argue, following de Boer and Riede and collaborators, that the loss of air sacs was crucial for the evolution of modern speech in its current form and that the disappearance of this organ has affected the inventory of primitive sounds. In addition, I will connect this acoustic aspect with a second one, directly related to the recently evolved codifying ability of humans. Importantly, this ability is strongly based on the property of duality of patterning. Finally, I will highlight another important point, more psychological in nature, for the creation of signals: the capability of perceiving different kinds of input – not just the typical linguistic sounds – that can nevertheless be categorized as linguistic signals.

### 3. On meaningless digits, duality and codifying systems

The linguistic signal has classically been identified as an arranged string of sounds. However, today we know that it can also be made up of bodily signs, as in sign languages. Language can be envisaged as both a thought system and a

communication system – and, contrary to what some scholars think, both are not mutually exclusive. Regarding the latter aspect, there is a wide agreement that humans are able to codify linguistic units made up of primitive sound/sign elements. The identification of this codification process goes back to Martinet (1960) and Hockett (1958, 1960). It was Hockett who identified the linguistic property of *duality of patterning* or simply “duality”. As such, duality allows the codifying system to combine a discrete set of primitives into strings of elements (these can be for example sounds, movements or lights). Each string codifies a particular *meaningful unit*. Hockett called the primitive elements “cenemes” and the final codified and externalizable units, “pleremes”. According to Hockett, in human language morphemes (and not words) are the most equivalent to pleremes, because they are discrete units codifying a meaningful unit. Morphemes have a phonological “envelope” that facilitates their externalization. The influence of Shannon’s Information Theory in Hockett’s conception of duality is deep (Barceló-Coblijn, 2012; Fortuny, 2010), so that a degree of abstract thought is required to understand his vision of language as communication system.

To understand better the relationship between the loss of air sacs and signal expansion, duality must be taken into account. For example, in Table 1 we find a description of a possible code. It is very simple: it has only two primitive meaningless elements (*cenemes*), namely A and B. It is thus a quite simple binary code. The codifying system must be able to codify meaningful units, in order to be externalized. Once codified, the system will then have of an inventory of *meaningful signals* (or *pleremes*) at its disposal – it is important not to confuse meaningful “units” with meaningful “signals”; the former will be codified into the latter. For example, the meaningful unit denoting “plurality” is codified, in general, into the meaningful signal “-s” in Catalan language (e.g., *llop* > *llops*, “wolf > wolves”), whereas in Italian language it is used, in general, a vowel (e.g., *lupo* > *lupi*, “wolf > wolves”). Hence, both Catalan and Italian have the same meaningful unit, but different signals. Hockett did not offer any label for the meaningful units. Some scholars call them *lexical items* (Barceló-Coblijn, 2012; Boeckx, 2008; Ott, 2009). In Table 1, code-words/pleremes may have up to 4 possible lengths (from 1 to 4 digits). As we have seen, the reduced inventory of cenemes forces the codifying system to reuse these primitive digits once and again.

2-digits code (cenemes: A, B)							
Length of the code-word		Code-words (pleremes, meaningful units)				$N$	Set
1 digit-long	$2^1$	A, B				2	$\alpha$
2 digits-long	$2^2$	AA, BB, AB, BA				4	$\beta$
3 digits-long	$2^3$	AAA AAB BAB ABA BBB BBA ABB BAA				8	$\gamma$
4 digits-long	$2^4$	AAAA	BBBB	ABBB	BAAA	16	$\delta$
		AAAB	BBBA	BBAA	AABB		
		AABA	BBAB	BAAB	ABBA		
		ABAA	BABB	BABA	ABAB		

Table 1: A binary code and several possible code-word lengths. Sets refer to all possible combinations determined by the length and the number of digits. The number of digits is always 2 and the length of the code-words are the exponents.

Interestingly, the signal expansion of a system depends on the previous expansion of two sets: the set of *meaningful units* and the set of *meaningless digits*. The latter will be the combinable material for the codification of the meaningful units, so that new signals can be created.

It is crucial to know the kind of system and the constraints it imposes on the length of the signal. Some codes allow only one type of length (for instance, in Table 1, only  $\alpha$  or only  $\beta$ ). Importantly for the discussion of language evolution however, other codes could allow different signal lengths (for instance,  $\alpha$  and  $\beta$ , or  $\alpha$  and  $\beta$  and  $\gamma$ ). As we see, the maximal Combinatorial Space (CS) of the code is related to the sum of the number of digits elevated to the length of the code-word:

(1)

Code  $C_k = \alpha + \beta = 2^1 + 2^2 = 6$  possible code-words

Code  $C_i = \beta + \gamma = 2^2 + 2^3 = 12$  possible code-words

When a codifying system has duality, it usually codifies some signals from among those that are possible. However, in animal systems, like *H. sapiens*, duality rarely makes use of all possible combinations, exhausting them. In the case of human languages, this is certainly the case (many possible combinations do not come up as morphemes). This fact increases *redundancy*, a property that enhances the probabilities that a message will be decoded, and hence communication will be successful. Nevertheless, this fact makes the code more prone to changes (an aspect leading to the evolution of multiple languages). Imagine that DNA is a conventional code. It has 4 digits (G, C, U, A), but the code-words can only be 3-digits-long: thus, the combinatorial space is  $4^3 = 64$  possible code-words. It has been proven that all possibilities exist and represent different codons. Thus, this is a robust code that does not allow much manoeuvrability.

So far, we have seen how a codifying system with the property of duality operates. Other systems could behave differently, for example, attributing one different ceneme/digit to each meaningful unit. For small codes, this is no problem. However, in order for the code to be expanded, it will eventually need more and more *memory*.

As I will explain, the loss of air sacs – along with other evolutionary changes – has made the expansion of the set of *cenemes*, i.e. the meaningless digits possible in humans (also called the *alphabet* of the code). However, the disappearance of air sacs did not directly affect the set of public signals, since this depends on the codifying characteristics of the system (see next section). Interestingly, we will see that the very same morphological change has happened in many species, though it has not always been accompanied by an increase in cenemes. Before approaching evolutionary explanations, I will summarize the most important factors intervening in human signal design.

#### 4. Factors in the composition of human signals

Human language allows different kinds of combinations, thus providing different lengths in the code-words (the morphemes). But this fact, although affected by the number of digits (say, phonemes or signs), does not directly affect the final number of signals. The codifying system is in charge of this task.

When air sacs disappeared, the possible number of sounds was altered, progressively increasing because, in the genus *Homo*, it was also followed by a change in the vocal tract (Boë et al., 2002; Boë et al., 2004; Barceló-Coblijn, 2011a). However, the number of phonemes – psychological entities – also had to increase. Nonetheless, this aspect was dependent upon cognitive changes, not just physiological ones. However, the number of public signals was still dependent upon the codifying system.

Some animals have the capacity of producing more oral sounds than they actually do (for example, the nightingale can produce around 1000 notes, but this bird usually sings around 200 songs; see Hurford, 2011). It is probably ecological factors, among others, that do not push them to exhaust the possibilities of their oral capacity.

If our ancestors, instead of developing a system that allows combination of several code-word lengths, had evolved towards a rigid system based on a single, though larger code-word length, then the number of signals would have probably increased. However, the CS would still be much more reduced and dependent on memory resources. Let us see an example, using the information from Table 1:

If a Code  $C_1$  (with a  $CS = 2^2 = 4$ ) increases by one the number of digits (cenemes), but the length remains untouched, then  $CS = 3^2 = 9$  possible code-words. However, if Code  $C_1$  evolves to Code  $C_2$  (with a  $CS = \beta + \gamma = 2^2 + 2^3 = 12$  possible code-words) – i.e., allowing two different lengths –, the CS of possible code-words is larger than in  $C_1$ . Of course, a combination of both more digits and lengths would increase the number of code-words enormously, as is the current case of human languages.

Additionally, humans can combine pleremes, a further step in the evolution of language (the so-called *compounds* like “fire-fly”). In this case, it is *productivity* (Hockett, 1961) – somehow similar to *recursion* (Hauser, Chomsky, & Fitch, 2002) – that has been argued to be involved in compounding (Rosselló, 2006; Barceló-Coblijn, 2012a).

Human public linguistic minimal signals, the morphemes, seem to be affected by several different factors: Firstly, the set of basic meaningless digits (or cenemes), which is determined by a complex relationship between the mind and body. For example, in oral languages, the morphology of the vocal tract has to go hand in hand with a proper mental representation and categorization of the sounds – hence intervening both speech- and hearing-related brain areas (e.g. Stowe, Haverkort, & Zwartz, 2005; Ackermann, Hage, & Ziegler,

2014). For this reason, a belch is not categorized as linguistic sound. Even some possible but quite unusual speech sounds such as the clicks of Bantu languages, would not be categorized in the “linguistic” category the first time, if the adult hearer has no previous experience with them (see next section on phonology). As we have seen, it is this first factor that is affected by the loss of air sacs in the genus *Homo*, in a complex series of evolutionary changes affecting both the physical morphology of the vocal tract and neural connectivity and activity.

Secondly, the length of the code-word (in the case of morphemes) is affected by Zipf’s statistical law (Zipf, 1936), which states that the length of a word is inversely proportional to its rank in a frequency table. Note that, in this case, “words” are the object, not “morphemes”. In English (especially with words of Anglo-Saxon origin) it is easy to find a 1:1 relationship between words and morphemes; in many cases one morpheme equals one word (therefore many people believe that “duality creates words”). However, this 1:1 equivalence is certainly more problematic in Romance languages, for example, where mono-morphemic words are much less common. As noted above, when a word is made up of more than one morpheme, an additional mechanism has intervened: *recursion* (as defined in Hauser, Chomsky, & Fitch, 2002; but see Barceló-Coblijn 2012a for some problems of their definitions).

Thirdly, the list of *meaningful units* is strongly connected to cognition. It is the brain/mind that cognizes the list of meanings. In human languages it has been proposed that there is a Conceptual-Intentional module that is responsible for the creation of meaningful units (Chomsky, 1995, 2000). In the field of philosophy of language there is an even stronger thesis, according to which human thought cannot be separated from syntax, so that thoughts and meanings are directly affected and structured by syntax (Hinzen, 2006, 2011; Gomila, 2011).

And finally, there is the intervention of the codifying system. If it has the property of duality, it has the possibility of combining digits in many different ways. Thus, a code with duality makes available a more creative and structurally different set of signals, than a simpler code that can make use of repetition only (for example a code with a single digit  $\beta$ , where all signals are the same but just a little bit longer:  $\beta$ ,  $\beta\beta$ ,  $\beta\beta\beta$ , etc.). Importantly, a code with duality needs a smaller channel capacity (Shannon, 1948; Shannon & Weaver, 1949), and in psychological terms, there is an optimization of memory resources. Another

side-effect of this kind of codes is an increase of redundancy, optimizing the message deciphering (Hockett, 1961, 1987).

Up to here, we have seen some important factors affecting the design of human linguistic signals and how they interrelate to each other. Most of these factors are not related *per se* to human cognition and, although the psychological nature of phonemes has been mentioned while describing the first factor, I would like to go more deeply into this latter question, stressing a really important aspect of human phonology that linguists have become aware of recently.

#### 4.1 Another factor in linguistic signals

There is an important aspect that is directly related to digits: the cognitive device that interprets and categorizes them as linguistic digits or cenemes. In the previous section, it has been noted that most languages make use of sounds, which are categorized as phonemes by human cognition. However, the presence of sign languages highlights an important aspect: human “phonology” – traditionally related to sounds (from Greek, *phonos* “sound, voice” and *logos* “word, speech, subject of discussion”) – also processes signs, or even lights. There also exist the so-called whistle-languages, like the Gomerian Spanish whistle language (GSWL) (Classe, 1957). It has been proven that whistlers of GSWL do activate the areas of the brain normally associated with spoken-language function (Carreiras et al., 2005). Hence, “phonology” is in fact a human capacity able to process different kinds of inputs as possible linguistic cenemes. It is for this reason that it has been observed that phonology is probably *substance-free* (Maihlot & Reiss, 2007). Humans would seem to have developed a kind of *signal*-logy capable of processing and classifying signs, sounds, or even lights into linguistic categories. This fact could somehow weaken the view that the human vocal tract evolved “for” communication, given that hominins could have made use of other kinds of signs, and not just the typical linguistic sounds.

### 5. An evolutionary reflection about the loss of air sacs and communication

In the previous section I have pointed out the apparent substance-free condition of phonology as a possible counter-argument to the view of vocal-tract evolution due to communicative pressure. Evolutionary theory has long ago observed that there are at least four factors in organic evolution: mutation, genetic drift, migration and natural selection. We should add other factors like introgressive hybridization (Mallet, 2005, 2008) and phenotypic accommodation (West-Eberhard 2003, 2005). I think that research in the evolution of air sacs could be benefited by an increase of the number of intervening evolutionary factors. De Boer (2012) has observed that the increase and enhancement of the vocal system once air sacs disappeared, suggests that natural selection could have favoured individuals lacking air sacs. This observation, argues de Boer, would fit humans. From this point I would like to extend the analysis on the evolution of this organ. Without making any claims against the intervention of natural selection – which ultimately selects phenotypes, rather than changes at the genomic level – , it is possible to make this hypothesis compatible with an enriched theoretical approach taking into account additional factors to natural selection. The theoretical approach I put forward also aims to include more than one species, and eventually, all mammals.

Let us consider the work by Hewitt and colleagues (2002) on the presence or absence of air sacs, in which a broad descriptive study of this trait has been carried out. The authors offer four cladograms where four macro-families of primates (124 species altogether) are depicted. I have summarized this information in the following table:

Species	<i>n</i> of species	With air sacs	Without air sacs
Strepsirrhini	8	35	3
Cebids and callitrichids	25	10	15
Cercopithecoidea (macaques, papio and cercopithecini)	40	9	31
Colobines, Hominoidea:	11	8	3
Gibbons	9	6	3
Hominoids	5	4	1
Total	128	72	56

Table 2 : Laryngeal air sacs in primates, according to Hewitt et al. (2002).

Additionally, studies on both cetaceans and primates suggest that air sacs are a fairly common trait so that the most parsimonious conclusion is, that in great apes the lack of air sacs is a derived trait, and their presence an archaic trait. Given that, among great apes, air sacs are only absent in *H. sapiens* – pathological air sacs aside (Giovanniello et al., 1970) – such an evolutionary novelty can be considered an autapomorphy of our species. Regarding the rest of the extinct hominins of the genus *Homo* that eventually cohabited with *H. sapiens*, this feature is shared, at least, with *H. neanderthalensis* and *H. erectus*. Hence, we could talk of an apomorphy of the whole genus *Homo*, rather than an autapomorphy of a single species. However, to do that we still need to be sure about *H. habilis*' hyoid bone shape.

Reidenberg and Laitman (2008) have also considered the possibility that laryngeal air sacs in mysticets could have an older origin in an archaic laryngeal ventricle, which would have moved away over time. Thus, we are dealing with an ancient feature in mammals, and therefore the observed functions are manifold:

Firstly, regarding cetaceans, Reidenberg and Laitman (2008) take into consideration (1) the increment of vocal resonance; (2) generating

vocalizations underwater; (3) to prevent drowning; (4) to elongate the sound length; (5) to reuse the air once and again in vocalizations, given that they are mammals and have a limited access to air; (6) pumping air from the sacs to the lungs would allow them to use the same air volume again and again, making possible multiple vocalizations underwater; finally (7), the authors concede the possibility that, given that it is the energy of vibration which is transmitted through water, there could be a functional coincidence with the laryngeal ventricles of terrestrial mammals. Reidenberg and Laitman also point out a structural function (8), since the presence of the air sacs affects the general density of the head.

Secondly, regarding primates, the debate on the possible functions of air sacs still endures: (1) Negus (1949) argued that they were used for saving exhaled air; (2) Hewitt, MacLarnon and Jones (2002) in turn proposed a reduction of hyper-ventilation; (3) Fitch and Hauser (2003) believe that, on the one hand, primate air sacs can generate a new sound source; (4) on the other hand, these authors also proposed that this organ makes possible the production of stronger and longer lasting calls; (5) Hayama (1970, 1996) in turn, thinks that they are useful for softening pressure.

So, many morphological differences, in primates as well as in cetaceans, warn us that the functions of air sacs can vary from species to species. Probably, the first function they ever had will never be known. This multifactorial aspect of air sacs points to an extreme dependency relation between organ and (1) the environment and (2) the evolutionary history of each species. Altogether this suggests that function cannot be used as the first evolutionary argument. This seems also to affect the argument on communication. Many primates have lost their air sacs, and no improvement has been attested so far – at least, not in the same way as in humans. Moreover, field data have shown that there are species that could make use of air sacs to expand their signal repertoire. For example, Campbell's monkeys, like many primates with air sacs, produce the "boom" call, which is basically the sound of air sacs (Ouattara et al., 2009). As it was mentioned in the introduction, this is also a possibility for Siamangs. On another front, McComb and Semple (2005) have found a positive correlation between the number of public signals and group size – although, when observed case by case, there are several exceptions. Importantly, these authors state that,

It is important to note that the direction of causality cannot be inferred from correlational analyses, therefore it is not possible to say whether evolutionary increases in vocal repertoire sizes directly preceded or followed increases in levels of sociality (McComb & Semple, 2005, p. 383)

All this taken together makes me somewhat wary of focusing the evolutionary explanation on a single factor such as communication or natural selection, or size group. Instead, I think that a suitable explanation can embrace other additional factors too. For instance, species-specific mutations could have arisen independently in many mammals, bringing on the loss of air sacs. This is especially conceivable regarding the phylogenetic distance between primates, ungulates and cetaceans. Additionally, it would be interesting to take the developmental patterns of each species into account, recalling Evo-Devo ideas that go back to Alberch et al. (1979), Alberch (1989) and Oster & Alberch (1982). These proposals put forth the hypothesis that small variations could lead species in one direction or another, yielding different phenotypic morphologies in each case: for example, a group of primates that splits and evolves into several differentiated species could develop any of the attested air sacs morphologies, or even the lack of the trait. Different phenotypes are also possible even when two species have the very same genes (Pigliucci et al., 1996). Much more probable than a mutation is methylation. The loss of air sacs could also be due to the methylation of some of the genes involved in the development of this organ. As it is well known, gene methylation has an important effect in gene expression, since methylation can “deactivate” a gene. Methylation has also been referred as “reprogramming” (Mann & Bartolomei, 2002).

The variability in functions in addition to the great quantity of primates that have lost air sacs also suggests that – in some cases – such an evolutionary change could have neutral consequences in terms of survival. A subject that develops lacking air sacs still has the rest of the body and brain intact. It still can communicate, though in a rather different way. To the author’s knowledge, there is no contrasted information about the reactions between coexisting species, which are close phylogenetically though different concerning the presence or absence of air sacs.

This altogether suggests that the loss of air sacs is a common change that may happen in mammals in general, and that could probably be rather neutral,

given the number of species which have survived the loss of this trait. Once a species has evolved and has fixed the developmental path that yields the loss of air sacs, the intervention of natural selection and other ecological factors will certainly have their effect on the species. However, I would not eliminate but rather diminish the role of fitness and survival regarding the presence or absence of air sacs. Probably, as has been put forward, survival depends on many other factors that are not affected by the presence or absence of this organ. What would seem to be logical, as was pointed out in the introduction, is that there has been co-evolution of speech and hearing, and that when air sacs disappeared, auditory perception had to be ready to process the new available sounds.

### Conclusions

In the present work I have reviewed issues on air sacs in relation to communication and signal design. I have noted that there is a direct relation between the loss of air sacs and the codifying system, and particularly with the property of duality. Whereas duality, as a property, has an effect on the combinability of digits and the structure of public signals, the absence of air sacs has had an incremental effect on the number of primitive elements with which a codifying system with duality can codify meaningful units. Finally, I have observed that human phonology may be considered substance-free, meaning that it deals with sounds (including whistles), signs or even lights. This point should be taken into account when theorizing about the evolution of air sacs in humans and their relation to communication. The presence and loss of air sacs are common traits in primates or in other species such as cetaceans or ungulates, suggesting that air sacs are an ancient feature in mammals. Finally, I have observed that nowadays, many different functions for this organ have been attested. Because air sacs have disappeared in many species, it is difficult to base the evolutionary explanation of such a loss only in functional terms. Therefore, I have included other factors also intervening in evolution, minimizing the role of natural selection, in favour of a multifactorial evolutionary explanation.

## ACKNOWLEDGEMENTS

I would like to thank an anonymous reviewer for constructive comments, observations and references. This work has been supported by the grant FFI2010-20759 from the Ministerio de Ciencia e Innovación (Spain), the Balear Government and FEDER funds.

## REFERENCES

- Ackermann, H., Hage, S. R. & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *15*, 1-84
- Alberch, P. (1989). The logic of monsters: Evidence for internal constraint in development and evolution. *Geobios*, *12*, 21-57.
- Alberch, P., Gould, S. J., Oster, G. F., & Wake, D. B. (1979). Size and shape in ontogeny and phylogeny. *Paleobiology*, *5*(3), 296-317.
- Barceló-Coblijn, L. (2011a). A biolinguistic approach to the vocalizations of *H. neanderthalensis* and the genus *Homo*. *Biolinguistics*, *5*(4), 286-334.
- Barceló-Coblijn, L. (2011b). Co-evolution of hearing and speech in the Genus *Homo*: The acoustic effect of air sacs on vocalizations and the consequences of their disappearance. *Spanish Journal of Psychology*. Proceedings of the 4th Iberian Conference on Perception, CIP-Mallorca 2011, 6th-8th July 2011 in Palma (Balearic Islands).
- Barceló-Coblijn, L. (2012a). Evolutionary scenarios for the emergence of recursion. *Theoria et Historia Scientiarum: International Journal for Interdisciplinary Studies IX*, 173-201
- Barceló-Coblijn, L. (2012b). [Neandertals. Why could they have spoken and why not?]. *Neandertals. Per què podrien haver parlat i per què no*. Doctoral dissertation. Palma (Mallorca): Universitat de les Illes Balears.
- Boë, L.J., Heim, J.L., Abry, C., Badin, P. (2004). Neandertal vocal tract: Which potential for vowel acoustics? *Interaction Studies* *5*, 409-429.
- Boë, L.J., Heim, J.L., Honda, K., Maeda, S. (2002). The potential Neandertal vowel space was as large as that of modern humans. *Journal of Phonetics* *30*(3), 465-484.
- Boeckx, C. (2008). *Bare Syntax*. Oxford: Oxford University Press.

- de Boer, B. (2012). Loss of air sacs improved hominin speech abilities. *Journal of human evolution*, 62(1), 1–6.
- de Boer, B. (2008). The Joy of Sacs. In A. Smith, K. Smith, & R. Ferrer i Cancho, (Eds.), *The Evolution of Language: Proceedings of the 7<sup>th</sup> international Conference (EVOLANG7)*, 415–416. Barcelona.
- Capasso, L., Michetti, E. & D’Anastasio, R. (2008). A *Homo erectus* hyoid bone: possible implications for the origin of the human capability for speech. *Coll. Anthropology* 32(4), 1007-1011.
- Carreiras, M., López, J., Rivero, F., Corina, D. (2005). Linguistic perception: neural processing of a whistled language. *Nature*, 433(7021), 31-32.
- Chomsky, N. (1995). *The minimalist program*. MIT Press.
- Chomsky, N. (2000). Minimalist Inquiries: The Framework. In Martin, R., D. Michaels, & J. Uriagereka, J. (Eds.) *Step by Step: Essays on Minimalist Syntax in honour of Howard Lasnik*, Cambridge, Mass: MIT Press.
- Classe A. (1957). The whistled language of La Gomera. *Scientific American*, 196, 111-124.
- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, 4(7), 258–267.
- Fitch, W. T., Hauser, M.D. (2003). Unpacking “Honesty”: vertebrate vocal production and the evolution of acoustic signals. *Acoustic communication*, 404.
- Fortuny, J. (2010). On the duality of patterning. In Zwart, Jan-Wouter and De Vries, Mark (Eds.) *Structure Preserved: Studies in syntax for Jan Koster*, 164:131–140. LinguisticAktuell/ Linguistics Today xxiii. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Gomila, A. (2011). *Verbal Minds: language and the architecture of mind*. Editorial, Elsevier.
- Hauser, M. D., Chomsky, N. & Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579.
- Hauser, M. D. & Fitch, W.T. (2003). What are the uniquely human components of the language faculty? In M.H. Christiansen & S. Kirby (Eds.), *Language Evolution: The States of the Art*, Oxford: Oxford University Press, 158-181.
- Hewitt, G., MacLarnon, A., & Jones, K.E. (2002). The functions of laryngeal air sacs in primates: a new hypothesis. *Folia Primatologica: International Journal of Primatology* 73(2-3), 70–94.

- Hinzen, W. (2006). *Mind Design and Minimal Syntax*. Oxford: Oxford University Press.
- Hinzen, W. (2011). Emergence of a systemic semantics through minimal and underspecified codes. In C. Boeckx & A. di Sciulo (Eds.), *The Biolinguistic enterprise: new perspectives on the evolution and nature of the human language faculty*, Oxford: Oxford University Press, 417–439.
- Hockett, C. F. (1958). *A Course in Modern Linguistics*. New York: Macmillan.
- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203, 89–96.
- Hockett, C. F. (1961). Linguistic Elements and Their Relations. *Language* 37(1), 29–53.
- Hockett, C. F. (1987). *Refurbishing Our Foundations: Elementary Linguistics from an Advanced Point of View*. John Benjamins Publishing Company.
- Love, Alan C. (2007). Functional homology and homology of function: Biological concepts and philosophical consequences. *Biology & Philosophy* 22(5), 691–708.
- MacLarnon, A. (2011). The anatomical and physiological basis of human speech production: adaptations and exaptations. In K. R. Gibson & M. Tallerman (Eds.), *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press.
- Mann, M. R., & Bartolomei, M. S. (2002). Epigenetic reprogramming in the mammalian embryo: struggle of the clones. *Genome Biology* 3 (2), reviews1003.reviews1001.
- Maihot, F., & Reiss, C. (2007). Computing long-distance dependencies in vowel harmony. *Biolinguistics* 1, 28–48.
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20(5), 229–237.
- Mallet, J. (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1506), 2971–2986.
- Martinet, A. (1960). *Éléments de linguistique générale*. 4th ed. Armand Colin.
- McComb, K. & Semple, S. (2005). Coevolution of vocal communication and sociality in primates. *Biology Letters* 1, 381–385.
- Negus, V. E. (1949). *The comparative anatomy and physiology of the larynx*. 1st ed. W. Heinemann Medical Books.
- Pigliucci, M., Schlichting, C.D., Jones, C.S., Schwenk, K. (1996). Developmental reaction norms: the interactions among allometry, ontogeny, and plasticity. *Plant Spec. Biol.* 11, 69–85.

- Oster, G. & Alberch, P. (1982). Evolution and Bifurcation of Developmental Programs. *Evolution*, 36(3), 444-459
- Ott, D. (2009). The Evolution of I-Language: Lexicalization as the Key Evolutionary Novelty. *Biolinguistics* 3(2), 255-269.
- Reidenberg, J.S., Laitman J.T. (2008). Sisters of the sinuses: cetacean air sacs. *Anatomical Record*, 291(11), 1389-1396.
- Riede, T., Tokuda, I.T., Munger J.B., & Thomson, S.L. (2008). Mammalian laryngeal air sacs add variability to the vocal tract impedance: Physical and computational modelling. *The Journal of the Acoustical Society of America*, 124(1), 634-647.
- Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L. F., Maricic, T., Good, J.M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E.E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M.V., Dereviako, A.P., Hublin, J.J., Kelso, J., Slatkin, M., & Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053-1060.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423, 623-656.
- Shannon, C.E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. 1ST ed. University of Illinois Press.
- Stowe, L.A., Haverkort, M., & Zwarts, F. (2005). Rethinking the neurological basis of language. *Lingua*, 115(7), 997-1042.
- Stringer, C. (2012). The status of *Homo heidelbergensis* (Schoetensack 1908). *Evolutionary Anthropology*, 21, 101-107.
- West-Eberhard, M.J. (2003). *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.
- West-Eberhard, M.J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Science USA*, 102, 6543-6549.
- Zipf, G.K. (1936). *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. New York: Routledge.

# From Holophrase to Syntax: Intonation and the Victory of Voice over Gesture

*Teresa Bejarano*<sup>†</sup>  
tebefer@us.es

## ABSTRACT

In the origin of syntax, primitive, holophrastic signs had to be weakened (original, drastic ‘bleaching’) and to lose their previous status of whole message. The original syntax was probably thema/rhema syntax. The earliest themas repeat the hearer’s message: the speaker embeds the hearer’s message in his own message. In this way a holophrase could be weakened, and turn into a part of a syntactic combination. This pregrammatical, interpersonal ‘recursive embedding’ is embodied in sensorimotor processes. The upper level is embodied in the intonation; the lower level, in the articulatory-phonetic word. This decoupling of intonation and articulatory pattern—i.e. the emergence of intonation capable of comprising more than one word—facilitated the weakening of previous holophrases and the genesis of syntax. In time, that facilitation determined the preeminence of voice over gesture, regardless of whether or not that preeminence existed before syntax.

Keywords: bleaching (semantic weakening), embodied cognition, holophrastic sign; intonation, recursive mind.

In the origin of syntax primitive, holophrastic signs had to be weakened (original, drastic ‘bleaching’) and to lose their previous status of whole message. The original, pregrammatical syntax was probably the predication in response to the hearer’s previous message. This previous message is embedded in the speaker’s message. In this way, the hearer’s holophrastic sign could be weakened and turn into a part of a syntactic composition. This is developed in Section 1.

Section 2 focuses on a more concrete issue. In original syntax, the hearer’s holophrastic message is only repeated at articulatory level, not at intonation

<sup>†</sup> University of Sevilla, Spain.

level. The intonational unit is already able to comprise two words. This involvement with audiomotor resources could mitigate the difficulties of original syntax.

Section 3 extracts a consequence from Section 2. I begin by acknowledging the similarity between verbal sign and pantomimic gesture: both involve motor imitation. It is even likely that gestures were as important as voice for a long period of time. However, when the syntax arose, intonation would have given the voice the decisive edge.

1. Thema/rhema-syntax and the hearer's previous message.  
Interpersonal recursive embedding

The original syntax was probably thema/rhema-syntax (Aitchison, 1998; Givón, 1979; Hurford 2007 and 2012, p. 187, and chapter 9; Jackendoff, 2003; Tomasello, 1999). Since they were coined by the Prague Circle of the Twenties, the concepts of thema and rhema have been frequently reformulated. In this way, differences have been—more or less consistently—established between three dimensions—given/new and topic/comment, and also background/focus (see Hinterwimmer, 2011). But, since these reformulations refer to the grammaticalization level peculiar to our present-day language, this paper is not going to deal with them. I take the predication in response to the hearer's message as the basic type of this syntax. The *thema* mirrors what is known by the hearer; therefore, in the basic type, it mirrors the particular combination of knowledge and ignorance previously expressed by the hearer. The *rhema* adds what the hearer does not know, i.e. what the limited knowledge of the hearer needs in order to become similar to that of the speaker.

Thus, the previous message uttered by the hearer expresses a degree of knowledge different from the speaker's. In present-day language, previous messages revealing a particular combination of knowledge and ignorance or mistake can be questions or wrong predications, but also requests for something unavailable, or calls to someone absent. Those first two cases—we could say—would lead us into a vicious circle if we projected them onto the historical origin of syntax. By contrast, previous messages in the form of calls or requests could be expressed in an entirely holophrastic language.

But the key question is: Did ‘parts of sentence’ exist in the original holophrastic language? Heine and Kuteva (2007, p. 300) say that «at layer 1 (holophrastic, i.e., one-word utterances) there was only one kind of category, namely nouns». However, they acknowledge that «a category ‘noun’ does not make much sense unless contrasted with other kinds of word categories», and conclude: «What exactly the reconstruction of a category ‘noun’ means with reference to layer 1 is a question that is open to further research». In my view, the original holophrastic sign is as close to (or as far from) the meaning ‘give me’, as it is to the meaning ‘hammer’, for example, and, therefore, it is neither of those two meanings. I will return to this central issue in a moment.

Now, let us look at the first syntax of children. Of course, a large part of this first syntax will not provide any clues to the historical origin. A lot of a child’s utterances and of the messages directed at him are aimed at helping the child to learn grammaticalized language. For example, the mother says ‘Look who’s there!’, ‘There!’, and, then she, or the child, adds, ‘Let’s go!’, and finally the mother recapitulates ‘Let’s go there!’. That is, after ‘There!’ and ‘Let’s go!’, we arrive at the grammatical composition: ‘Let’s go there!’ However, in the historical origin there would have been no need for the conversion of those two messages into one syntactic composition, since in this case the second message is unambiguously understood: Here, the automatic inference of the link between the two messages is enough. Obviously, utterances like these cannot throw any light on the historical origin of syntax. However, other kinds of utterances of children are more promising.

Let us focus on the following example (which I really observed). An adult and a child are playing with wooden blocks. Later, the adult requests the child, who has the box with the blocks in his hands, ‘Give me more blocks! More!’. The child, who still has the box in his hands, sees that it is empty and says ‘More, no’.

In the initial, pregrammatical syntax of children, the rhema often consists of a particular kind of ‘metalinguistic’ (Horn, 1989) or ‘metarepresentational’ (Wilson, 2000) negation: the negation does not refer to a predicate but to the previous message of the hearer. In the example, the ‘no’ serving as rhema resembles the negation used for rejections of someone else’s requests or invitations (Dimroth, 2010), but with that ‘no’ what is rejected is the false belief involved in the speech act of the hearer.<sup>1</sup>

<sup>1</sup> Children well under the age of 4 can perceive wrong beliefs of other individuals. Mainly from Onishi and Baillargeon, 2005, this claim is no longer a fantasy for the researchers in the ‘Theory of Mind’.

In this example the child repeats the hearer's message. If the hearer's previous message were not repeated, the negation would be interpreted as a refusal of the request. In other words, here the only way for the child to indicate—or, at least, suggest—that the 'no' is uttered to correct or update the belief involved in the hearer's previous message is to repeat the core of this message. On the contrary, in the case of B rejecting a previous request of A, or in the case of B supporting or continuing ('Let us go!') the previous message ('There!'), there is no need to go beyond holophrases.

In short, only when 1) the belief involved in the hearer's previous message must be corrected or updated, and 2) there is not yet any grammaticalization—when both conditions concur—, the repetition of the hearer's previous message is compulsory. In order to see the effect of grammaticalization, we can think of the questions. Thanks to our sophisticated grammar resources, our questions designate exactly the unknown we are asking about. In this manner, the answer will unambiguously be interpreted as the unveiling of that unknown, although only the rhema need be present. But we are interested in the first condition—in the original need of syntax. In a child, the repetition of a previous message and the correction of the hearer's false belief tend to co-relate, however inexact the co-relation. (Children will also use 'More, no' to reject the command/invitation to eat more pap, although in this case a rejecting holophrase would have been enough.) But my point is that in the historical origin, when there had never been any previous combination of words, such a repetition would show clearly the novelty of the process—the correction of the hearer's belief.

Let's see again the essential difference: The holophrase in which the hearer's previous message is supported or rejected *versus* the original, indispensable syntax (whose first part is a copy of the hearer's message). In the first case, the previous message has already become a favourable or unfavourable element of the current reality. In the second one, the previous message continues to be just a message.

Returning to the analysis of the wooden-block example, the 'more' uttered by the first speaker (a 'more' that could be expressed in holophrastic language) becomes the 'more' which functions as the *thema* for the second speaker. Both 'more' have the same articulatory pattern, meaning and concrete referent, and both involve the false belief that there is still some left of what has been requested. However they are also different. The second 'more' does not work as a request. Instead, it is the platform to which the 'no' will be attached.

In my view, the second ‘more’, which functions as a thema, would not be simply part of a message. In the original syntax, that ‘more’ still maintains its whole-message status, since it mirrors the hearer’s message. Admittedly, that whole-message status is now embedded in another message, but that does not mean that it has disappeared.

In other words, there is recursive embedding of someone else’s message in one’s own message. (‘Message inside Message’: ‘X inside X’.)<sup>2</sup> Certainly this embedding is not a grammatical, Chomskyan recursion.<sup>3</sup> However it can be included in Corballis’ (2011) ‘recursive mind’, if we add that primitive recursions would be interpersonal. Thus, pregrammatical, original syntax would be an early stage of the recursive mind.<sup>4</sup>

Certainly the classic ‘embedding of one message in another’ is grammatical reported speech. But this is really an embedding of sentences, and must have been a much later conquest. In reported speech neither the speaker, nor the place nor the time of the original message are present. Consequently, these elements must be provided through linguistic means. More importantly, this classic embedding requires both syntactic subordination and the meaning ‘say/tell’, which most probably appeared at a much later time than the point when humans started saying things. By contrast, the predications of answer—the ‘More, no’ example—do not have such demanding requirements. In short, the kind of ‘embedding of one message in another’ considered up to now as paradigmatic could in no way have been the original one. Despite these differences, both kinds—this is what mainly interests me—share three crucial features: interpersonal repetition, syntactic advance and the weakening of the embedded message. Let us see the weakening. The predication ‘The Earth is flat’ that is reported in ‘The ancients said that the Earth was flat’ is seen by the speaker as a mere mental state—a mental state of somebody else.<sup>5</sup> In ‘More, no’,

<sup>2</sup> Levinson (2013, p. 154) also connects recursion and pragmatics, but in a different way. Thus, the alternative explanation that Legate et al. (2014, p. 524) propose in their criticism against Levinson cannot be applied to my examples.

<sup>3</sup> Even ‘recursive Merge’—«the most simple recursion», according to Hurford (2012, p. 51), although other authors, as Bickerton (2009, p. 6), do not accept this redefinition of recursion—requires more than two words. Anyway, there is an advantage (which has always fascinated me) for the older definition of recursion. «This shares an essential ingredient with some uniquely developed human trait—language about language, thought about thought, converse about each other’s thoughts»: Hurford (2012, p. 575). Or, in a more general formulation, recursive processes are crucial for human mind: Corballis (2011).

<sup>4</sup> But the earliest recursion could be prelinguistic (Bejarano, submitted).

<sup>5</sup> It is thanks to this loss of referential strength that the whole composition can be true. This weakening—a Fregean point—has been reformulated in an evolutionary perspective (Gilbert, 1991; Jerison, 1988).

the ‘more’ is also weakened: it has stopped working as a request, and it is seen by the speaker as a mere repetition of the hearer’s message.

The step from holophrastic language to syntax must have been very difficult. The meaning of words had to be integrated into an upper-level unit and contribute to a new meaning which had never been learnt before. This is a familiar issue—Humboldt’s “infinite use of finite means”, Chomsky (1965)—, which can be associated with “the expanded working-memory” (Coolidge & Wynn, 2005). But, in my view, we must also look at a second, more basic source of difficulties: Holophrastic signs, in order to become true words, have to suffer a semantic weakening. If the genuine word (the noun ‘hammer’, for example) can serve as efficiently for any function—to request, narrate, ask, say how you say it, etc.—, it is because its meaning no longer includes its function. By contrast, the holophrastic sign was inseparable from its function and its strength. The request was not the meaning of a use of that sign, but it was the meaning of the sign itself. The loss of that strength and that self-sufficiency—the gain in versatility—is tantamount to the origin of syntax. In other words, ‘semantic weakening’ or ‘bleaching’—a central concept in grammaticalization studies—can be extrapolated to the very origin of syntax, i.e. to the original, interpersonal recursion. The original, drastic bleaching would have been the crucial difficulty. Schafer and colleagues (2013) have found that «words that can be associated with many words are underrepresented in the comprehension vocabulary when a weak central coherence can be attributed to the patient». In a similar way, the change from comprehension of holophrastic sign to comprehension of genuine word would require an enormous increase of central coherence.<sup>6</sup>

Obviously, there are two hypotheses about this issue. On the one hand, many authors—not only Fodor (1975) and his extreme innatism, but also Givón (2002) or Hurford (2003) – accept that, previously to language learning, thought already possesses a compositional, syntax-like format. On the other hand, the original syntax would derive from the bleaching of holophrases that is achieved in the dialogue.

According to this second hypothesis, prelinguistic perception and thought lack syntax-like compositionality. It is indisputable that within a prelinguistic perception there are agents, action, quality. However within a perception none

<sup>6</sup> The difference between words and holophrastic signs can be reinforced if we accept—cf. Taylor (2012)—that the most properly linguistic part of the meaning of a word is constituted by the past episodic links of this with other words.

of these features would be addressed separately. That is, at the conscious perception level the scene would be captured as a unity, not as a composition of features, each receiving separate attention. Of course, the construction of perception integrates many features and each of these features is captured by different groups of neurons. Some features, such as line direction or colour, are captured in the first stages of visual processing. (In these features we know for sure that a ‘perceptual binding’ takes place and it is thanks to it that we do not perceive ‘red’ and ‘square shaped’ separately but a ‘red square’.) By contrast, other features—which can well be called ‘concepts’—are captured at a later processing level of the scene. However, my point is that neither of these features is attended to in the way the successive meanings within a syntactic composition are attended to.

In addition, according to the second hypothesis, these features are not the original root of words, in spite of the fact that they are present in the animal’s mind. Primitive holophrastic signs had no relation to these features (nor had they any relation to the ‘parts of the sentence’), because their aim was not to mirror reality. They were only used for calling or requesting. Certainly, after interpersonal recursion and consequent bleaching, meanings would have come nearer to those features, and, as a result, these features could progressively be separated from the perceived whole.<sup>7</sup> For example, a holophrastic sign whose only original function was to act as a vocative to call a particular individual, after the genesis of syntax, would have become a word for that individual, in increasingly closer resemblance to the prelinguistic ‘concept, or recognition pattern, of that individual’. However, according to the second hypothesis, we cannot presume that that close resemblance was in the origins of language. *Separately attendable* concepts did not exist in prelinguistic minds. Instead, their genesis took place at a later stage and was mediated by interpersonal communication. It is thanks to these concepts that we can ‘decompose and recombine’— «we disassemble the world, and then create alternative versions of reality» (Tattersall, 2009, p. 586). Certainly, syntactic combinations produce new wholes, some of which can in the end be similar to prelinguistic perceptual

<sup>7</sup> This separation might have also been generated at the level of features captured in the first stages of visual processing. “Artificially produced geometric forms (75000 bp) serve as an externally derived supernormal correlate of the patterns processed by the early visual system leading to a proto-aesthetic pleasure (or perceptual fluency not determined by familiarity)” (Hodgson, 2014). See also Arden et al. (2014; I emphasise p. 1845 – “Do not mention any of the body parts”).

wholes. However the new whole, even if it mirrors a trivial scene, has been “created” by human skill, is an embedding, upper-order unity.

In other words, in holophrastic language there was no neutral meaning which could subsequently be used either to represent the world (‘declarative’ use), or to try to change it (‘imperative’ use). Instead, there was request intonation or call intonation within which phonetic articulation would have progressively developed to make more specific the object of the request or call. The road to neutral, versatile meaning was long and arduous.

To be more exact, instead of two opposite hypotheses it would be better to say that the second one makes the way suggested by the first one more indirect and complex. ‘Concepts similar to those of an animal mind + communicative signs representing no concept at all + bleaching > Syntax’: the two intermediate steps are exclusive to the second hypothesis. On one hand, this hypothesis may seem less parsimonious, on the other, this hypothesis would better correspond with the enormous difference between human and non-human primates.

To make a more concrete contrast between the hypotheses, let us concentrate on denials. According to the first hypothesis (more specifically, according to a non-innatist version of this hypothesis), denials might have originated from the surprise caused by the absence of an expected element (Davidson, 1982, p. 318): the surprise would have led to the denial of the wrong expectation or prediction held previously. Accordingly, the origin of denials would have been intrapersonal. On the other hand, the second hypothesis supports the increasingly accepted idea that the surprise or ‘prediction error’ is dealt with using a much less costly resource than a denial: the incorporation of the latest information into the perceptive content will suffice. There is no need to deny the wrong prediction, because this disappears when the latest information arrives. Consequently, in the light of the second hypothesis, denials would have originated in interpersonal communication. In the beginning, i.e., before syntax, only received requests would be denied. Later, in concurrence with the genesis of syntax, the speaker would deny the hearer’s wrong beliefs: denial as rhema. In time, true negative statements can become mere representations of reality.<sup>8</sup>

<sup>8</sup> Certainly many exemplars of this type of statements become mere representations of reality. However, if the interpersonal origin of the type is forgotten, the Bergsonian question arises: where does the characteristic negated come from?

I do not believe we have the resources today to reject either the existence of prelinguistic syntax or the opposite. The only thing that we can do in a case such as this is ‘wager and see’.<sup>9</sup> Thus, I wager on the original bleaching and the interpersonal genesis of syntax.

The question then arises as to whether there was anything that could mitigate those difficulties in the original bleaching. Looking for an answer, Section 2 will make the above proposal more concrete. ‘Embedding your holophrastic message not only in my message but also in my intonation’: This is the new schema.

## 2. The two levels of embedding and their respective audiomotor resources: Embodied embedding

At the articulatory level, the two ‘more’ are identical. But, apart from that identity, there is an intonation difference between them. One of the functions of intonation, both now and—even more importantly—before grammaticalization, is to indicate the end of the message. At the holophrastic stage, the end of the message coincides, obviously, with the end of the word: the intonation unit and the articulatory-semantic unit run parallel. But later, with the emergence of syntax, the intonation pattern becomes capable of comprising two or more words.

<sup>9</sup> When did “declarative” communication emerge in evolution? Given that I do not consider the cries of vervets to be declarative communication, let me be more precise: When—where in the bridge between our nonhuman ancestors 7 million years ago and modern human—did declarative pointing arise? Please note that, in the case of an early emergence of declarative pointing, declarative communication would already have been deeply established when the first articulated holophrases—undoubtedly, very close to human language—began and would therefore have most likely impregnated (against the second hypothesis) these holophrases. Could this be a path for the second hypothesis soon to become a falsifiable hypothesis? I am inclined to accept the three following claims. 1) The white of eye is adaptive because it facilitates the understanding of pointing gestures as communicative. 2) The declarative pointing did not arise later than the imperative one. In my view, chimpanzees can realize that a caged—or physically disabled—companion desires an object, without them understanding as communicative the—really communicative—unlearned gesture of the companion of getting as close as possible to the desired object. So, the true understanding of pointing would appear in evolution simultaneously with the declarative use of gestures. 3) The origin of the white of the eye could be soon clarified. In short, I suggest linking the first hypothesis with an early date for the white of the eye and the second one with a late date. In Bejarano (2011, chapters 3, 4, 6) and Bejarano (submitted), I propose that the true understanding of pointing gestures, since it involves a demanding requisite, arose extremely late (in Neanderthal, if not in Sapiens). If, on the contrary, paleogenetics discovers that the white of the eye had already appeared earlier, then that result would be unfavourable regarding the second hypothesis. It is true that this suggestion depends on other theoretical crossroads and, consequently, it does not really lead to the falsifiability of the second hypothesis. However, it can be seen as a step in that direction.

In ‘More, no’, the intonation comprises the whole of the two-word message, whereas the one-word message of the hearer is only repeated at articulatory level and not at intonation level. Let us say it more concretely. In ‘More, no’, ‘more’ is identical to the hearer’s previous message, and remains a complete message. But at the same time, that ‘more’ occupies the first part of the intonation pattern and therefore has become a part of a message. In other words, the two levels of that recursive embedding respectively have recourse to two different audiomotor resources. The upper-order level of embedding is represented in the intonation, and the lower-order level, in the phonetic articulation of the word.

Thus, original syntactic recursion would have been not only interpersonal, but also embodied. ‘Embodied cognition’ is a multifaceted approach, and some of its elements are controversial. But there is a wide acceptance of its oldest, quasi-Piagetian nucleus. It is in this sense that I use ‘embodied’. The recursive embedding would be intrinsically associated with those two sensorimotor patterns, and in this way it would be facilitated, both in children (see Keitel et al., 2013) and in historical genesis.

If my proposal is correct, then the decoupling of intonation and articulated-semantic pattern, i.e., the emergence of intonation capable of comprising more than one word, is a crucial milestone.<sup>10</sup> We do not know when the decoupling took place. But there is little doubt that it had a firm basis.

The evolutionary root of intonation probably consists of primate cries, in which the continuum of intonation intensity mirrors the continuum of emotional intensity. Later, after the genesis of syntax, intonation certainly becomes less rooted and less emotional. But in the beginning, its production and reception are disconnected from learning. On the other hand, articulatory patterns depend on imitative learning of new motor sequences, and we know that neither syllabic abilities, nor—in the general motor field—the imitation of new sequences (which is the only useful motor imitation) are developed in non-human primates. In short, there is a strong contrast between intonation and articulation. Side by side, however, as was to be expected, antecedents of articulation are finally beginning to appear in non-human primates (vs. songbirds: Rosselló, 2014), some facial expressions with a speech-like rhythm

<sup>10</sup> Obviously, this decoupling does not mean a disconnection. The stressed syllable of the rhema becomes the peak of the intonation of the sentence. More important, as a reviewer wrote, “tone languages show that the division of audiomotor resources is partly influenced by culture. Still, it is probably right to say that there’s a strong underlying basis for the division”.

(Ghazanfar & Takahashi, 2014), or antecedents of voiceless consonants (Lameira et al., 2014).

We must also take into account the hemispherical specialization of the brain. Articulatory, learned patterns are mostly associated with the left hemisphere: Broca's discovery has been qualified over time, but its essential truth remains. By contrast, intonation (or "the system involving recognition of pitch contours") does not seem to have such a strong hemispherical preference (Peretz & Hyde, 2003). This must have facilitated—or rather, consolidated—the crucial decoupling.

Both in children and in history, the articulatory-phonetic pattern would have been the first and foremost cause for the decoupling of articulation and intonation. With regard to articulatory sequences, de Boer and Zuidema (2010) emphasize a very interesting distinction: 'true segmentation in the producer' vs. (more primitive) 'superficial combinatorial structure'. They characterize the superficial one as «combinatorial structure that can be observed by an outside observer in a system of signals, but that is not actively used by the agents using the signals» (Boer & Zuidema, 2010, p. 144). Very probably, this distinction is valid both for primaevial language and for children. Sequences of babbling, and even the earliest holophrases, would be 'superficial combinatorial structures'. Returning to our argument, the 'true' learning of articulatory-phonetic sequences is very different from (originally emotional, holistic) intonation, and this difference could have triggered the decoupling of intonation and articulated pattern. Nowadays (but, in my view, not in the very historical origin!) protodeclaratives (pointing + holophrase without any innate intonation), which only communicate in order to learn (Begus & Southgate, 2012) or teach words, facilitate that decoupling in children.

I am focusing on the consequences of this decoupling. «The division of labor between prosody and segmental sounds may shed an important light on the evolution of language» (Chafe, 2000, p. 253). This Section has suggested one consequence of that 'division of labor': original syntactic recursion would have been an embodied, facilitated recursion. Now, from that consequence I am going to extract another consequence.

### 3. Intonation, syntax and the victory of voice over gesture

The articulatory patterns of words are produced by motor imitation of models on the part of the learner. The same motor imitation involved in the learning of words also takes place in pretending (communicative pantomime and children's symbolic play): In pretending, since the reproduction is executed without the proper object or context, any action becomes a genuine motor imitation. This similarity between words and iconic gestures is in fact a very strong one. Only in these two types of patterns is motor imitation maximally implemented. In technical tasks, although they have been learnt from a model, other factors besides motor imitation intervene. Goal-imitation (aka 'mere emulation'), commonly associated with animals, also occurs in humans. More precisely, it has been observed that accuracy in genuine motor imitation decreases when there are objects, or, stated in the reverse, the temptation to turn to mere emulation grows stronger when the movement is acting on objects (Gattis et al., 1998). Moreover, the size, shape, weight, etc. of objects impose constrictions on the degrees of freedom of the motor system regardless of the model. In short, it is only in imitations performed without the handling of objects that the motor pattern is really under the control of the model. Thus, articulatory-phonetic imitation and communicative pantomime are very closely related. It is likely that gestures were as important as voice for a long period of time, or perhaps even more important (see Arbib 2012; Corballis, 2002; Donald, 1991). Also, although pointing is different from iconic gestures, handedness (left hemisphere) in toddlers' pointing is connected to this issue (Vauclair & Cochet, 2013).

However, when the need for syntax arose, intonation would have given the voice the decisive edge. Intonation and articulated pattern, which are heard at exactly the same time and in the same sequential linearity, nevertheless may not coincide in their respective units, which, I have suggested, facilitates the embedding of a whole in a new whole, and, consequently, syntax. Deprived of that edge, gestures have eventually become merely a minor, subordinated means.<sup>11</sup>

The rest of the advantages traditionally invoked in favour of the voice do not seem at all decisive. If vocal language can take place in the dark, gestural language can claim to be able to take place in silence, which can sometimes be

<sup>11</sup> But «they can perform many of the functions of natural language» (Goldin-Meadow, 1999, p. 420), even in the thought itself (Cartmill et al., 2012; Warburton et al., 2013).

useful. Busy hands—that is, hands that cannot perform other tasks – are certainly a disadvantage for any gestural language, but only in extended productions, which are unlikely until relatively recent historical times. In addition, analogy does not necessarily have to have been a limiting factor for gestures. Continuous use would have made them progressively less analogical (Tomasello, 2008, p. 223; Smith & Höfler, in press).

In short, intonation would be the major cause responsible for the overwhelming success of the voice. The emergence of spatial verbal deictics—‘there’, for example—is a sign of this success. At first these deictics would have been unnecessary; a pointing gesture can fulfil their function. However, they came into being.<sup>12</sup> Most probably the reason is this: spatial indications had to be verbalized if they were to be integrated into a syntactical combination. This happens quite soon in children. Certainly children, in the very beginning of their post-holophrastic stage, produce some one-word messages which, accompanied by a pointing gesture towards an object, designate a quality or fact about the object: Here, there are not yet any verbal deictics. But children soon begin to include verbal deictics, mainly in contrastive functions. In short, syntax provokes a strong preference for the vocal channel.

Given the growing primacy of voice, language only by gesture has been relegated to communication with foreigners or the deaf. True languages of the deaf, which are predominantly manual, use another commonality, besides their shared relationship with motor imitation. This second commonality is that one can see one’s own hands (more easily than any other part of one’s own body), and one can hear one’s own voice.<sup>13</sup> This self-perceptiveness greatly facilitates imitative learning, since the model and the reproduction adopt a common perceptual format for the learner.<sup>14</sup> However, despite these two commonalities of vocal language and manual language of the deaf, there are significant differences. In manual languages of the deaf, the prosody is certainly produced not only by face and head, but also by hands. At the end of an intonational phrase, the hands carry phrase-final lengthening (Sandler et al., 2014, p. 254). However even this manual prosody is unable to embed the sequence of gestures-words. On the contrary, intonation can embed the sequence of words. «Intonation is spoken, and that connects it directly with the stream of speech» (Bolinger, 1983, p. 70). For that reason, oral language and manual language

<sup>12</sup> “At layer 4”: Heine & Kuteva (2007).

<sup>13</sup> During her learning, Helen Keller also perceived the results of her movements.

<sup>14</sup> Is there another consequence of self-perceptiveness? See Bejarano (2011, chapter 9).

are differently efficient in the process of providing the upper order, in which the different meanings must be embedded, with embodiment. In vocal language, the facilitation of syntax is more efficient.

Obviously, recursion and syntax are not unknown to the deaf. Indeed, deaf people have used their linguistic abilities in the socio-cultural process of generating new, independently developed languages (Senghas et al., 2004; Sandler et al., 2014). In more general terms, recursion, which was interpersonal and embodied in origin, subsequently became intrapersonal (to use the Vygotskian term) and began to rely on historical developments—grammaticalization or writing. But my point is that in the origin of syntax—in the original bleaching—, not only the repetition of the hearer’s message but also the two different audiomotor resources of vocal language were decisive resources.

#### ACKNOWLEDGEMENTS

I am very grateful to an anonymous reviewer whose comments have helped me in a crucial way.

#### REFERENCES

- Aitchison, J. (1998). *The Seeds of Speech*. Cambridge: Cambridge University Press.
- Arden, R. et al. (in press). Genes Influence Young Children’s Human Figure Drawings and Their Association With Intelligence a Decade Later. *Psychological Science*.
- Begus, K. & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, 15, 611-617
- Bejarano, T. (2011). *Becoming human: From pointing to syntax*. Amsterdam: Benjamins.
- Bejarano, T. (submitted). In Your Shoes When You Look at Me: Clear-cut human uniqueness.
- Bickerton, D. (2009). Syntax for Non-Syntacticians. In D., Bickerton & Szathmáry (eds.), *Biological Foundations and Origin of Syntax*. Cambridge: MIT Press, 3–14.

- Bolinger, D. (1983). Intonation and Gesture. *American Speech*, 58, 156–174
- Cartmill, E., Beilock, S. & Goldin-Meadow, S. (2012). A word in the hand: Action, gesture, and mental representation in human evolution. *Philosophical Transactions of the Royal Society, B*, 367, 129–143
- Chafe, W. (2002). A memoir. *Historiographia Linguistica*, 29, 245–261.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Coolidge, F. & Wynn, T. (2005). Working memory, its executive functions, and the emergence of modern thinking. *Cambridge Archaeological Journal*, 15, 5–26.
- Corballis, M. (2002). *From Hand to Mouth*. Princeton: Princeton University Press.
- Corballis, M. (2011). *The Recursive Mind: The Origins of Human Language, Thought and Civilization*. Princeton: Princeton University Press.
- Davidson, D. (1982). Rational animals. *Dialectica*, 36, 317–327.
- de Boer, B. & Zuidema, W. (2010). Multi-agent simulations of the evolution of combinatorial phonology. *Adaptive Behavior*, 18, 141–154.
- Dimroth, C. (2010). The acquisition of negation. In L. R. Horn (ed.), *The Expression of Negation*. de Gruyter, 39–72.
- Donald, M. (1991). *Origins of Human Mind. Three Stages in the Evolution of Culture and Cognition*. Cambridge: Harvard University Press.
- Fodor, J. (1975). *The Language of Thought*. Cambridge: Harvard University Press.
- Gattis, M., Bekkering, H. & Wohlschläger, A. (1998). When actions are carved at the joints. *Behavioral and Brain Sciences*, 21, 691–692.
- Ghazanfar, A. & Takahashi, D. (in press). Facial Expressions and the Evolution of the Speech Rhythm. *Journal of Cognitive Neuroscience*.
- Gilbert, D. (1991). How mental systems believe. *American Psychology*, 46, 107–119
- Givón, T. (1979). *On understanding grammar*. New York: Academic Press.
- Givón, T. (2002). The visual information-processing system as an evolutionary precursor of human language. In T. Givón & B. F. Malle (Eds.), *The Evolution of Language out of Pre-language*. Amsterdam: Benjamins, 3–50.
- Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3, 419–429
- Heine, B. & Kuteva, T. (2007). *Genesis of Grammar*. Oxford: Oxford University Press.

- Hinterwimmer, S. (2012). Information Structure and Truth-Conditional Semantics. *Semantics*, 1875–1908.
- Hodgson, D. (2014). Decoding the Blombos Engravings. *Cambridge Archaeological Journal*, 24, 57-69.
- Horn, L. (1989). *A Natural History of Negation*. Chicago: Chicago University Press.
- Hurford, J. (2003). The neural basis of Predicate-Argument Structure. *Behavioral and Brain Sciences*, 26, 261-283.
- Hurford, J. (2007). The origin of noun phrases. *Lingua*, 117, 527–542.
- Hurford, J. (2012). *The origins of grammar*. Oxford: Oxford University Press.
- Jackendoff, R. (2003). *Foundations of Language*. Oxford: Oxford University Press.
- Jerison, H. (1988). Evolutionary neurology and the origin of language. In M. Landsberg (Ed.), *The genesis of language*. de Gruyter, 3-9.
- Keitel, A. et al. (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology*, 116, 264–277.
- Lameira, A., Maddieson, I. & Zuberbühler, K. (in press). Primate feedstock for the evolution of consonants. *Trends in Cognitive Sciences*.
- Legate, J., Pesetsky, D. & Yang, C. (2014). Recursive misrepresentations: A reply to Levinson. *Language*, 90, 515-528.
- Levinson, S. (2013). Recursion in Pragmatics. *Language*, 89, 149-162.
- Onishi, K. & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308, 255–258.
- Peretz, I. & Hyde, K. (2003). What is specific to music processing? *Trends in Cognitive Sciences*, 7, 362–367.
- Rosselló, J. (2014). On the separate origin of vowels and consonants. In B. de Boer, & T. Verhoef (Eds) *Evolang X - Workshop on Signals, Speech and Signs*, 29-33 <http://ai.vub.ac.be/~bart/ProceedingsSignalsSpeechSigns> (accessed 16 July 2014).
- Sandler, W., Aronoff, M., Padden, C. & Meir, I. (2014). Language emergence. In N. Enfield, P. Kockelman & J. Sidnell (Eds.), *The Cambridge Handbook of Linguistic Anthropology*. Cambridge: Cambridge University Press, 246-278.

- Senghas, A., Kita, S. & Özyürek, A. (2004). Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science*, *305*, 1779–1782.
- Schafer, G., Williams, T. & Smith, P. (2013). Which Words are Hard for Autistic Children to Learn? *Mind & Language*, *28*, 661–698.
- Smith, A. & Höfler, S. (in press). Metaphor as the driving force of language evolution. In *Proceedings of EVOLANG 2014*.
- Tattersall, I. (2009). Becoming Modern *Homo sapiens*. *Evolution: Education & Outreach*, *2*, 584–589.
- Taylor, J. (2012). *The Mental Corpus: How language is represented in the mind*. Cambridge: Oxford University Press.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge: MIT Press.
- Vauclair, J. & Cochet, H. (2013). Hand preference for pointing in toddlers. *Developmental Psychobiology*, *55*, 757–765.
- Warburton, E. et al. (2013). The Cognitive Benefits of Movement Reduction. *Psychological Science*, *24*, 1732–1739.
- Wilson, D. (2000). Metarepresentation in Linguistic Communication. In D. Sperber (Ed.), *Metarepresentations*. Oxford: Oxford University Press, 411–441.



# The Gradual Evolution of Language

*Michael C. Corballis*<sup>†</sup>  
m.corballis@auckland.ac.nz

## ABSTRACT

Language is commonly held to be unique to humans, and to have emerged suddenly in a single “great leap forward” within the past 100,000 years. The view is profoundly anti-Darwinian, and I propose instead a framework for understanding how language might have evolved incrementally from our primate heritage. One major proposition is that language evolved from manual action, with vocalization emerging as the dominant mode late in hominin evolution. The second proposition has to do with the role of language as a means of communicating about events displaced in space and time from the present. Some have argued that mental time travel itself is unique to human, which might explain why language itself is uniquely human. I argue instead that mental time travel has ancient evolutionary origins, and gradually assumed narrative-like properties during the Pleistocene, when language itself began to take shape.

Keywords: evolution, gesture, hippocampus, language, mental time travel, mirror system, speech.

## 1. Introduction

He thought he saw a Rattlesnake  
That questioned him in Greek.  
He looked again, and found it was  
The Middle of Next Week.  
“The one thing I regret,” he said,  
“Is that it cannot speak!”

—from *The Gardner’s Song*, by Lewis Carroll

Language poses a substantial problem for the theory of evolution. It is a complex faculty, yet seemingly unique to our species. Members of other species

<sup>†</sup> School of Psychology, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand.

communicate with one another, but their communications seem to lack both the open-ended quality and propositional structure of human language, and their vocal communications also lack the purposeful, intentional character of human discourse. Nonhuman species do not seem to exchange information about events, or about the nature of the world; they do not tell stories. According Darwin's theory of evolution through natural selection, complex traits emerge incrementally, in piecemeal fashion, yet language seems distinctive in that appears to be fully formed as a complex, integrated whole. The task of explaining how language might have emerged through Darwinian evolution is therefore a challenging one.

Some have responded to the problem by proposing that language did not evolve in Darwinian fashion, but instead emerged in a single step. In the view of Noam Chomsky, the most prominent of contemporary linguists, this was indeed a profound change, and one that could not have occurred incrementally. It was, moreover, merely a by-product of a more fundamental restructuring of the human mind, involving two "paths," one intellectual and one moral:

One path seeks to understand more about language and mind. The other is guided by concerns for freedom and justice. There should be some shared elements, in particular, what the co-founder of modern evolutionary theory, Alfred Russel Wallace, called 'man's intellectual and moral nature': the human capacities for creative imagination, language and symbolism generally, interpretation and recording of natural phenomena, intricate social practices and the like, a complex of capacities that seem to have crystallized fairly recently among a small group in East Africa of which we are all descendants, sometimes called simply 'the human capacity'. The archaeological record suggests that the crystallization was sudden in evolutionary time. Some eminent scientists call the event 'the great leap forward', which distinguished contemporary humans sharply from other animals (Chomsky, 2007, p. 3).

Despite Chomsky's reference to Wallace, the notion of the "great leap forward" is profoundly at odds with evolutionary theory, which holds that evolution occurs in small increments. But Chomsky's objection to an evolutionary account runs deeper. He suggests that language cannot have evolved through natural selection because the symbols and concepts we use have no external reality, and therefore could not have been shaped by environmental contingencies. The emergence of what he calls I-language—the internal language of thought—must have occurred entirely within the brain, without reference to the external

environment. In this respect, «natural language diverges sharply ...from animal communication, which appears to rely on a one–one relation between mind/brain processes and ‘an aspect of the environment to which these processes adapt the animal’s behavior’» (Chomsky, 2007, p. 10). Language as a form of communication is in this view simply a by-product of a fundamental restructuring of human thought. It depends on the invention of E-languages—external means of communication—that allow one person’s thoughts to map onto the thoughts of another. Language depends on so-called “theory of mind,” the understanding of what others are thinking, rather than on reference to the external world.

This notion of a restructuring—the “great leap forward”—has been supported by a number of archaeologists, based on evidence from hominin artifacts of a profound transformation at some point within the past 100,000 years, perhaps even as recently as 50,000 years ago (Klein, 2008), with striking advances in technology and evidence of symbolic representation. The archaeologist Ian Tattersall (2012) writes:

Our ancestors made an almost unimaginable transition from a non-symbolic, nonlinguistic way of processing information and communicating information about the world to the symbolic and linguistic condition we enjoy today. It is a qualitative leap in cognitive state unparalleled in history. Indeed, as I’ve said, the only reason we have for believing that such a leap could ever have been made, is that it *was* made. And it seems to have been made well *after* the acquisition by our species of its distinctive modern form (Tattersall, 2012, p. 199).

Such proclamations have an almost Biblical sweep, perhaps owing more to wishful thinking than to a critical appraisal of the evidence. Indeed, not all are agreed that a sudden transformation took place within the past 100,000 years. Some have proposed a more gradual development of tools and other artifacts from the Middle Pleistocene, which dates from around 750,000 years ago (McBrearty, 2007; Shea, 2011). The “great leap forward” also seems to deny Neanderthals human-like cognition and language, yet the Neanderthals had brains as large as those of humans, perhaps slightly larger, and we know that early humans did interbreed to some extent with the Neanderthals before they died out some 30,000 years ago (Green et al., 2010). A recent review suggests that Neanderthal language

and culture may not have differed substantially from those of *Homo sapiens*, even raising some question as to they were actually distinct species (Johansson, 2013). Yet if we ignore the limited amount of interbreeding, we must go back some 500,000 years to find the common ancestor in the Middle Pleistocene. It seems unlikely that a sudden genetic transformation within the past 100,000 years so transformed *Homo sapiens* as to create a species cognitively and morally unique, and sharply differentiated from the equally large-brained Neanderthals.

The notion of the great leap forward, then, is counter to evolution by natural selection, as proposed by Darwin. Indeed, if true, it might even herald the demise of the Darwinian theory itself, for Darwin (1859) himself wrote in *Origin of Species* as follows:

If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down. But I can find no such case (Darwin, 1859, p. 158).

Is language, then, the case that Darwin feared? But as Pinker and Bloom (1990) pointed out, we should not give up an evolutionary account lightly; to them the idea that language evolved through natural selection is inescapable:

The only successful account of the origin of complex biological structure [such as language] is the theory of natural selection, the view that the differential reproductive success associated with heritable variation is the primary organizing force in the evolution of organisms (Pinker & Bloom, 1990, p. 708).

## 2. Toward an evolutionary account

### 2.1 Nonhuman animals don't lack concepts

As noted above, Chomsky argues that human concepts and symbols have no direct reference to the external world, and that this distinguishes them from various sounds and movements that underlie animal communication. Evidence from animal communication suggests otherwise. Even Darwin (1872) noted that animals can effortlessly learn to associate myriad sounds and signals with 'general ideas or concepts' (p. 83). These sounds and signals, moreover, typically bear no direct relation to the aspects of the world they signal, and in that respects are like

spoken words. For instance Cheney and Seyfarth (1990) famously observed that vervet monkeys produce a variety of different calls to indicate different predators, such as a snake, leopard, or eagle, suggesting that the calls have different meanings. It is now known that many other primate species, including chimpanzees, produce different calls to express different meanings (see Cheney & Seyfarth, 2005, for review).

Although primate calls suggest a degree of conceptual representation, they seem to lack the flexibility and intentionality of human language, where there seems no limit to the number of things—objects, qualities, actions, emotions—that we can name. But the conceptual repertoire of other animals is much more evident in their ability to comprehend than in their vocal productions. Savage-Rumbaugh et al. (1998) reported that Kanzi, a bonobo, was able to follow spoken instructions in English, made up of several words, at a level comparable to that of two-and-a-half-year-old child. Kanzi is now said to understand some 3,000 spoken words (Raffaele, 2006). Even domestic dogs can rapidly learn the meanings of spoken words, even though they cannot themselves articulate them. A border collie known as Rico responds accurately to spoken requests to fetch different objects from another room, and then either place the designated object in a box or bring it to a particular person. If Rico is asked to fetch an object with an unfamiliar name, he fetches an object he has not encountered before and thereafter knows that object's name—a phenomenon consistent with what has been termed “fast mapping” (Kaminski et al., 2004). Based on similar studies, another border collie called Chaser is said to respond meaningfully to the spoken names of 1022 objects (Pilley & Reid, 2011).

To Chomsky, the “great leap forward” was fundamentally a transformation of the conceptual capacity, with communication a secondary process. Animal research nevertheless suggests a continuity between humans and other animals, and therefore an evolutionary continuity, in the conceptual repertoire. The primary limitation evident in nonhuman animals' vocal calls is therefore not so much one of representing the world, but rather one of communicating about it. Cheney and Seyfarth (2005) remark that «Animals' limited vocal repertoires are particularly puzzling because they appear to have so many concepts that could, in principle, be articulated» (p. 142).

## 2.2 A vocal limitation

The inability of animals, with the exception of some birds, to produce anything resembling speech, is therefore primarily a vocal limitation rather than one of conceptual understanding. Primate vocal calls probably evolved as “honest signals,” for the most part innately specified, with relatively little scope for intentional control or even vocal learning. This may mean that vocalization was not a natural platform for language itself, which is heavily dependent on the capacity to learn new signals and produce them intentionally. Jane Goodall once wrote that «(t)he production of sound in the absence of the appropriate emotional state seems to be an almost impossible task for a chimpanzee»(Goodall, 1986, p.125). David Premack, another pioneer in the study of chimpanzee behavior, suggests that even chimpanzees, our closest nonhuman relatives «lack voluntary control of their voice» (Premack, 2007, p.13866).

These conclusions do need some qualification. Cheney and Seyfarth (2005) draw attention to examples of primates modifying their vocalizations, sometimes even suppressing them, depending on the audience. Vervet monkeys seldom give alarm calls when they are alone, and are more likely to do so in the presence of kin than of non-kin. Chimpanzees modify their screams when under attack, depending on the severity of the attack and their status relative to that of nearby chimps (Slocombe et al., 2010), and when encountering food chimps emit different kinds of grunts depending on the type of food (Slocombe & Zuberbühler, 2005). Such examples, though, suggest subtle changes within call types rather than the generation of new call types (Egnor & Hauser, 2004). Some modifications involve the face and mouth rather than voicing itself. For instance chimpanzees can modify emitted sounds to attract attention by vibrating their lips, as in the “raspberry” sound (Hopkins *et al.*, 2007), and this call can be imitated by naïve animals in captivity (Marshall *et al.* 1999)—although these sounds depend on movement of the lips rather than of the larynx. Reviewing these and other examples, Petkov and Jarvis (2012, p. 5) write that

[...] we would interpret the evidence for vocal plasticity and flexibility in some non-human primates as limited-vocal learning, albeit with greater flexibility via non-laryngeal than laryngeal control. But they do not have the considerable levels of laryngeal (mammalian) or syrinxal (avian) control as seen in complex vocal learners.

Complex vocal learning, and therefore speech itself, appears to have evolved late in primate evolution, and was possibly restricted to the hominins—and perhaps even to our own species. I discuss this in more detail below, but first it is useful to point out that an inability to speak need not mean that language itself is ruled out, as the signed languages of the deaf remind us.

### 2.3 Manual communication

A more solid basis for intentional communication of learned signals may come from the hands rather than the voice. Great apes have not learned anything approaching speech, but attempts to teach them simplified forms of sign language have been moderately successful. The bonobo Kanzi communicates by pointing to arbitrary signs on a keyboard, representing objects and actions (Savage-Rumbaugh et al., 1998), and his keyboard now has over 300 signs, which he supplements these by inventing gestures of his own. The gorilla Koko is said to use and understand over 1000 signs (Patterson & Gordon, 2001). These examples demonstrate little in the way of grammatical competence, but at least show intentional use of gesture to represent objects and actions, and some limited competence at combining a few gestures to create simple requests. The productive vocabularies of Kanzi and Koko probably still fail to match their conceptual understanding—although this is also true of humans. We often find ourselves at a loss for words.

The evolutionary origins of intentional communication are probably to be found in the primate mirror system, a network in the brain that is active both when the animal observes a particular action and when it performs that action (Rizzolatti & Sinigaglia, 2010). The motor component of this system in primates lies in Area F5, which is homologous to Broca's area, the cortical area critically involved in the production of speech. In nonhuman primates, though, this area appears unresponsive to the vocalizations of conspecifics, but it is responsive to the sounds produced by manual actions, such as the noise of a stick being dropped, or the cracking of peanut shells (Kohler et al., 2002). In this respect it does not merely “mirror” an action, as often supposed, but is involved more generally in perceptuo-motor learning. The relative deafness of the mirror system to vocal production may again reflect the non-intentional aspect of primate vocalization, necessary to maintain vocal calls as honest signals that can't be faked.

## 2.4 Interaction between hand and mouth

Although the primate mirror system does not appear to accommodate vocalization, the perception and control of hand and mouth are tightly integrated. This is true even of the primary motor cortex, located in the precentral gyrus. Since the pioneering work of Penfield and Rasmussen (1950), it has long been held that the primary motor cortex represents simple movements of the body, from which more complex movements are constructed. It is now known that some integrated movements, especially of hand and face, are organized within the precentral gyrus. In monkeys, stimulation of the rostral part of the precentral gyrus elicits coordinated hand-to-mouth movements (Graziano & Aflalo, 2007)—movements already evident in primate newborns (Allman, 2000). In a study of humans aged from 2 to 60 undergoing operations in which the precentral gyrus was exposed, Desmurget et al. (2014) found that stimulation of some sites elicited independent movements of mouth and arm (including hand and wrist), some sites elicited coordinated movements. These included gradual opening of the mouth while the closing hand moved toward the face, as though wanting to bring food to the mouth. This movement seems to be innately programmed; ultrasound recordings show that human fetuses suck their thumbs from as early as the 11<sup>th</sup> week of gestation—and more often such the right than the left thumb (Hepper et al., 2005). Bringing the hand to the mouth seems to be the first coordinated movement to appear in development.

More complex movements, which require learning, depend on premotor areas, of which area F5 is an example. In the monkey, some neurons within this area respond to grasping with the mouth as well as to grasping with the hand (Rizzolatti et al., 1988). Petrides and Panya (2009) also identify neurons in the homologue of Broca's area in monkeys which control orofacial muscles, and identify connections from the parietal and temporal lobes that terminate in that area. They write that their findings “are consistent with suggestions that control of action and gesture may have preceded specialization for language” (Petrides & Panya, 2009, p. 13). Gentilucci has documented close correspondences between hand and mouth movements during speech itself; for instance, when uttering the syllable “ba” the mouth opens wider when the speaker grasps a larger object than a smaller one, or even when the speaker watches another person making these movements (Gentilucci & Corballis, 2006; Gentilucci et al., 2012).

Facial gestures play an important role in sign languages (Emmorey, 2002; Sutton-Spence & Boyes-Braem, 2001), and even normal speech retains a visible

component. This is illustrated by the McGurk effect: A syllable (such as *da*) is dubbed onto a mouth saying another syllable (such as *ba*), and people tend to “hear” what they see rather than what was actually voiced (McGurk & MacDonald, 1976). Other studies show the parts of the brain involved in producing speech are activated when people simply watch silent videos of people speaking (Calvert & Campbell, 2003; Watkins et al., 2003). Ventriloquists know the power of vision over what they hear when they project their own voices onto the face of a dummy by synchronizing the mouth movements of the dummy with their own pursed-lipped utterances.

The visible accompaniments of speech also include expressive movements of the hands and arms. Indeed the distinction between speech and signed languages is not absolute, since speech is universally accompanied by manual gestures, and the tight synchrony between the two suggests that they are controlled by a single integrated system (McNeill, 1985). Experiments show that gestures influence the understanding of speech, just as speech influences the understanding of gestures, so the interaction is mutual and obligatory, implying that speech and gesture are «two sides of the same coin» (Kelly et al., 2010, p. 260). Modern language may actually range from pure speech, as on radio or telephone, to pure manual gesture, as in signed languages. Moreover, if prevented from speaking, people naturally invent gestural communication, which can take on grammatical properties (Goldin-Meadow et al., 1996).

These considerations support the view that language itself evolved from manual communication, and may even have evolved to a level comparable to that of modern signed languages (Corballis, 2002). Language, then, may well have emerged before speech itself became the dominant mode.

## 2.5 Finding voice

If language can be traced to the intentional movements of the body, with emphasis on the hands and face, we need still to explain how the voice was incorporated. This was probably not a major step, since speech involves movements of the mouth that are not themselves vocal, and indeed overlap with movements involved in eating (MacNeilage, 2008). One view of speech is that it arose from the gradual shift of expressive facial movements into the mouth itself, where they are at least partly invisible. The addition of voicing was a device to render these movements accessible to the receiver—not through sight, but through sound.

Except for humans, primates are poor vocal learners—that is, as noted earlier, movements of the larynx are for the most part involuntary and not susceptible to learning (Petkov & Jarvis, 2012). Nevertheless it may have required a relatively small step in hominin evolution to bring the larynx under the control of the intentional motor system, and enable the learning of vocal sounds. One possibility is that the capacity for vocal learning was achieved through an extension of pre-existing motor pathways involved in movement—a possibility that arises from considerations of why some birds are vocal learners and some are not (Feenders et al., 2008). As in vocally articulate birds, so in humans. The analogue of Broca’s area in the macaque modulates movements of the mouth and face, but not of the larynx itself (Petrides et al., 2005), and it may have been a straightforward step to incorporate voicing into the system.

This might have been accomplished according to a parsimonious evolutionary principle whereby new cortical representations occur through the enlargement of older areas, with part of the enlarged area allocated to the new function and the remainder retaining the original function (Finlay et al., 2005). In endorsing such a mechanism for how some birds and humans evolved the capacity for vocal learning, Feenders et al. (2008, p. 21) write:

Our results are ... concordant with the gestural origin of spoken language hypothesis whereby the motor learning ability of gestures in humans and non-human primates has been argued to be the precursor behavior of the motor learning of speech/language.

Language, then, may not have depended on a “great leap forward” that transformed the manner in which concepts are represented and manipulated in the brain, but may rather have arisen gradually through the emergence of gestural communication systems tapping into representations of the world, and that themselves go far back in evolution. Late in primate evolution, and probably after the hominins separated from their common ancestry with great apes, this gestural system incorporated vocalization into a motor control matrix that already included manual and facial movements. Nonhuman primates are indeed manipulative creatures, using the hands for grooming and extracting food, and both hand and mouth for grasping, eating, and fighting. These activities provided a natural template for the evolution of intentional communication systems, going beyond the fixed systems of vocal calls.

But there is still one ingredient that is missing from this account.

### 3. Toward grammar

Nonhuman primates have the capacity to generate intentional gestures, and great apes have also demonstrated some capacity to create sequential signs with some semblance of grammar. But this capacity still seems to fall far short of the human ability to create grammatically complex, novel utterances, whether gestural or vocal. Of course grammatical language may well have preceded speech itself, and its origins may well lie in bodily gesture.

#### 3.1 Displacement and mental time travel

The grammatical, generative aspect of language may have been reliant on displacement, and the advantages to be gained from reference to events that are not-present (Bickerton, 2010; Corballis, 2011; Gärdenfors & Osvath, 2010); this in turn may have depended on the evolution of “mental time travel,” a term first used by Tulving (1985) and elaborated by Suddendorf and Corballis (1997, 2007). We humans, at least, carry the ability to consciously relive past events and imagine future ones, and indeed to construct entirely imaginary scenes with no reference to specific points in time (“Once upon a time”). The construction or reconstruction of events in the mind requires the arrangement of internal representations in a sequential fashion.

It has been argued that mental time travel itself is uniquely human (e.g., Köhler, 1925; Premack, 2007; Suddendorf & Corballis, 1997, 2007; Tulving, 1985), which itself might be taken as evidence that language itself emerged only in our species, or in the now extinct hominins that preceded us. Recent evidence suggests, though, that precursors of mental time travel may go far back in mammalian evolution. Mental time travel in humans depends critically on the hippocampus, a brain structure that is activated both when people “relive” past events or imagine future ones (Addis et al., 2011; Martin et al., 2011), and destruction of the hippocampus results in an inability to recover memories of personal events or the conjuring up of possible future ones (Corkin, 2013; Tulving, 2002; Wearing, 2005). But the hippocampus itself is not unique to humans, and plays similar roles in other species.

Individual neurons in the rat hippocampus fire when the animal is in particular locations in an environment, such as a maze, suggesting that the hippocampus is involved in the construction and activation of cognitive maps of the environment

(O'Keefe & Nadel, 1978). This activity occurs while the animal is exploring a maze, but also occurs in sharp-wave ripples (SWRs) sometime after the animal has actually been in the maze, either during slow-wave sleep (Wilson & McNaughton, 1984) or when the animal is awake but immobile (Karlsson & Frank, 2009). The paths indicated by the replay need not correspond to the actual paths taken in the maze, sometimes corresponding to the reverse of those actually taken (Foster & Wilson, 2006). Sometimes they correspond to paths the animal will take in the future (Pfeiffer & Foster, 2013), and sometimes to paths the animal never takes at all (Gupta *et al.*, 2010). The ripples are typically compressed in time, just as our own mental travels are generally much shorter than actual ones. Summarizing these and other findings, Dragoi and Tonegawa (2013, p. 6) write:

The existence of temporally compressed neuronal sequences that are independent of the recent experience of the animal could support the existence of the ability to travel mentally into its past as well as into the future, a sophisticated process that may underlie higher cognitive functions like memory recollection, navigational planning, imagining, cognitive map formation, and schema-based rapid learning.

And that is in the rat. This suggests, contrary to my earlier view (Corballis, 2011; Suddendorf & Corballis, 2007), that the basic ability to travel mentally in time and space goes far back in evolution (Corballis, 2013).

Of course remembered or imagined trajectories in a spatial environment do not have the complexity of human episodes, which include objects, people, animals, emotions, happenings—as well as locations in space and time—and their combinations. Nevertheless it seems likely that mental time travel evolved gradually over the tens of millions of years that separate us from our common ancestry with the other mammals, and perhaps goes further back to common ancestry with birds, which also possess a hippocampal analogue (Macphail, 2002). In creatures that move freely over the surface of the earth—and beyond—a mechanism for spatial memory and mental exploration in time and space may have been a very early requisites for successful adaptation.

### 3.2 The emergence of narrative

Perhaps it was the capacity for narrative, in which imagined events are woven into coherent sequences, that separated human mental time travel from that of other

extant species, and led to the capacity to communicate our travels to others. The literary scholar John Niles (2010) suggests that our species should be renamed *Homo narrans*—the story tellers. Cosentino and Ferretti (2014) suggest that language itself, in the form of connected discourse, depended on an advanced capacity for navigation in space and time. The elaboration of mental time travels into narratives probably dates from the Pleistocene, dating from some 1.6 million years ago, when our hunger-gatherer forebears foraged over increasingly wide terrain for food. It would have been increasingly adaptive to relay the foraging experiences to others. Children, too, would have benefited from the tales told by the adult hunters and foragers, gaining knowledge about food sources and hunting techniques before themselves graduating to the hunt. Sugiyama (2011) suggests that the importance of stories may explain the prolonged juvenile phase in humans. The juvenile phase began to increase with the earliest members of our genus, *Homo habilis*, early in the Pleistocene, reaching a peak with *Homo sapiens* (Locke & Bogin, 2006).

The Pleistocene also saw the emergence of our genus *Homo* as intensely social creatures, occupying what has been terms the “cognitive niche” as an adaptation to a scattered and dangerous environment (Tooby & De Vore, 1987). This created safety in numbers, and also more effective methods of foraging and problem solving. The evolution of language as a way to exchange complex information was perhaps the most critical accomplishment in the establishment of the cognitive niche (Pinker, 2003). It was not simply a matter of exchanging practical information. Stories are the product of imaginary mental travels, not only into past or future and into the minds of others, but also into realms of fantasy and play. Stories continue to dominate our lives, whether in the form of plays, novels, operas, television soaps, or bedtime tales told to children.

Story telling may have originated in pantomime, as our forebears acted out their experiences, and perhaps even their fantasies. Some degree of pantomime is evident in the natural communications of apes, including orangutans (e.g., Russon & Andrews, 2011), although it is unlikely that gestural communication evolved a story-like structure until the emergence of bipedal hominins, and perhaps not even until the emergence of the genus *Homo* in the Pleistocene (Thompson, 2010). A critical development may have been the development of stone tool-making industries, perhaps at the transition from the earlier Oldowan industry to the later, more complex Acheulian industry dating from around 1.7 million years ago. Stout et al. (2008) asked three archeologists experienced in early tool technologies to make Oldowan tools or Acheulian tools, or simply strike cobbles together, and in

each case measured the their brain activity using a PET scanner with a slow-decaying radiological trace. Only in the case of Acheulian tool-making did the brain activity overlap with language circuits, including Broca's area. The active areas also corresponded to the known mirror system. Unlike Oldowan tools, Acheulian tool-making involves planned sequential action. Whether pantomime evolved from the telling of stories to sequential action, or sequential action came first, is a moot point. Perhaps it was a matter of co-evolution, as the brain itself evolved to accommodate to increased cognitive and social demands.

### 3.3 From pantomime to language

Pantomime, though, is inefficient, and can be streamlined by a process of conventionalization (Burling, 1999), whereby pictorial or iconic representations are gradually replaced by arbitrary signals, shaped to convey information with maximum efficiency and minimum effort. In losing their pictorial quality, though, these arbitrary signals must be sustained and taught by the linguistic community, and transmitted between generations. The process of conventionalization is evident in the development of sign languages. For example, in American Sign language the sign for *home* was once a combination of the sign for *eat*, which is a bunched hand touching the mouth, and the sign for *sleep*, which is a flat hand on the cheek. Now it consists of two quick touches on the cheek, both with a bunched handshape, so the original iconic components are effectively lost (Frishberg, 1975). Albeit with a lack of modern cultural sensitivity, Darwin (1872) remarked

on the practice of the deaf and dumb and of savages to contract their signs as much as possible for the sake of rapidity. Hence their natural source or origin often becomes doubtful or is completely lost; as is likewise the case with articulate language (Darwin, 1872, p. 62).

In this view, speech itself can be regarded as the product of conventionalization, in which the element of pantomime is effectively lost, and spoken words are sustained through cultural transmission, albeit subject to mutations than have resulted in the some 7,000 different languages in today's world. Language, then, is not due to the sudden emergence of symbolic representation, as maintained by Chomsky and others, but is rather the end process of a gradual process by which the elements of discourse have lost their iconic connection with external reality. Grammar, too, might be a result of this process. Pantomimic sequences might well have retained

the sequential structure of the events they represent, but new conventions would then have been introduced to render the sequence of symbols coherent. Conventions are needed to differentiate different elements—as Pinker (2003, p. 27) put it, “who did what to whom, when, where and why”. This gives rise to such grammatical contrivances as case, tense, mood, and the like. The emergence of grammar, then, can also be seen as an incremental process, building up as the means of communication became less pantomimic and more arbitrary in format.

This process in turn might well have allowed language to move from the telling of stories involving concrete objects and physical actions to more abstract accounts, of which this very article is an example. But our ability to deal in abstract concepts depends heavily on metaphor, treating the abstract as we treat real-world objects and events (Lakoff & Johnson, 1980). We *grasp ideas*, talk of being *at a crossroad in a relationship*, or *stumble toward a solution*. Gallese and Lakoff (2005) suggest how metaphorical language might have arisen from elaboration of the mirror system itself.

Grammatical language, then, need not have arisen from a great leap forward that bestowed a procedure such as Chomsky’s unbounded Merge, but may have emerged from increasingly complex representations of real-world events in the form of internal narratives, and the establishment of gestural systems to communicate them. As these systems became less pantomimic and more conventionalized, rules were needed to establish convey properties of the narratives—who did what to whom, when, where and why. Such an account is similar to Chomsky’s in that it is the structure of our thoughts that fundamentally controls how we communicate them, but that structure evolved gradually through primate evolution, probably accelerating during the Pleistocene, rather than as a single leap within the timespan of our own species.

## Conclusions

Many of those who have speculated about the evolution of language have been misled by the assumption that language is identified with speech. The grammatical sophistication of signed languages should remind us that this is not the case. Speech bears misleading witness to language evolution because most animals do not have the intentional vocal control or capacity for vocal learning necessary for articulate speech. Animals capable of vocal learning include some birds, cetaceans, pinnipeds, and at least one elephant (Petkov & Jarvis, 2012), but conspicuous

among those with little if any capacity for vocal learning are the primates. Yet great apes and even dogs appear able to learn to comprehend human vocalizations, even if unable to produce them. The limitation imposed on nonhuman animals is therefore in large part one of vocal production, and not of understanding, nor even one of conceptual representation.

The more likely platform for language evolution is the sensorimotor cortex, which mediates the learning of motor sequences. In primate evolution, this was dominated by bodily movement, and especially manual and facial action, suggesting that language itself evolved through modifications of sensorimotor cortex. The mirror system, in particular, is the most likely candidate, since it maps the perception of bodily movement onto its production—a natural basis for communication. In our hominin forebears, the incorporation of vocalization into this system seems to have been late, and possibly restricted to *Homo sapiens*, although one might surmise that it was gradual rather than sudden.

Perhaps the most important feature of language, as distinct from other forms of animal communication, is that it permits communication about the non-present, such as past events, imagined future events, or hypothetical and even impossible events. The ability to mentally represent such events includes mental time travel, whereby we form internal scenarios. Although some have suggested that mental time travel is itself uniquely human, evidence from hippocampal recordings in the rat suggest that the ability to replay, pre-play, or even construct purely imaginary trajectories in spatial environments goes far back in evolution. The complexity of these imagined events no doubt increased through evolutionary time, perhaps gaining a narrative-like character in our hominin forebears after they split from the great apes.

The adaptive advantages of communicating internal narratives perhaps surfaced during the Pleistocene, when our forebears were forced away from a forested environment to a foraging existence on the open savanna. This created an imperative for increased social bonding and the exchange of information about food sources and other components of foraging expeditions. At first, this exchange might have depended on pantomime, a gestural replay of events that occurred or of future plans. Through time, pantomime would have been gradually replaced by more efficient gestures, and sustained by convention rather than by pictorial or iconic resemblance to what is communicated. Speech was the end product of this process of conventionalization, creating a medium that was energy-efficient, and that freed the hands and rest of the body for other activities, such as tool manufacture. Removal of the iconic component also allowed different languages to

mutate, increasing bonding within groups but denying access of information to other groups. There are now some 7,000 different languages.

This article is not of course the final word on language evolution, and many details need to be worked out. My main purpose, though, is to suggest that an evolutionary account is at least plausible, and we do not need to postulate any kind of miraculous “great leap forward.”

## REFERENCES

- Addis, D.R., Cheng, T., Roberts, R.P. & Schacter, D.L. (2011). Hippocampal contributions to the episodic simulation of specific and general future events. *Hippocampus*, *21*, 1045–1052.
- Allman, J.M. (2000). *Evolving brains*. New York: Scientific American Library.
- Bickerton, D. (2010). On two incompatible theories of language evolution In R.K. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language*. Cambridge: Cambridge University Press, 199-210.
- Burling, R. (1999). Motivation, conventionalization, and arbitrariness in the origin of language. In B.J. King (Ed.), *The origins of language: What nonhuman primates can tell us*. Santa Fe, NM, School of American Research Press.
- Calvert, G.A., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience*, *15*, 57-70.
- Cheney, D.L., & Seyfarth, R.M. (1990). *How monkeys see the world*. Chicago: University of Chicago Press.
- Cheney, D.L., & Seyfarth, R.M. (2005). Constraints and preadaptations in the earliest stages of language evolution. *The Linguistic Review*, *22*, 135-159.
- Chomsky, N. (2007). Bilingual explorations: Design, development, evolution. *International Journal of Philosophical Studies*, *15*, 1-21.
- Corballis, M.C. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.
- Corballis, M.C. (2011). *The recursive mind: The origins of language, thought, and civilization*. Princeton, NJ: Princeton University Press.
- Corballis, M.C. (2013). Mental time travel: The case for evolutionary continuity. *Trends in Cognitive Sciences*, *17*, 5-6.
- Corkin, S. (2013). *Permanent present tense: The man with no memory, and what he taught the world*. London: Allen Lane.

- Cosentino, E., & Ferretti, F. (2014). Communication as navigation: A new role for consciousness in language. *Topoi – An International Review of Philosophy*, *33*, 263-274.
- Darwin, C. (1859). *The origin of species*. London: John Murray.
- Darwin, C. (1872). *The descent of man, and selection in relation to sex*. London: John Murray.
- Desmurget, M., Richard, N., Harquel, S., Baraduc, P., Szathmari, A., Mottolesi, C., & Sirigu, A. (2014). Neural representations of ethologically relevant hand/mouth synergies in the human precentral gyrus. *Proceedings of the National Academy of Sciences*, *111*, 5718-5722.
- Dragoi, G., & Tonegawa, S. (2014). Selection of preconfigured cell assemblies for representation of novel spatial experiences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20120522. <http://dx.doi.org/10.1098/rstb.2012.0522>
- Egnor, S.E.R. & Hauser, M.D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, *27*, 649-654.
- Emmorey, K. (2002). *Language, cognition, and brain: Insights from sign language research*. Hillsdale, NJ: Erlbaum
- Feenders, G., Liedvogel, M., Rivas, M., Zapka, M. Horita, H., Hara, E., Wada, K., Mouritsen, H., & Jarvis, E. D. (2008). Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS ONE*, *3*, e1768. doi:10.1371/journal.pone.0001768
- Finlay, B.L., Cheung, D., & Darlington, R.B. (2005). Developmental constraints on or developmental or developmental structure in brain evolution. In Y., Munakata & M., Johnson (Eds.), *Attention and performance XXI: Process of change in brain and cognitive development*. Oxford: Oxford University Press, 131-162.
- Foster, D.J., & Wilson, M.A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, *440*, 680-683.
- Frishberg, N. (1975). Arbitrariness and iconicity in American Sign Language. *Language*, *51*, 696-719.
- Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensorimotor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*, 455-479.
- Gärdenfors, P., & Osvath, M. (2010). Prospection as a cognitive precursor to symbolic communication. In R.K., Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language*. Cambridge: Cambridge University Press, 103-114.
- Gentilucci, M., Campione, G.C., De Stefani, E. & Innocenti, A. (2012). Is the coupled control of hand and mouth postures precursor of reciprocal relations between gestures and words? *Behavioural Brain Research*, *233*, 130-140.

- Gentilucci, M., & Corballis, M.C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews*, *30*, 949-960.
- Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psychological Review*, *103*, 34-55.
- Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behaviour*. Cambridge, MA: Harvard University Press.
- Graziano, M.S., & Afshar, T.N. (2007). Mapping behavioral repertoire onto the cortex. *Neuron*, *56*, 239-251.
- Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M. et al. (2010). A draft sequence of the Neanderthal genome. *Science*, *328*, 710-722.
- Gupta, A.S., van der Meer, M.A.A., Touretzky, D.S., & Redich, A.D. (2010) Hippocampal replay is not a simple function of experience. *Neuron*, *65*, 695-705.
- Hepper, P.G., Wells, D.L., & Lynch, C. (2005). Prenatal thumb sucking is related to postnatal handedness. *Neuropsychologia*, *43*, 313-315.
- Hopkins, W.D., Tagliapietra, J.P. & Leavens, D.A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour*, *73*, 281-286.
- Johansson, S. (2013). The talking Neanderthals: What do fossils, genetics, and archaeology say? *Biolinguistics*, *7*, 35-74.
- Kaminski, J., Call, J. & Fischer, J. (2004). Word learning in a domestic dog: Evidence for “fast mapping.” *Science*, *304*, 1682-1683.
- Karlsson, M.P. & Frank, L.M. (2009). Awake replay of remote experiences in the hippocampus. *Nature Neuroscience*, *12*, 913-918.
- Kelly, S.D., Özyürek, A. & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*, 260-267.
- Klein, R.G. (2008). Out of Africa and the evolution of human behavior. *Evolutionary Anthropology*, *17*, 267-281
- Köhler, W. (1925). *The mentality of apes*. New York: Routledge & Kegan Paul. (Originally published in German in 1917).
- Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*, 846-848.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

- Locke, J.L., & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral & Brain Sciences*, *29*, 259-325.
- MacNeilage, P.F. (2008). *The origin of speech*. Oxford: Oxford University Press.
- Macphail, E.M. (2002). The role of the avian hippocampus in spatial memory. *Psicologica*, *23*, 93-108.
- Marshall, A.J., Wrangham, R.W. & Arcadi, A.C. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour*, *58*, 825–830.
- Martin, V.C., Schacter, D.L., Corballis, M.C., & Addis, D.R. (2011). A role for the hippocampus in encoding simulations of future events. *Proceedings of the National Academy of Sciences*, *108*, 13858-13863
- McBrearty, S. (2007). Down with the revolution. In Mellars, P., Boyle, K., Bar-Yosef, O., & Stringer, C. (Eds.), *Rethinking the human revolution*. McDonald Institute Monographs. Cambridge: McDonald Institute for Archaeological Research, 133–152.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*, 350-371.
- Niles, J.D. (2010). *Homo narrans: The poetics and anthropology of oral literature*. Philadelphia: University of Pennsylvania Press.
- O’Keefe, J. & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Patterson, F.G.P., & Gordon, W. (2001). Twenty-seven years of project Koko and Michael. In B.M.F. Galdikas, N.E. Briggs, L. K. Sheeran, G. L. Shapiro, & J. Goodall (Eds.) *All Apes Great and Small, Vol. 1: African Apes*. New York: Kluver, 165-176.
- Penfield, W., & Rasmussen T. (1950). *The cerebral cortex of man*. New York: Macmillan.
- Petkov, C. I., & Jarvis, E.D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, *4*, article 12.
- Petrides, M., Cadoret, G., & Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca’s area. *Nature*, *435*, 1235-1238.
- Petrides, M., & Panya, D.N. (2009). Distinct parietal and temporal pathways to the homologues of Broca’s area in the monkey. *PLoS Biology*, *7*, e1000170, <http://www.doi:10.1371/journal.pbio.1000170>.

- Pfeiffer, B.E., & Foster, D.J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*, 74-80.
- Pilley, J. W., & Reid, A.K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*, 184-195.
- Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M.H. Christiansen & S. Kirby (Eds), *Language evolution*. Oxford: Oxford University Press, 16-37.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral & Brain Sciences*, *13*, 707-784.
- Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences (USA)*, *104*, 13861-13867.
- Raffaele, P. (2006). Speaking bonobo. *Smithsonian Magazine*, November 2006. Online at: <http://www.smithsonianmag.com/science-nature/speakingbonobo.html> (accessed 25 August 2014).
- Rizzolatti, G., Camardi, R., Fogassi, L., Gentilucci, M., Luppino, G. & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental Brain Research*, *71*, 491-507.
- Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*, 264-274.
- Russon, A., & Andrews, K. (2011). Orangutan pantomime: Elaborating the message. *Biology Letters*, *7*, 727-630.
- Savage-Rumbaugh, S., Shanker, S.G. & Taylor, T.J. (1998). *Apes, language, and the human mind*. New York: Oxford University Press.
- Shea, J.J. (2011). *Homo sapiens* is as *Homo sapiens* was. *Current Anthropology*, *52*, 1-35.
- Slocombe, K.E., Kaller, T., Call, J., & Zuberbühler, K. (2010). Chimpanzees extract social information from agonistic screams. *PLoS ONE*, *5*, e11473.
- Slocombe, K.E., & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology*, *15*, 1779-1784.
- Stout, D., Toth, N., Schick, K., & Chaminade, T. (2008). Neural correlates of early Stone Age toolmaking: Technology, language and cognition in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 1939-1949.
- Suddendorf, T., & Corballis, M.C. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs*, *123*, 133-167.
- Suddendorf, T., & Corballis, M.C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*, 299-351.

- Sugiyama, M.S. (2011). The forager oral tradition and the evolution of prolonged juvenility. *Frontiers in Psychology*, 2, Article 133.
- Sutton-Spence, R., & Boyes-Braem P. (Eds.) (2001). *The hands are the head of the mouth: The mouth as articulator in sign language*. Hamburg: Signum-Verlag.
- Tattersall, I. (2012). *Masters of the Planet: The Search for Human Origins*. New York: Palgrave Macmillan.
- Thompson, T. (2010). The ape that captured time: Folklore, narrative, and the human-animal divide. *Western Folklore*, 69, 395-420.
- Tooby, J. & DeVore, I. (1987). The reconstruction of hominid evolution through strategic modeling. In W.G. Kinzey (Ed.), *The evolution of human behavior: Primate models*. Albany, NY: SUNY Press, 183-238.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1-12.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1-25.
- Watkins, K.E., Strafella, A.P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41, 989-994.
- Wearing, D. (2005). *Forever today*. New York: Doubleday.
- Wilson, M.A., & McNaughton, B.L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265, 676-679.

# Embodied Pragmatics and the Evolution of Language

*Erica Cosentino*<sup>†</sup>  
ericacosentino@libero.it

## ABSTRACT

In the evolutionary theory, a central tenet is that complex forms evolved from simpler ones, according to a bottom-up process. When it comes to the evolution of language, however, a bottom-up approach is problematic. In this case, such an approach often assumes that minimal units that are inflexibly associated to their meaning come first, where the wider discourse is only a later product. In the present paper, I shall argue that we need to assume a top-down perspective on language evolution, which claims that the wider discourse is the evolutionary starting point rather than the final achievement. This approach involves the necessity to focus on the pragmatic abilities of our ancestors and on the biological mechanisms underlying them. Combining a top-down model of language evolution with an embodied account of cognition, I shall argue that a basic mechanism of *affordance* perception supports core pragmatic processes by enabling the individual to determine not only her own action possibilities in the physical environment but also the action possibilities of others and, thereby, enabling her to determine other people's intentions. As a result, I shall introduce the notion of an embodied pragmatics as a key to account for the evolution of language.

Keywords: embodied theory, affordances, language evolution, pragmatics, discourse.

## 1. Introduction

The present paper is based on a double methodological claim. First, I assume that one's own theory of language has to be constrained by evolutionary considerations. It follows from this assumption that a model of language which is not consistent with the theory of evolution is not a good model and has to be

<sup>†</sup> University of Calabria, Rende, Italy.

abandoned (“your theory of language depends on your theory of language evolution”; cf. Ferretti, 2010). Second, I maintain that also the opposite statement is true, namely one’s own theory of language evolution has to be constrained by an empirically informed theory of language, given that the type of questions for a theory of language evolution to answer depends on what we think language is (“your theory of language evolution depends on your theory of language”; cf. Jackendoff, 2010). The structure of the paper follows from these methodological considerations.

In the first paragraph, I shall introduce two alternative models of language which, on the basis of different predictions concerning the role of discourse-level contextual information in semantic processing, characterize the interpretative process as a two-step or a single-step procedure. I shall discuss the results of an experimental investigation that supports a single-step model, showing that the context of discourse has pervasive top-down effects on the processing of individual words and sentences. In the second paragraph, I shall address the question of how the empirical investigation of modern language can shed light on the issue of language evolution. I shall argue that a two-step model of language processing is incompatible with evolution and that evolutionary considerations lead to assume a top-down perspective according to which pragmatic processes enabling inferential communication were necessary to the evolution of language. In the third and last paragraph, I shall show that a top-down model of language evolution is consistent with an embodied – bottom-up – account of cognition. Focusing on the mechanism of *affordance perception*, I shall finally argue in favor of the notion of an *embodied pragmatics* as a key to account for language evolution.

## 2. Two models of language

Since the Chomskian revolution in linguistics (Chomsky, 1957), the study of human language in cognitive sciences has been influenced by the idea that the sentence is not only the core unit of syntactic analysis but also the core unit of language interpretation. This notion is often been conjoined with the minimalist semantic claim that sentences encode meaning by means of a context-free rule-based combination of lexical–semantic features of the words within a sentence (Grice, 1975; Borg, 2004; Cappelen & Lepore, 2005).

According to the minimalist or literalist, the process of semantic composition, which consists in putting together the semantic values of the parts to establish the semantic value of the whole, determines what is explicitly said by the speaker or “sentence’s meaning”. This step of the comprehension process is considered necessary, and it corresponds to the level of sentences’ truth conditions. The role of the context at this level is limited to cases of indexicality and anaphors resolution and should be traceable to syntactic elements in the logical form of sentences. That is, pragmatic processes of saturation and disambiguation suffice to determine the literal interpretation of the sentence. All other pragmatic processes involved in the interpretation of the utterance, which contribute to determine “speaker’s meaning”, are secondary, and presuppose the identification of the meaning literally expressed.

From this point of view, then, linguistic interpretation is construed as a two-step procedure. First, the literal interpretation of the sentence is computed by combining fixed word meanings in ways specified by the syntax, and second, information from prior discourse, world knowledge and other sources of extra-linguistic information are used to integrate sentence meaning. According to this perspective, language processing proceeds in a bottom-up fashion, incorporating contextual information only after establishing phrase or sentence local meaning.

The two-step model is challenged by the idea of free pragmatic enrichment according to which contextual information can be immediately incorporated into the truth-evaluable sentence meaning such that global context and lexical content contribute to sentence meaning at once, leading to a single-step model. Contextual information may be then used in a top-down fashion, such that the local contribution of individual words or sentences is a function of the construction of a situational interpretation at the global meaning level (Cosentino et al., 2013).

In order to adjudicate between these two models, in a recent experimental study (Cosentino et al., 2014) we tested their predictions regarding the semantic integration of a sentence in a discourse. A crucial aspect of a two-step perspective on interpretation is that local semantic constraints have precedence over global contextual factors and, consequently, local semantics cannot initially be overruled by the wider context. Single-step models on the contrary, assume that there is no such priority of local constraints over contextual factors such that, in principle, the wider context of discourse has an

immediate effect on the interpretation of the unfolding linguistic information. More to the point, we used the electroencephalography (EEG) to record event-related brain potentials (ERPs) while people read short stories. We focused on the N400 component, a negative ERP deflection peaking around 400 ms after stimulus onset and larger over centro-parietal electrodes. The N400 has become particularly relevant in language studies given its close relation to the processing of word meanings in context (see for a review Kutas & Federmeier, 2011).

In our experiment, subjects were presented with stories in which a human character selected an object to accomplish a specific goal. The combination between object and action could be either appropriate or not with respect to (1) the local telic component of the meaning of nouns (i.e., the function or purpose of an object coded in the lexical entry), and (2) the context-driven affordances of the object (i.e., its action possibilities). Two-step models predict that local semantics cannot initially be overruled by the wider context so that a telic violation such as “She uses the funnel to *hang* her coat” will always be considered inappropriate, regardless of the wider discourse. In our study we found that when this anomalous combination is preceded by a conventional context, it elicits indeed a typical N400 effect, indicating that the subject is experiencing interpretative problems. However, if the very same combination is preceded by a modified/unconventional context, the latter has the power to neutralize the violation, as reflected by the reduction of the N400. Additionally, an unconventional context can also make a locally acceptable combination (e.g. using a funnel to *pour* water into a container) globally incongruent with reference to the context-driven affordances of the object. These findings challenge the two-step model, showing that contextual information has top-down effects on linguistic processing as predicted by the single-step model. The latter, then, can better account for our results.

This conclusion is consistent with recent work in the field of lexical pragmatics (Wilson & Carston, 2007), which emphasizes that the distinction between semantics and pragmatics can be applied also at the level of individual words or phrases rather than whole sentences. In such a view, the meanings of words are often pragmatically adjusted and fine-tuned in context, in accordance with speakers’ needs and gleaning opportunistically to what they know about the world, their interlocutors, and previous discourse. Understanding a word in context may involve the construction of an ‘ad hoc’ concept or occasion-specific sense, which is based on encoded concepts,

contextual information and pragmatic expectations (Barsalou, 1987, 1993; Wilson & Carston, 2007; see also Recanati, 2004; Glucksberg, 2001). The results of the construction of an occasion-specific sense is the modulation of the lexically encoded meaning via narrowing or broadening, that is, the linguistically-specified meaning can be made either more specific (e.g. *drink* used to mean ‘alcoholic drink’) or more general (e.g. *square* used to mean ‘squarish’).

Importantly, most current approaches to lexical pragmatics maintain that the occasion-specific senses created by the pragmatic interpretation of individual words and phrases are components of the proposition explicitly expressed by the speaker. According to this contextualist thesis, there is no level of semantic content that is independent of pragmatic processes (Recanati, 2004; Carston, 2002; Wilson & Carston, 2007; Wilson & Sperber, 2004). Sentences’ truth conditions are pragmatically constructed: the output of syntactic processes is subpropositional, and it provides only a template or schema for building propositions by means of pragmatic processes. These pragmatic processes are not restricted to saturation, which is mandatory and bottom-up (in order to understand what is said in “She is John’s sister”, it is necessary, at least, to determine to whom the speaker refers by the pronoun “she”), but include also top-down processes driven by the context such as narrowing, broadening and “free enrichment” (“She took out the key and opened the door” can be freely enriched, or specified, by “She opened the door with the key”). The latter processes are not triggered by any linguistic property of the utterance and are entirely pragmatically motivated. Pragmatics then is not only concerned with what is implicitly meant by the speaker, but it also heavily shapes what the speaker explicitly says. Whereas two-step models do not allow for the immediate integration of contextual information, single-step models are more flexible and allow for the priority of either local or contextual factors to be established case by case, depending on the context. In the next paragraph I will suggest that the idea that meaning is immediately contextualized has significant implications not only for the study of actual language processing but, even more importantly to our current aims, for the question of the evolution of language.

### 3. A top-down model of language evolution

Although human languages are very rich codes, they are not optimal. While in an optimal code a signal is strictly associated with only one meaning, in human languages the same linguistic expression can take on different meanings in different contexts. In the previous section, we have argued that meaning is immediately contextualized such that a single-step model of language processing, which allows for top-down effects of contextual information, can better account for language understanding. This claim has striking implications for the issue of language evolution given that empirically-informed models of modern language functioning should provide constraints to build plausible models of the evolution of language (see Jackendoff, 2010).

A preliminary trivial remark is that a full-fledged code cannot be assumed to be the evolutionary starting point. Therefore, if contextual factors play a major role in the interpretation of completely developed current linguistic codes, they must have been even more crucial in the evolutionary history of language, when those codes were not yet evolved. I am suggesting that the rising verbal communication could not work unless initial linguistic expressions were contextually constrained, just like current linguistic expressions are. It means that the issue of the evolution of language must be addressed assuming a top-down approach. In order to better appreciate the repercussions of this statement and help situating our proposal within other recent theories of language evolution, it might be useful to frame the notion of a top-down model of language evolution with reference to the debate on the passage from protolanguage to language.

The ability to process complex languages thanks to the existence of syntactic abilities is considered by many to be a uniquely human feature. It is hypothesized that the common ancestor of chimpanzees and humans was not in possess of this capacity. Therefore, evolutionary linguists face the problem of explaining the gap between a non-linguistic ancestor and our linguistic (i.e., syntactical) species. The notion of protolanguage has been commonly invoked as a stable intermediary stage in the evolution of language: “[t]he hypothesis of a protolanguage helps to bridge the otherwise threatening evolutionary gap between a wholly alingual state and the full possession of language as we know it” (Bickerton, 1995, p. 51).

Protolanguage has been defined under two competing accounts: the synthetic account and the holistic account. Under the *synthetic* account

(Bickerton, 1990, 1995; Tallerman, 2007), protolanguage had a limited set of word-like symbols which could be used to convey simple, atomic meanings, effectively the ancestors of modern nouns and verbs. Under the *holistic* account (Wray, 1998, 2000; Arbib, 2005), protolanguage units represented complex propositions, similar to modern sentences but lacking in internal morphological structure. Both accounts assume that protolanguage had distinguishable units, but their disagreement over the level of complexity of those units leads to different ideas of how protolanguage could have developed into modern language. Clearly, from this point of view, the origin and evolution of syntactic language is guided by evolutionary pressures to evolve more and more efficient systems of communication. A detailed review of the current positions concerning the nature of protolanguage is beyond the scope of this paper. However, it might be useful to bear in mind that my argument relies on a notion of protolanguage that seems to be more consistent with the holistic account. In fact, I argue that a linguistic unit conveys a far more complex meaning than assumed by supporters of the synthetic account and, as we will see, this perspective raises a relevant issue concerning which cognitive abilities are involved in inferring such a meaning.

What is even more relevant to the present paper is that this complex meaning is not fixed but rather pretty flexible and needs to be interpreted according to broader contextual constraints. From this perspective then, instead of focusing on the precursors of human syntactic abilities, I adopt a pragmatic perspective on the evolution of language and argue that the origin of modern syntactic language can only be explained by considering the protolanguage in terms of protodiscourse, namely in terms of the pragmatic processes that determine the contextually appropriate interpretation of linguistic expressions (see Ferretti, 2013 for a related perspective). This statement raises the question of how contextual constraints can lead to the appropriate interpretation of an utterance in the evolutionary scenario. What abilities were involved? Also, since we are interested in biological as opposed to cultural evolution, what were the core biological mechanisms that enabled these abilities? In order to answer these questions we need to specify, first of all, what a pragmatic model of language evolution is meant to explain and why it should be a better framework for analyzing the evolutionary issue.

According to the classical model of communication, or “code model” (Shannon and Weaver, 1949), communication is described by the twin acts of encoding and decoding. Speakers encode the meaning into a succession of

sounds and then transmit the signal. Receivers then decode the message in order to be able to share the thoughts of the speaker. From this point of view, the key question is to look for a specific mechanism that can transform the sounds uttered by the speaker into meaning. In opposition to this view, and exploring the implications of a pragmatic perspective for the evolution of language, Origgi and Sperber (2000) and Sperber and Origgi (2010) have argued that the function of language is not to encode the speaker's meaning but rather to offer some hints that can help to infer that meaning. Their claim is that the inferential model is necessary for a structured code to develop. They make the convincing claim that for coded communication to work, speaker and listener have to share exactly the same code. Any difference that affects an individual's code is likely to produce a mismatch between his signals and those of his conspecifics compromising the success of communication. In this situation then, a linguistic mutant who introduces modifications of the code, including the addition of new signals, would have counter-adaptive effects because these modifications would only be advantageous once they are shared by the population.

In the context of the inferential communication though, the situation would be very different. The inferential model in fact does not assume that the code needs to be shared. The code, actually, plays a little role here because the inferential model allows communication to precede grammar such that a fragmentary and ambiguous coding is generally sufficient, in the context, to unequivocally indicate the speaker's meaning. Then, not only the inferential model is necessary for the code to change but, what is more, if we imagine a protolanguage that consists only of words associating sound and meaning without any syntactic structure, such a poor and fragmentary code could be of use only to beings capable of inferential communication. For such individuals, the role of linguistic expressions in the communication process is just that of providing evidence of the intended message which is sufficient for reconstructing a full-fledged meaning, the speaker's meaning. This is a crucial notion within an inferential model of communication. Following Grice (1957), speaker's meaning is defined as a complex communicative intention aimed to achieve a certain effect upon the mind of the hearer by means of the hearer's recognition of the very intention to achieve this effect. In these terms, then, pragmatic interpretation is ultimately an exercise in mind-reading, the ability to infer the mental states of others. In the next section I will focus on this ability providing a new insight on its evolutionary foundation and its pragmatic

function. In particular, while Sperber and Wilson (2002) argue that the interpretation process involves a dedicated “metacommunicative” system, with its own specific principles and mechanisms, I will suggest that the interpretation process is strongly linked to other mechanisms rather than being encapsulated. More to the point, I will argue that motor activity takes part in the construction of meaning.

#### 4. Embodied evolutionary pragmatics

In the last decades there has been much debate about the nature of the ability to ascribe mental states to others and the putative mechanism underlying it. Most of the research in this area has focused primarily on high-level cognitive processing, such as understanding that others can have different desires and beliefs from one’s own (Baron-Cohen et al., 1985). However, from an evolutionary perspective, it is relevant to ask how these complex abilities have originated from more basic capacities, and what the impact of the latter was on human socio-communicative abilities and language evolution. While I adopt a top-down model of language evolution, I argue that the pragmatic processes that allow for the immediate contextualization of meaning are based on very basic cognitive abilities and biological mechanisms. It means that top-down effects are supported by processes and biological mechanisms that can be described assuming a bottom-up perspective. Specifically, I suggest that the notion of the pragmatic origins of language has to be combined with an embodied account of cognition which claims that “high-level” cognitive processes are rooted in “low-level” processes such as action and perception (Pecher & Zwaan, 2005). From this point of view, pragmatic processes are based on sensory-motor processes that enable individuals’ physical interactions with their environment.

In the last twenty-five years, a growing number of studies have emphasized that language and cognition are deeply rooted in the experience that results from possessing a body with certain physical features and a specific sensory-motor system. Most important to our current aims, humans’ ability to attribute mental states to others may also be interpreted in these terms. In particular, “motor theories of social cognition” aim to derive human social cognition from human motor cognition showing that the ability to understand other people’s minds crucially involves the capacity to understand other people’s intentions

by observing their actions. An action is a goal-directed sequence of bodily movements initiated and monitored by a ‘motor intention’. Thus, understanding a perceived action requires at least representing the agent’s motor intention and, possibly, also her social and communicative intentions (Jacoboni et al., 2005; but see Jacob & Jeannerod, 2005 for a critique). Given that human adults readily explain actions ascribing certain beliefs to the agent, it is possible that this lower-level of “theory of mind” is a prerequisite for high-level processing such as representing others’ desires and beliefs (Blackemore & Decety, 2001; Gallese, 2003; Wolpert et al., 2003).

The idea is that the mechanism for inferring other people’s intentions from observed actions might depend on the same mechanism that allows interpreting the consequences of one’s own actions as being produced by one’s own intentions: the intentions of an actor are estimated via a process of *simulation*, in which the actions of the actor are interpreted as if they were one’s own actions, and thus relating them to one’s own intentions (Goldman, 2006). The recently discovered “mirror neurons” may be the cortical mechanism supporting the simulation process, as they are involved in coupling observation and execution of goal-related motor actions responding both when a particular action is performed and when the same action, performed by another individual, is observed. By automatically matching the observed movements of an agent onto one’s own motor repertoire without executing them, the firing of mirror neurons in the observer’s brain simulates the agent’s observed movements and thereby contributes to understand the perceived action (Gallese & Goldman, 1998; Gallese et al., 2004). Because mirror neurons provide motor, not purely perceptual, representations of actions, they must be crucially involved in establishing whether represented actions are executable, that is, consistent with the rules of the motor system. This step – establishing whether the action is possible or not – is crucial for an individual to understand actions. Here I would like to focus on this specific component and on the mechanism of *affordance perception*. Given that this is a very basic mechanism, the attempt to link it to more sophisticated mental processes deserves great attention if we are interested in the foundations of social cognition and, ultimately, of communication.

Affordances are commonly defined as qualities of an object or an environment that, in combination with a particular bodily structure, determine possibilities for actions (Gibson, 1979; Glenberg et al., 2013). That is, for an individual to detect an affordance is to perceive an opportunity for action which

eventually depends on many factors such as the perceptual characteristics of the object or environment, the features of one's body, and the situation at hand. In opposition to more traditional accounts that see perception purely as an input system and action as a separate output system, the notion of affordance incorporates motor aspects in perception, thus emphasizing the close connection between action and perception.

It is interesting to note that the role of affordance perception in language evolution has been previously discussed in the context of theories of the gestural origins of language with reference to the hypothesis of the involvement of the mirror system in language evolution ("the mirror system hypothesis"; see Arbib et al., 2014 for a recent review). According to this hypothesis, a "mirror for actions" system, concerned with both generating an action which is appropriate to object's affordances and recognizing the action being performed by another individual, provided the evolutionary basis for the emergence of a "mirror for words and constructions" system. The hypothesis builds on the distinction between a dorsal path concerned with the *how* of converting affordances of an object into motor parameters for interacting with it and a ventral path involved in determining which actions are to be executed by taking into account not only *what* the object is but also context, task and more. The crucial argument is that word recognition based on how the word is articulated must be differentiated from interpreting the meaning of the word via a ventral pathway. In particular, the dorsal "mirror for actions" system played a particularly relevant role in language evolution supporting attempts to reproduce the articulatory form of words that were not in one's motor repertoire. According to this hypothesis then, a mirror and production system for praxic actions provided the evolutionary basis for the emergence of a mirror and production system for words and larger utterances as articulatory structures. In this way, the mirror system allowed for complex imitation to be transferred from manual skills to a new communicative domain. Although supporters of this hypothesis state that the mirror system involved in affordance perception has little to say about the notion of intentional communication as an exercise in mindreading, I would like to emphasize that the role of affordance perception in the passage from a manual to a vocal form of communication is consistent with the role of this mechanism within a motor account of social cognition.

Even if affordances have been generally studied with respect to the ability of an individual to determine his own action possibilities in the environment,

there is also a wealth of research supporting the notion that affordances perception is fundamental to the success of social interactions. Humans' ability to understand and predict the movements of others in social contexts (Marsh, Richardson, & Schmidt, 2009; Ramenzoni et al., 2008a, 2008b) and to engage in ongoing interactions (Davis et al., 2010; Richardson, Marsh, & Baron, 2007) depends to large extent on it. Perceiving affordances is a key step in the process of building representations of both one's own intended actions and the potential actions of the agent with whom one is interacting. These representations are used to make predictions and estimates of the social consequences of the represented actions and, when these actions come to execution, they are perceived as social signals which can confirm or not the agent's predictions and, possibly, modify the agent's beliefs and desires. What is relevant to our current aims is that in the conceptual framework of the simulation hypothesis, the same sensory-motor processes that allow individuals to determine their own action possibilities in their environment are also recruited to estimate the action possibilities of other people and, thereby, to understand their actions. Affordance perception, then, may play a crucial role in the evolution of socio-communicative abilities.

In support of this hypothesis, it has been recently suggested that an impaired mechanism of affordance perception may be the cause of the known impairments in socio-communicative skills of individuals with Autism Spectrum Disorders (ASDs; Linkenauger et al., 2012). Specifically, subjects with ASDs have impairments in both social and motor skills. Whether they are different deficits or may originate from a common etiology is a highly controversial issue. An interesting recent study has provided data which seems to support the intriguing hypothesis that an impaired ability to perceive affordances is the putative mechanism underlying both impairments. In this study, in order to test the ability to perceive one's own affordances, a well-developed experimental paradigm was used (Ishak, Adolph, & Lin, 2008; Linkenauger et al., 2012; Warren & Whang, 1987) that consisted in asking individuals with ASD and normotypical control subjects to estimate whether they could perform simple actions without overt feedback. Then, subjects were asked to actually perform these actions and their accuracy in perceiving affordances was assessed. In particular, subjects performed three tasks: (1) graspability, (2) reachability, (3) aperture. In the graspability task, they were asked to estimate the extent of their grasping ability with respect to foam board blocks of different sizes; after they finished all estimates, the experiment

assessed whether they could actually grasp the object: a successful grasp was defined as the ability to lift the block completely off the table. For the reachability task, participants observed the experimenter moving a plastic chip away from them or toward them and they were instructed to tell the experimenter when they thought the chip was at the limit of their reach; their actual reaching abilities were subsequently assessed. In the aperture task, subjects had to anticipate if their hand could fit through a hole whose size was gradually manipulated, and after the estimates, the smallest aperture through which they could indeed fit their hand was assessed.

It is relevant to note that these tasks are well-suited to explore affordance perception in individuals with ASDs as they do not involve other skills that might be as well compromised in those individuals. For example, even if autistic people do have general motor problems, they are as accurate as typically-developed children in terms of hand and arm movements and their visual estimation abilities (e.g., size and distance estimation) are even superior. Moreover, deficits in these tasks cannot be attributed to clinical comorbidities since motor planning is intact in other clinical populations such as hyperactivity disorder, and finally affordance perception is not generally associated with imagination or creativity but rather with perceptual-motor integration, so even if individuals with ASDs may have limitations in imagination this cannot be the source of their impaired ability to perceive affordances. Taking into account these remarks, it is particularly interesting that difficulties with affordance perception tasks have been shown to correlate with scores on the Social Communication Questionnaire (SCQ) (Rutter et al., 2007), a specific measure of autistic socio-communicative abilities based on parents' report of the lifetime presence of a child's autistic symptoms. This questionnaire is composed of three subscales, communicative ability, reciprocal social interaction, and restrictive/repetitive behaviors, and highly correlates with "gold standard" diagnostic measures (such as the ADOS). A high SCQ score indicates more autistic symptoms. Relevant to our current aims, higher SCQ scores predict poorer results in the affordance perception tasks for ASD participants with the correlation being driven by communication and reciprocal social interaction subscales.

As mentioned above, in the light of the notion of simulation, impairments in perceiving one's own affordances may also affect the capacity to perceive the action possibilities of other individuals and, consequently, to understand their actions, leading to striking limitations in the social and communicative

domains. Of course, the issue of the role of affordance perception in socio-communicative deficits of autistic individuals needs to be further investigated and we should be very careful in drawing conclusions from this data. However, what this data suggests is that the rather counterintuitive notion of the involvement of a mechanism of sensory-motor integration in communication seems at least to have some plausibility. Exploring this relation might have extremely interesting implications for theories of language evolution given that affordance perception is a very basic mechanism but still may be involved in the critical pragmatic function of inferring communicative intentions, then allowing for top-down contextual effects. These effects, in turn, crucially constrain the interpretation of linguistic expressions enabling the construction of appropriate interpretations and, ultimately, guarantying the success of communication.

## 5. Conclusions

The aim of this paper was to argue in favor of a top-down model of language evolution in which pragmatic abilities are grounded on sensory-motor processes of action and perception. We have shown that the mechanism of affordance perception that is concerned with the physical interactions between an individual and his environment is also crucially involved in understanding the actions of other individuals and, consequently, in the evolution of socio-communicative abilities that were at the foundation of the pragmatic origins of language. By combining a top-down approach to language evolution with a bottom-up perspective of the cognitive mechanisms underlying language functioning, we have suggested that the notion of an embodied pragmatics may have a key role in understanding the evolutionary origins of language.

## REFERENCES

- Arbib, M.A., Casser, B., & Barrès, V. (2014). Language is handy but is it embodied? *Neuropsychologia*, 55, 57-70.
- Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28(2),105-124.

- Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a theory of mind. *Cognition*, 21, 37–46.
- Barsalou, L.W. (1987). *The instability of graded structure: Implications for the nature of concepts*. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge: Cambridge University Press, 101–140.
- Blakemore, S.J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, 2, 561–567.
- Barsalou, L.W. (1993). *Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols*. In A.F. Collins, S.E. Gathercole, M.A. Conway, & P.E. Morris (Eds.), *Theories of memory*. Hillsdale, NJ: Erlbaum, 29–101.
- Bickerton, D. (1990). *Language and species*. Chicago, IL: University of Chicago Press.
- Bickerton, D. (1995). *Language and human behaviour*. London: University College London Press.
- Borg, E. (2012). *Pursuing Meaning*. Oxford: Oxford University Press.
- Cappelen, H., & Lepore, E. (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell.
- Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton & Co.
- Cosentino, E., Baggio, G., Kontinen, J., Garwels, T., & Werning, M. (2014). *Lexicon in action: N400 effect on affordances and telicity*. In P. Bello et al. (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 2079–84.
- Cosentino E., Adornetti I., Ferretti F. (2013). *Processing Narrative Coherence: Towards a top-down model of discourse*. Open Access Series in Informatics (OASICS). Special Issue “2013 Workshop on Computational Models of Narrative”, 32, 61–75.
- Davis, T.J., Riley, M.A., Shockley, K., & Cummins-Sebree, S. (2010). Perceiving affordances for joint actions. *Perception*, 39, 1624–1644.
- Ferretti, F. (2010). *Alle origini del linguaggio umano. Il punto di vista evoluzionistico*. Laterza, Roma-Bari.

- Ferretti F. (2013), Navigation, Discourse and the Origin of Language. In Brambilla P., Marini A. (Eds.), *Brain evolution, language, and psychopathology in schizophrenia*, New York: Routledge, 22-32.
- Gallese, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philosophical Transactions of the Royal Society B*, 358, 517–528.
- Gallese V., Keysers C. & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8, 396–403.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 12, 493-501.
- Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the Revolution to Embodiment: 25 Years of Cognitive Psychology. *Perspectives on Psychological Science*, 8, 573-585.
- Glucksberg, S. (2001). *Understanding Figurative Language: From Metaphors to Idioms*. Oxford: Oxford University Press.
- Goldman, A.I. (2006), *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*, Oxford: Oxford University Press.
- Grice, P. (1975). *Logic and conversation*. In Cole, P., & J.L. Morgan (Eds.), *Speech Acts*. New York: Academic Press, 41–58.
- Ishak, S., Adolph, K.E., & Lin, G.C. (2008). Perceiving affordances for fitting through apertures. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1501–1514.
- Jackendoff, R. (2010). Your theory of language evolution depends on your theory of language. In R. Larson, V. Déprez, & H Yamakido (Eds.), *The Evolution of Human Language: Biolinguistic Perspectives*. Cambridge: Cambridge University Press, 63–72.
- Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in Cognitive Sciences*, 9(1), 21-25.
- Jacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology* 3, e79, pp. 1–7.

- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647.
- Linkenauger, S., Lerner, M.D., Ramenzoni, V., & Proffitt, D. (2012). A perceptual-motor deficit predicts social and communicative impairments in individuals with autism spectrum disorders. *Autism Research*, *5*, 352 - 362.
- Marsh, K.L., Richardson, M.J., & Schmidt, R.C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*, 320-339.
- Origgi, G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the Human Mind: Language, Modularity and Social Cognition*. Cambridge: Cambridge University Press, 140-169.
- Pecher, D., & Zwaan, R. A. (Eds.) (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. New York: Cambridge University Press.
- Ramenzoni, V.C., Riley, M.A., Shockley, K., & Davis, T. (2008a). Carrying the height of the world on your ankles: Encumbering observers reduces estimates of how high an actor can jump. *Quarterly Journal of Experimental Psychology*, *61*, 1487-1495.
- Ramenzoni, V.C., Riley, M.A., Shockley, K., & Davis, T. (2008b). An information-based approach to action understanding. *Cognition*, *106*, 1059-1070.
- Recanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.
- Richardson, M.J., Marsh, K.L., & Baron, R.M. (2007). Judging and actualizing intrapersonal and interpersonal affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 845-859.
- Rutter, M., Bailey, A., & Lord, C. (2005). *SCQ: The social communication questionnaire manual*. Los Angeles: Western Psychological Services.
- Sperber, D., Origgi, G. (2010). A pragmatic account of the origin of language. In R. K. Larson, V. Déprez, & H. Yamakido, (Eds.) *The Evolution of Human Language: Biolinguistic perspectives*. Cambridge: Cambridge University Press, 124-132.
- Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage?. *Lingua*, *117*, 3, 579-604.

- Warren, W.H., & Whang, S. (1987). Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 371–383.
- Wilson, D., & Carston, R. (2007). *A unitary approach to lexical pragmatics: relevance, inference and ad hoc concepts*. In N. Burton-Roberts (ed.), *Advances in Pragmatics*. Basingstoke: Palgrave Macmillan, 230-260.
- Wilson, D., & Sperber, D. (2004). *Relevance theory*. In L. Horn, & G. Ward (eds.), *The Handbook of Pragmatics*. Oxford: Blackwell, 607-632.
- Wolpert, D.M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society London B*, 358, 593–602.
- Wray, A. (2000). *Holistic utterances in protolanguage: The link from primates to humans*. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *The evolutionary emergence of language*. Cambridge: Cambridge University Press, 285-302.
- Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 18, 47–67.

# The Evolution of Sentential Structure

*Peter Gärdenfors*<sup>†</sup>  
peter.gardenfors@lucs.lu.se

## ABSTRACT

The aim of this article is to present an evolutionarily grounded explanation of why we speak in sentences. This question is seldomly addressed, neither in the Chomskian tradition nor in cognitive linguistics. I base my explanation on an analysis of different levels of communication. I identify four levels: praxis, instruction, coordination of common ground and coordination of meaning. The analysis will be focused on the evolutionary benefits of communicating about *events* as a way of coordinating actions. A cognitively grounded model of events will be outlined. My central thesis is that the communicative role of sentences is to express events.

Keywords: sentence structure, pragmatics, semantics, evolution of language, common ground, coordination of meaning, events, event construals, cooperation, indirect reciprocity.

## 1. Why do we speak in sentences?

In evolutionarily early forms of communication, the communicative *act* in itself and the context it occurs in were presumably more important than the expressive form of the act (Clark, 1992; Winter, 1998; Gärdenfors, 2010). As a consequence, the pragmatics of natural language is the most basic from an evolutionary point of view. When communicative acts become more varied and eventually conventionalized during hominin evolution and their contents become detached from the immediate context (Gärdenfors, 2000), one can start attending to the expanding meanings of the acts. Then semantics becomes salient. Finally, when linguistic communication becomes even more conventionalized and combinatorially richer, certain markers, a.k.a. syntax, are

<sup>†</sup> Lund University, Sweden; Stellenbosch Institute of Advanced Study, Wallenberg Research Centre at Stellenbosch University, Marais Street, Stellenbosch 7600, South Africa.

used to disambiguate the contents when the context and the common ground of the interlocutors are not sufficient. According to this view, syntax is required only for the subtlest aspects of communication – pragmatic and semantic features are more fundamental.

This view of the evolutionary order of different linguistic functions stands in sharp contrast to much of mainstream contemporary linguistics. For followers of the Chomskian school, syntax is the primary study object of linguistics; semantic features are added when grammar is not enough; and pragmatics is a wastebasket for what is left over (context, deixis, etc.).

Clark (1996, p. 56) calls the Chomskian perspective the *production tradition* (focusing on the products of language) and the perspective that puts pragmatics in focus the *action tradition*. These two approaches to the evolution of language generate quite different research questions. It seems that never shall the twain meet.

There is, however, one linguistic unit that is central to both approaches: the *sentence*. In the Chomskian tradition it is taken for granted that the central goal of linguistic production is to generate sentences with a minimal structure of a noun phrase and a verb phrase. And the core linguistic data concern whether certain combinations of words are grammatical. Also in the pragmatic-semantic tradition, the sentence plays an important role. Furthermore, in analytic philosophy, sentences are central units, being the bearers of truth-values. In the tradition since Frege, a sentence expresses a proposition. But also in more cognitively oriented semantics, sentences are seen as natural units (e.g., Langacker, 1987, 2008; Talmy, 2001; Croft, 1991; Goldberg, 1995, Levin & Rapaport Hovav, 2005).

Since the sentence is so central to both research traditions, it is surprising that nobody asks *why* this unit exists. The question becomes more pressing when one compares with what is communicated by language-trained apes and other animals. Their communication never, or just by chance, exhibits sentential structures. Kanzi and his colleagues only bundle signs together without concern for whether the collection forms a sentential structure (Savage-Rumbaugh & Lewin 1994). For example, Greenfield and Savage-Rumbaugh (1990) succeed in finding a few rough patterns. Kanzi more often places the verb before the object – “hide nut” instead of “nut hide” in accordance with a language such as English. When he combines two verbs, for example “tickle bite” (which does not occur in English), he wants to do the actions in the order he mentions them. However, Kanzi’s grammatical patterns

are far from consistent, and they tally poorly with the grammatical competence that Chomsky's theory of language postulates.

My aim in this article is to present an evolutionarily grounded explanation of why we speak in sentences. I will follow what Clark (1996) calls the action tradition and base my explanation on an analysis of different levels of communication. The analysis will be focused on the evolutionary benefits of communicating about *events*. My central thesis is that the communicative role of sentences is to express events.

## 2. Levels of communication

The obvious goes without saying. If all partners in a cooperating group perform their tasks as expected by the others, there is no need for communication. Cooperation takes place on the level of praxis. It is only when an instruction, a correction or a coordination is needed that communication plays a pragmatic role. The basic level of communication is therefore for solving problems of coordinating actions. For example, if A is carrying a heavy box, but his path is blocked by a closed door, and B does not realize the situation, A typically *instructs* or *requests* B to open the door.

However, there are situations when the communicators misunderstand each other, because of badly formulated instructions or because they have different mental models of the world. For example, if A commands B to open the door and B sees several possible doors, he replies "Which door". Then A and B move to the level of coordination of common ground (Clark, 1992), that is, to agree on which door A wants B to open. When this is accomplished, they return to the level of instruction and B can perform the desired action. Coordination of common ground can also be done as a preparation for future collaboration. As I argue below, this aspect is central from an evolutionary perspective. Everyday talk about what other people do or have done also belongs to this level.

There is a third, more severe form of misunderstanding that occurs because the addressee does not understand an expression used by the speaker, or does not understand it in the same way. For example if you say "I'll talk to the chair" and you mean the chairperson, while *chair* for me just means physical objects, I will not understand your intention. On this final level – the level of semantic

coordination – the communicators must negotiate their use of expressions until they find a sufficient agreement.

For these reasons, following Winter (1998), I want to distinguish three levels of communication, in addition to a ground level of human interaction:

Level 0: *Praxis*. On this level people interact with each other without using intentional communication.

Level 1: *Instruction*. On this level coordination of action is achieved by instruction.<sup>1</sup>

Level 2: *Coordination of common ground*. On this level people inform each other in order to reach a richer or better coordination. It can also be achieved via questions.

Level 3: *Coordination of meanings*. On this highest level, people negotiate the meanings of words (labels) and other communicative elements.

The four levels are used in a hierarchical manner. When one level does not function properly, a break in the communication is signalled and it moves to the next higher level. When the problem is solved the communicators signal an acknowledgment and return to the level below. One example of this is the coordination of which door to open, that was presented above. In this case the communication goes from the level of instruction to coordination of common ground and then back again. Another example, going from the second to the third level, is that if *A* is telling *B* something and uses a word that *B* does not understand, *B* can signal this and they move to the level of coordination of meaning. When this is accomplished, they return to the level of coordination of common ground. Considering the evolution of communication, it is also reasonable that the four levels emerged in the same order as well.

Clark's (1992, 1996) work on *common ground* and *uptakes* can be seen as analyses of some central forms of coordination on level 2. First, the utterances in a conversation introduce new referents or new information about referents. This, together with the participants' expectations about the other's previous knowledge, forms the common ground that the subsequent conversation can take for granted. Second, a participant often introduces a proposal for a joint project in the conversation. This proposal can be taken up by the interlocutor

<sup>1</sup> This level corresponds to the language games introduced by Wittgenstein (1953).

(or it can be rejected). The proposal and the uptake then lead to a coordination of the continuing communication.

As an illustration of coordination of knowledge about the facts of the world, consider how *definite reference* is achieved (Clark 1992, p. 107). An example is a simple communicative act such as explaining to a tourist where to find a restaurant she is looking for: it involves a complex series of further requests and information extensions, as well as corrections, nods, and interjections. Creating such a reference is a coordination problem that is rarely reduced to uttering the right word at the right time. What is required instead is a process of mutual adjustment between speaker and addressee converging on a mutual acceptance that the addressee has understood the speaker's utterance. The process is highly iterative, involving a series of reciprocal reactions and conversational moves usually concluded by assent signals. Conversational adjustments toward mutual agreement typically resort to both the discrete resources of spoken language and the continuous resources of gesture, intonation, and other bodily signals.

### 3. The evolutionary roles of coordination of common ground

During the evolution that led to *Homo sapiens*, our hominin ancestors developed new forms of cooperation that made it possible to organize their societies in new ways. It is generally agreed that hominins evolved in open landscapes that favoured a long-ranging life style (Preuschoft & Witte, 1991; Hilton & Meldrum, 2004). As a part of their adaptation they changed their diet from predominantly vegetarian to more protein and fat based. The first culture along the Homo lineage is associated with the finds at Oldowan (Isaac, 1982). The Oldowan lifestyle was in a way signified by an extension in time and space. For example, there were long delays between the acquisition and the use of the tool, as well as considerable geographical distances between the sources of tool raw material sources and killing sites.

### 3.1 Referring to absent objects

In this type of environment, it became increasingly important to jointly refer to objects that are not present on the scene.<sup>2</sup> If the common goal is present in the actual environment, for example food to be eaten or an antagonist to be fought, the collaborators need not communicate before acting. If, on the other hand, the goal is distant in time or space, then a common representation of it must be obtained before cooperative action can be taken. For example, building a shared dwelling requires coordinated planning of how to obtain the building material and advanced collaboration in the construction. The possibility of achieving joint attention to absent entities opens up for new forms of cooperation. This introduces selective pressures towards a communicative system that makes it possible for members of a group to share mental representations of non-present entities (Gärdenfors & Osvath, 2010; Gärdenfors, Brinck & Osvath, 2012).

Symbolic communication is based on the use of representations as stand-ins for entities, present or just imagined. This form of communication is “displaced” (Hockett, 1960) or “detached” (Gärdenfors, 2003), since it typically refers to non-present entities or events.<sup>3</sup> Use of such representations replaces the use of environmental cues in communication. If somebody has an idea about a goal she wishes to attain, she can use language to communicate her thoughts. In this way, language makes it possible for us to coordinate common grounds.

A wide range of communicative tasks can be performed by single words or a combination of a few words (or iconic signs). There are two main communicative situations, however, both unique to humans, where sentential structures play a crucial role: (i) cooperation for future goals and (ii) narratives, in particular gossip.

### 3.2 Communication for future goals

Planning for future collaboration, essentially a task of coordinating goals, requires several forms of coordination of commons ground: coordination in

<sup>2</sup> As regards cooperation among animals, there is no evidence that such communication is used.

<sup>3</sup> Iconic communication can also be used for displaced communication, but, unlike arbitrary symbols, the icons then exhibit some similarity with what they stand for (Gärdenfors, 2003).

space (often outside the present visual field), joint reference to absent objects, coordination of goals and coordination of actions. Such planning depends on forming joint intentions, an advanced form of intersubjectivity presumably unique to humans (Gärdenfors, 2003, 2008; Tomasello et al., 2005). A *joint plan* can be described as a combination of forming a joint intention and coordinating actions.

In previous work (Brinck & Gärdenfors, 2003; Gärdenfors, 2003, 2004; Osvath & Gärdenfors, 2005, Gärdenfors & Osvath, 2010), it has been argued that symbolic language makes it possible to efficiently cooperate about future goals. Along the same lines, Tylén et al., (2010, p. 6) write:

Analogous to the way that manual tool use has been shown to enlarge the peripersonal space by extending the bodily action potential of arm and hand in space ..., linguistic symbols liberate human interactions from the temporal and spatial immediacy of face-to-face and bodily coordination and thus radically expand the *interaction space*.

I submit that the evolution of symbolic language generated evolutionary advantages for the individuals of a society built around cooperation toward future goals.

The transition from an animal signalling system to a symbolic language was, most likely, not made in one step. Bickerton (1990) and other researchers (e.g., Dessalles, 2007) propose that there was a stage in the evolution of language when a protolanguage, containing only the semantic components of language, was used. According to Bickerton, *Homo erectus* mastered a protolanguage and it is not until *Homo sapiens* that one finds a language with a grammatical structure. It is possible that the coordination of common ground required for forming a common plan for future actions can be achieved in a communication system that lacks syntax, that is, in a proto-language. Nevertheless, some sentential structure is necessary since a joint plan involves a series of coordinated future events. As I argue below, describing the planned events requires communication that refers to actions as well as the agent and patient of the action.

### 3.3 The evolutionary role of gossip and other forms of narratives

In social species, individuals often face a decision whether to cooperate or not. In the analyses of prisoners' dilemmas and similar games in standard game theory, it is taken for granted who the potential collaborators are. In practice, however, the most important question is: How do you know *who* to cooperate with? Here I agree with Dessalles (2007, p. 360): «Some of our ancestors who belonged to the first species of Homo, say, began to form sizeable coalitions. In such a 'political' context, finding good allies becomes essential». This type of information is an important special case of coordinating common ground.

Reciprocal altruism ("you scratch my back and I'll scratch yours"), is found in several animal species. *Indirect reciprocity* is a more extreme form of altruism: "I help you and somebody else will help me." The conditions for this to evolve as an evolutionary stable strategy have been modelled (e.g. Leimar & Hammerstein, 2001; Nowak & Sigmund, 2005). The key concept in Nowak and Sigmund's (2005) evolutionary model of indirect reciprocity is that of the *reputation* of an individual. An individual *i*'s reputation is built up by members of the society observing *i*'s behaviour towards third parties and then spreading this information to other members of the society. To wit, gossip becomes a way of achieving societal consensus about reputation (Dunbar, 1996; Slingerland et al., 2009). In this way the reputation for *i* being a 'selfless' helper can be known by more or less all the members of the group. The level of *i*'s reputation can then be used by any individual when deciding whether or not to assist *i* in a situation of need. Reputation is not, of course, something immediately visible to others in the way of such status markers as a raised tail among wolves. Instead each individual must keep a private account of the reputation of all others with whom she interacts. Semmann et al. (2005) provide a nice experimental demonstration that building a reputation through cooperation is valuable for future social interactions, not only within but also beyond one's social group. Tirole (1996) argues that not only individual reputation, but also collective reputation plays an important role in societies: «Countries, ethnic, racial or religious groups are known to be hard-working, honest, corrupt, hospitable or belligerent» (Tirole 1996, p. 1).

In general, the communication required for functioning forms of indirect reciprocity concerns different aspects of who you can trust. The information is often conveyed in the absence of the individual discussed – and it can hence be characterized as *gossip*. Gossip normally contains expressions of the form "X

did A to Y,” which involves identifying thematic roles such as agent, action and patient. Thus gossip plays a central role in the evolution of language according to the theory presented here, but it does not function as a replacement for grooming as Dunbar (1996) suggests.

### 3.4 Sentences are needed for the coordination of common ground

The considerations of this section provide some evolutionary reasons for why coordination of common ground is important for the forms of cooperation that seem more or less unique to humans. I have presented two forms of communication for cooperation where sentences are required: coordination of future goals and gossip that help you decide who to cooperate with. Pragmatically, they serve to coordinate the common ground of the interlocutors.

The important thing to note here is that describing planned actions as well as information about who did what to whom are special cases of describing *events*. I conjecture that the capacity to communicate about events is a watershed that distinguishes the communication of language-trained apes from that of humans. Both types of communication can be seen a special cases of *narratives*. In the following section, I will outline a cognitive theory of events that will support this position.

## 4. A cognitive model of events

Why then are events so central in human cognition? One central feature of events is that they are bearers of *causal relations*: An event typically contains information about an *agent* that performs an *action* related to a *patient* that leads to a *result*. Based on these components Gärdenfors and Warglien (2012) and Warglien et al. (2012) present a model of events and event categories in terms of conceptual spaces. I will here briefly outline this model.

A prototypical event is one in which the action of an agent generates a *force vector* that affects a patient causing changes in the state of the patient. The change of the properties of the patient can be described in terms of a *result vector*. As a simple example, consider the event of a person pushing a table. In this example, the force vector of the pushing is generated by an agent. The result vector is a change in the location of the patient – the table (and, perhaps,

a change in some other of its properties, e.g. it is getting warm and dusty). The result depends on the properties of the patient along with other aspects of the surrounding world: in the depicted event, e.g. frictions act as a counterforce to the force vector generated by the agent.

More technically, Warglien et al. (2012) and Gärdenfors (2014) formulate the following central criterion for event representations:

*The two-vector condition:* A representation of an event contains at least two vectors and at least one object – a *force* vector that represents the cause of the change and a *result* vector representing a change in properties of the object.

The structure of the event is determined by the mapping from force vector to result vector. We will call the central object of an event the *patient*. Even though prototypical event representations contain an agent, there are events without agents, for example, in events of falling, drowning, dying, growing and raining. In the limiting case when nothing happens, that is, when the result vector is the identity vector, the event is a *state*. However, identity result vectors can also be maintained by balancing forces and counterforces: for example, when a prop prevents a wall from falling. Given this representation of events, causation can then be modelled by identifying the cause with the force vector and the effect with the result vector.

The two-vector model can be seen as a form of basic schema that can be elaborated by specifying further components. To the minimal representation of an event required by the two-vector condition, a number of other entities (what linguists call thematic roles) can be added: agent, instrument, recipient, benefactive, etc.

The proposed model allows one to represent events at different levels of generality. There are subcategories of events, just as for object categories. For example, *pushing a door open* is that subcategory of *pushing a door*, where the force vector exceeds the counterforce of the door. *Pushing a door but failing to open* it is another subcategory, where the counterforce annihilates the force vector.

A limiting case of our event model, expressed linguistically by intransitive constructions such as “Victoria is walking” and “Oscar is jumping,” is when the patient is identical to the agent. In these cases, the agent exerts a force on

itself: in other words, the agent modifies its own location in space or its properties.

### 5. Sentences express events

The traditional account within analytic philosophy is that a (declarative) sentence expresses a *proposition*. Propositions are taken to be either true or false, that is, to have truth-values. On many accounts, a proposition is identified with a set of *possible worlds*. From the point of view of semantics, as I explain in Gärdenfors (2000, Section 3.3), this is putting the cart before the horse, since possible worlds are cognitively inaccessible entities. Furthermore, most of the examples of sentences in the philosophical discussion involve states – the classical example is “Snow is white.” In contrast to most philosophical theories, I do not assume that there is a semantic mapping between sentences and propositions. The reason is that the meaning of a sentence is to a large extent dependent on its *context*. For example, an ironic communicative act may drastically change the standard meaning of a sentence.

I will here not attempt to account for the drawbacks of the semantics of sentences within analytic philosophy. Instead I want to be more constructive and present the bare outlines of how sentences can be analysed on the basis of events modelled in conceptual spaces. If one takes a cognitive-communicative point of view, as presented in section 2, it is not so obvious why we express ourselves in sentences. Frege’s answer that the meanings of sentences are *thoughts* is simply not sufficient, since nobody knows how a thought is identified (independently of language). So what do sentences refer to?

My basic idea is that *sentences refer to events*. Furthermore, the focus should be on *utterances* rather than on sentences. Utterances are parts of a communicative context that contributes to the meaning, while in the philosophical (and much of the linguistic) discussion, sentences are often analysed as having a meaning that is independent of the context. However, I shall here use the word sentence to take in the role of utterance as well.

Any description of an event is based on a *construal*. The *attention* of the speaker is a selection mechanism for a cognitive event representation. There are, however, other aspects of how a construal is formed (see Croft & Wood 2000, Ch. 3; Langacker, 2008, Ch. 3, for a survey). One aspect is *perspective*. For example, if I and you are located on two sides of a house, I can

say that you are behind the house, if I put myself in the centre, or I can say that you are in front of the house, if I put the house and the direction of its main side in focus. Another aspect is *categorization*. A construal must select a level of generality to describe an object, for example, *terrier*, *dog*, *mammal*, or *animal*. Yet another aspect is the relation to the *common ground* in the communicative situation. For example, when selecting whether to use a pronoun, noun, or name to refer to an individual, the speaker must consider whether the individual or the name is part of the common ground. Speakers have conversational goals in producing construals. Consequently, the construals are contextual, depending on what the conversation partner already knows or believes or will find most interesting.

What minimal elements must a construal of an event contain? A generic way of describing an event is that “something happens to something.” According to the model of events outlined above, the something it happens to is the *patient* (sometimes identical with the agent). Furthermore “happen” is a placeholder for either the force or the result vector. This leads to the following thesis:

*Thesis about construals:* A construal of an event contains at least one vector (force or result) and one object.

On the basis of the notion of a construal of an event, I can now formulate a fundamental connection between the semantics of sentences (utterances) and events:

*Thesis about sentences:* A (declarative) sentence typically expresses a construal of an event.

From a communicative perspective, one can ask why sentences have such a fundamental status, in comparison to other compositions of words. I do not believe there is a unique answer to the question, but I will base my analysis on the levels of communication discussed in section 2. On level 1, instruction, sentences are often not required. When sitting at a dinner table “Salt!” may function as request, albeit not a polite one (polite requests are often concealed as questions). Or if standing in front of a door “Open!” may be an efficient speech act, since the addressee (the agent) and the object are contextually given. When it comes to level 2, coordination of common ground, the situation is different. Here the communication typically concerns agents, patients,

actions, and results that are not present in the context of the utterance. Then the thesis about sentences implies that at least an agent or a patient (expressed by a noun phrase) and a force vector or a result vector (parts of a construal of an event) are elements in what is expressed. Thus the two main components noun phrase and verb phrase have to be present in a linguistic description of an event. This generates a *semantic* explanation of why a combination of a noun phrase and a verb phrase is so fundamental.

The upshot is that on the coordination level of communication, sentences are indeed central units. On level 3, when coordinating their meanings of words, the partners may rely on *definitional generics* as a special tool (Lawler, 1973). For example, “Whales are mammals” and “A wrench is a tool for fastening nuts” are used to express elements of the meaning of *whale* and *wrench*. Grammatically, generics are sentences, but semantically they are atypical since they describe generic information about concepts rather than about events.

For this reason, sentences are natural units of a semantic theory, albeit not as central as philosophers and some linguists want it. In brief, the model of events and the thesis about construals explain the necessity of the central components of a sentence. They thus provide a motivation for a sentence being a *cognitive unit* of communication.

An analogy to a construal is perhaps visual perception leading to judgments of the form “category X is at location Y.” Searle (1983, p. 40) writes: «The content of the visual perception ... is always equivalent to a whole proposition. Visual experience is never simply *of* an object but rather it must always be *that* such and such is the case». He says that his experience of a station wagon must also be an experience of, for example, a station wagon in front of me. Given the thesis about construals, however, the analogy does not capture all kinds of event construals, in particular not the dynamic aspects. Talmy (1988, p. 61) summarizes the position succinctly:

All of the interrelated factors in any force-dynamic pattern are necessarily copresent wherever that pattern is involved. But a sentence expressing that pattern can pick out different subsets of the factors for explicit reference—leaving the remainder unmentioned—and to these factors it can assign different syntactic roles within alternative constructions.

For example, the sentences “Victoria hits Oscar” and “Oscar is hit by Victoria” describe the same event with the aid of two different construals, where Victoria and Oscar, respectively, are put in focus.<sup>4</sup>

Consequently, no simple mapping exists between the role taken in an event and the designation of subject, object or oblique. A sentence expresses a construal representing a particular focus on an event. In English (and many other languages), the most focused role is designated subject and the secondary focus is designated object. Givón (2001) calls these *primary* and *secondary topics*. He writes that topicality «is fundamentally a cognitive dimension, having to do with the focus on one or two important event-or-state participants during the processing of multi-participant clauses» (Givón, 2001, p. 198). As Croft (2012, pp. 252–253) notes, this phenomenon creates problems for all argument realization rules that are based on thematic roles.<sup>5</sup> In agreement with Givón (2001, p. 198), I see topicality not as directly part of event representation, but rather as a central element of the construal process. This setup avoids the problems that arise when event representation and construal are conflated.<sup>6</sup>

### Conclusion: From event thinking to sentence structure

This article has been written from what Clark (1996) calls the action tradition. In support of the position that pragmatics is evolutionarily primary, it is clear that most human cognitive functions had been chiselled out by evolution before the advent of language. Language would not be possible without these cognitive capacities, in particular having rich intersubjectivity, having a memory system that includes episodic memory, and being able to represent future goals (see Gärdenfors, 2003, 2007; Gärdenfors et al., 2012).

In summary, my thesis is that we communicate using sentential structures because human cooperation benefitted evolutionarily (and still benefits) from

<sup>4</sup> Croft (2012, p. 256) describes the passive voice as a *deprofiling* of the causal chain from the agent to the patient. This can be expressed in my terminology by saying that the patient is made the focus (or topic) of the event.

<sup>5</sup> Also Jackendoff (1987, p. 380) accords: «*Subject* is a syntactic relation, not a conceptual one, and syntactic subjects can hold a variety of [thematic] roles».

<sup>6</sup> Mapping subject to primary topic and object to secondary is not the only way of mapping the elements of an event onto language: In *ergative* languages, the agent is mapped onto the subject and the patient to object. Thus, in such a language, the sentence “Oscar is hit” is expressed as object plus verb.

communication about events. I have argued that the basic semantic referent of a sentence is an event (or a state as a special case). Other animals do not communicate about events since they neither have the mental capacities to cooperate about future goals, nor to cooperate via indirect reciprocity, and therefore they have no need for such communication.

My hypothesis that humans have more advanced cognitive representations of events than other species tallies well with the hypothesis that human are, more or less, the only species that has *episodic memory* that allows us to remember individual events and the order in which they have occurred (Tulving, 1985). This is the memory we use when we think of previous episodes and experiences we have encountered, or when we elicit from our memory events we have learned from conversations with others. Without episodic memory we cannot relate or recount anything. Tulving claims that only human beings have episodic memory, but this thesis has been contested by researchers in animal cognition (Osvath, 2010).

Despite all efforts, the apes' linguistic communication is very limited. In the best cases, they reach the level of a two-year-old human child. They seldom create combinations of more than two words. Most of what they communicate is about something they want. A typical example is when the chimpanzee Nim Chimpsky signs "Nim milk, give milk, Laura give Nim milk, more milk." The apes never *tell* anything, but already two-and-a-half-year-old human infant can tell rudimentary stories. Narration presumes event representations. Therefore, a speculative explanation of why language-trained apes do not tell stories is that they do not mentally represent event or do not represent them in a way that corresponds to the structure of sentences.

The narrative ability is central for human communication. But there seem to be no examples showing that Kanzi masters narration. We are still waiting for him to tell a story by the campfire. So, even if Kanzi understands many spoken utterances, they are coupled to a limited repertoire of communicative functions. They mainly consist of his obeying requests and expressing wishes. He remains on level 1 of communication and never engages in coordinating a common ground (level 2), let alone explains the meaning of words (level 3).

## ACKNOWLEDGEMENTS

I gratefully acknowledge support from the Swedish Research Council for the Linnaeus environment *Thinking in Time: Cognition, Communication and Learning* and from the Knut and Alice Wallenberg Foundation for support to the project *Culture, Brain and Learning*. Part of this work was done during a very fruitful stay at Stellenbosch Institute for Advanced Study.

## REFERENCES

- Bickerton, D. (1990). *Language and Species*. Chicago, IL: University of Chicago Press.
- Brinck, I. & Gärdenfors, P. (2003). Cooperation and communication in apes and humans. *Mind and Language*, 18, 484-501.
- Clark, H. (1992). *Arenas of Language Use*. Chicago: University of Chicago Press.
- Clark, H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Croft, W. (1991). *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: University of Chicago Press.
- Croft, W. (2012). *Verbs: Aspect and Argument Structure*. Oxford: Oxford University Press.
- Croft, W. & Wood, C. (2000). Construal operations in linguistics and artificial intelligence. In L. Albertazzi (Ed.), *Meaning and Cognition: A Multidisciplinary Approach*. Amsterdam: John Benjamins, 51-78.
- Dessalles, J.-L. (2007). *Why We Talk*. Oxford: Oxford University Press.
- Dunbar, R. (1996). *Grooming, Gossip and the Evolution of Language*. Cambridge: Harvard University Press.
- Gärdenfors, P. (2003). *How Homo Became Sapiens: On the Evolution of Thinking*. Oxford: Oxford University Press.
- Gärdenfors, P. (2008). Evolutionary and developmental aspect of intersubjectivity. In H. Liljenström, and P. Århem (Eds.), *Consciousness Transitions – Phylogenetic, Ontogenetic and Physiological Aspects*. Amsterdam: Elsevier, 281-385.
- Gärdenfors, P. (2010). Evolution and semantics. In P. C. Hogan (Ed.), *Cambridge*

- Encyclopedia of the Language Sciences*. Cambridge: Cambridge University Press, 748-750.
- Gärdenfors, P. (2014). *Geometry of Meaning: Semantics Based on Conceptual Spaces*. Cambridge, MA: MIT Press
- Gärdenfors, P., Brinck, I. & Osvath, M. (2012). Coevolution of cooperation, cognition and communication. In F. Stjernfelt, T. Deacon & T. Schilhab (Eds.), *New Perspectives of the Symbolic Species*. Berlin: Springer, 193-222.
- Gärdenfors, P. & Osvath, M. (2010). Prospection as a cognitive precursor to symbolic communication. In R. Larson, V. Déprez, & H. Yamakido (Eds.), *Evolution of Language: Biolinguistic Approaches*. Cambridge: Cambridge University Press, 103-114.
- Gärdenfors, P. & Warglien, M. (2012). Using conceptual spaces to model actions and events. *Journal of Semantics*, 29, 487-519.
- Givón, T. (2001). *Syntax, Vol I*. Philadelphia, PA: John Benjamins.
- Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press
- Greenfield, P. M. & Savage-Rumbaugh, E. S. (1990). Grammatical combination in Pan paniscus. Processes of learning and invention. In S. T Parker & K. H. Gibson (Eds.) *Language and intelligence in monkeys and apes: Comparative developmental perspectives*, Cambridge: Cambridge University Press, 540-578.
- Hilton C. E. & Meldrum D. J. (2004). Striders, runners and transporters. In D. J. Meldrum, & C. E. Hilton (Eds.), *From Biped to Strider: The Emergence of Modern Human Walking, Running and Resource Transport*. New York: Kluwer Academic, 1-8.
- Isaac, G. (1982). The earliest archaeological traces. In *Cambridge History of Africa, Vol. 1: From the Earliest Times to c. 500 BC*. Cambridge: Cambridge University Press.
- Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar, Vol. 1*. Stanford, CA: Stanford University Press,
- Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

- Lawler, J. M. (1973). Studies in English generics. *University of Michigan Papers in Linguistics*.
- Leimar, O. & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity, *Proceedings of the Royal Society London B*, 268, 745–53
- Levin, B. & Rappaport Hovav, M. (2005). *Argument Realization*. Cambridge: Cambridge University Press.
- Nowak, M. A. & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437, 1291–8
- Osvath, M. (2010). *Planning Primates: A search for Episodic Foresight*. Lund: Lund University Cognitive Studies 148.
- Osvath, M. & Gärdenfors, P. (2005). Oldowan culture and the evolution of anticipatory cognition. Lund: *Lund University Cognitive Studies* 122.
- Preuschoft H. & Witte H. (1991). Biomechanical reasons for the evolution of hominid body shape. In Coppens Y. & Senut B. (Eds.) *Origine de la Bipedie Chez les Homínides*. Paris: Cahiers de Paleoanthropologie Editions du CNRS, 55-79.
- Savage-Rumbaugh, E. S. & Lewin, R. (1994). *Kanzi: The Ape at the Brink of the Human Mind*. New York, NY: Wiley.
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Semmann, D., Krambeck, H.-J. & Milinski, M. (2005). Reputation is valuable within and outside one's own social group. *Behavioral Ecology and Sociobiology* 57, 611–616.
- Slingerland, I., Mulder, M., van der Vaart, E. & Verbrugge, R. (2009). A multi-agent systems approach to gossip and the evolution of language. In Taatgen, N.A. & van Rijn, H. (Eds.), *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*. Austin: Cognitive Science Society, 1609–1614.
- Talmy, L (2000). *Toward a Cognitive Semantics, Vol 1: Concept Structuring Systems*. Cambridge, MA: MIT Press.
- Tirole, J. (1996). A theory of collective reputations (with application to the persistence of corruption and to firm quality), *The Review of Economic Studies*, 64, 1-22.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28, 675-735.

- Tulving, E. (1985). How many memory systems are there? *American Psychologist* 40, 385-398.
- Warglien, M, Gärdenfors, P. & Westera, M (2012). Event structure, conceptual spaces and the semantics of verbs. *Theoretical Linguistics*, 38, 159-193.
- Winter, S. (1998). *Expectations and Linguistic Meaning*. Lund: Lund University Cognitive Studies 71.
- Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.



# The Origin of Languages. A Constrained Set of Hypotheses

*José Luis Guijarro*<sup>†</sup>  
joscluis.guijarro@uca.es

## ABSTRACT

As it seems impossible to find reliable evidence to back up hypotheses on the origin of our use of the linguistic tool in our acts of communication, I believe that we may start by pointing as accurately as possible to the processes involved, using a methodology that attempts to reach the levels of adequacy proposed by Chomsky, complemented by those suggested by David Marr. If we conclude that human communication and human language may have had different origins, we might find a new perspective which opens a vast field of research.

Keywords: Language, Communication, Evolution, History, Levels of Adequacy.

## 1. Introduction

### 1.1. The lamp and the key

A man was walking home late one night when he saw the Mulla Nasrudin searching under a street light on hands and knees for something on the ground.

“Mulla, what have you lost?” he asked.

“The key to my house,” Nasrudin said.

“I’ll help you look,” the man said.

Soon, both men were down on their knees, looking for the key.

After a few minutes, the man asked, “Where exactly did you drop it?”

Nasrudin waved his arm back toward the darkness. “Over there, near my house.”

<sup>†</sup> University of Cádiz, Spain.

The first man jumped up.

“Then why are you looking for it here?”

Because there is more light here than near my house.”<sup>1</sup>

## 1.2 Unsolved problems that have arisen when musing about the origin of languages

It is widely assumed that we haven't got sufficient and reliable data to build a real scientific theory to account for and explain the origin of human languages<sup>2</sup>. We cannot collect any evidence, so the story goes, as none exists in today's physical world, and our conclusions are impossible to prove. Therefore, the only thing we can do when thinking about these matters is to muse, muse...and muse again.

However, there is musing and *musings*: What I am trying to say is that, since we cannot rely on physical evidence, we may however constrain our musings in such a way that they point to clear notions with which we may be apt to construct a likely model that works. If we were able to do that, our model could be then considered a true scientific theory, whose proof relies on this functioning aptitude. In other words, as we do with other realms of the so-called mind, we can attempt a *simulation* of the processes involved in the making of our linguistic tool.

Before that, however, we need to clear up the muddled field in which most of those musings have landed, and attempt to use constraining methodological tools to arrive at somewhat more precise and accurate concepts.

In this paper, I will try to do just that.

From my point of view, almost every musing about the origin of languages, that I am aware of, concentrates its efforts, like Nasrudin, on an illuminated area which seems clear and self-evident to researchers: human languages arose and subsequently evolved as *the* solution to our species *communicative* need. There had to be, it has been claimed, a simple proto-language which slowly developed into a full language in the sense we know it today.

An important problem that appears with this view is that nobody has ever found any token of this supposed protolanguage, not even in the most, so-called, primitive societies that have been discovered so far. All the languages

<sup>1</sup> See: [http://www.getnewvisions.com/teaching\\_stories/how\\_to\\_read\\_ts.html](http://www.getnewvisions.com/teaching_stories/how_to_read_ts.html)

<sup>2</sup> The use of the plural, language*s*, is consciously adopted for reasons which will be made clear below.

spoken by the human species that we know of show similar degrees of complexity in their different level structures, so that none is apt to become a likely model for that alleged proto-language.

There have been two main ways to try and avoid that problem without really solving it: on the one hand, it has been proposed that the developing of pidgins into creoles might shed some light on the evolution of the human language. Pidgins are indeed less complex than the natural languages of humans, and, furthermore, they arose for communicative needs. Finally, when these pidgins became creolized, they acquired a universal sort of complexity which made them natural to our species, like every other language. In other words, it is assumed that our proto-language was a kind of pidgin, which then flourished into a primeval creole.

On the other hand, some researchers, like Chomsky and followers (see, for instance, Hauser et al., 2002), decided that there is too much ignorance at present to start musing about human protolanguages. So, they have proposed that, as far as they were concerned, language arose suddenly within the human species, and this is the only interesting object to be considered in the present state of our linguistic knowledge – proto-languages, according to them, are just-so-stories not worth caring about.

A less obvious problem (thus, hidden from general consideration) is that other species also have coded languages they use in their communicative acts. Languages which may be prewired in the species or that may be learned from their co-specifics. Is this sort of coded communicative behavior less evolved than our linguistic codes? A related difficulty that would need to be considered is that some human coded signals seem to point to altogether different conceptual spaces all the time (cf. Wilson, 2008). The coded word RED, for instance does not point to the same hue, when talking about, say, strawberries, the nose of a drunkard, the shame which shows in a face, the color of a kind of hair, the stop sign of a traffic light, a political preference, or the state of a bank account. How come? If RED is a coded sign it must have a fixed relationship with a given thing. It can't keep changing all the time. If it does, as it seems obvious, it is a problem to which we must find a solution in our musings. However, in canonical theories about the origin of languages, it is seldom, if ever, tackled.

In this paper, I will try to make all these problems disappear, not by ignoring them, as some have done in the past; no, I am sure those problems arise from a basic misrepresentation of the processes involved. So, in the same

way we make our fists disappear by opening our hands, if we adopt a different point of view, and try to be coherent with it, those unanswered questions will no longer exist in our theoretical landscape.

## 2. Constraining musings

I have found that attempting to seriously reach the three levels of adequacy proposed by Noam Chomsky (1957) is a good tool kit to start cleaning the muddled array of musings that are offered today. In order to make it more accurate, I have decided to complement it with the three requisites David Marr (1982) proposed, in his turn, in order to say something valuable about human cognitive faculties. I am not sure whether or not either Chomsky or Marr would be happy with my decision, but the fact is that, for me, the combination of both seems to be, in practice, a highly efficient means of constraining my own musings on the matter.

### 2.1 Language or Languages?

Let me start by wondering about the poorness of the English vocabulary on the fundamental way to point to our subject matter. English has only one word, *language*, while my native language, Spanish, has three, *lenguaje*, *lengua* and *idioma*, and French has two, *langue* and *langue*. There are perhaps other languages with more terminological pointers, but the three Spanish ones suffice for the time being. Although our three terminological pointers are used in a rather loose way, especially by Spanish linguists or philologists (who are strongly influenced by the French tradition and, nowadays, by the Anglo-Saxon research), the fact is that they seem to be not totally synonymous for “normal” people. Thus, some uses of one, but not the other two words are typically exclusive or, at least, preferred in some constructions (Guijarro, 1998).

Be that as it may, when the origin of language is discussed, Spanish minds are at a loss, for we don’t really know whether the origin of *lenguaje*, the origin of *lengua*, or the origin of *idioma* is going to be tackled. Moreover, are they the result of the same process? Do they have a different origin? What?

It is a blessing for my clarifying efforts that English is so lexically poor in this field, for it forces me to try and point sharply to given conceptual spaces, trying to reach the first Chomskyan level of adequacy, the observational one.

What are we observing, what conceptual place are we signaling with our three terminological pointers that English does not have?

Let me translate our word *lenguaje* as language 1 (or L1, for short) and state explicitly that when I use this pointer I am referring to the cognitive device that allows us to abstract away and formalize (*i.e.*, to give a mental form to) elements of our environment, so that we may store them in our minds, manipulate and use them, as needs be. This meaning is in accordance with the meaning of the “language” of computers, which is primarily a way to store and retrieve information in a machine.

How about *lengua*, or language 2 (L2, for short)? When I speak of L2, I’m pointing to a mental module in Fodor’s terms. That is to say, a dedicated automatic device through which certain incoming information is processed according to strict constraints, and is, then, responsible for producing a set of mental structures that universally underlie our *idiomas*.

*Idioma*, or language 3 (L3, for short) is, thus, the term with which I point to the final state of L2, once it has been formatted by experience in a given linguistic environment. It corresponds to what are known as “natural languages” like English, Spanish, Italian, Swahili, or Lingala.

We have now tried to reach the observational level of adequacy. When I use the English pointer “language”, then, I am signaling three different conceptual spaces. How may I describe them, *i.e.*, how can I reach the second level of adequacy?

It is here that I have found Marr’s three requisites a convenient practical tool. It has helped me to refine the descriptive adequacy effort along three main lines:

- 1) *The computational line.* What are the basic operations that describe the functioning of the observed element? In the case of L1, the basic operations would be those that concentrate on salient aspects of different elements in our environment, abstracting them away and using them as props to mentally reconstruct their relevant identities. In other words, we should research the possibilities and workings of the processes involved in concept formation. When talking about L2, the universal principles and parameters proposed by Chomsky (1981) would offer an accurate description along the computational line. Finally, the computations needed to describe L3 would be those that form the core grammar of any given natural language.

- 2) *The representational line.* One needs to describe the social and/or personal representations that are currently available and those we are going to favor. L1 seems to be widely represented as the other side of the communicative coin, *i.e.*, language/communication. However, once we have made the three term distinction, this is hardly a good representation of L1. My own representation of this element is as one side of the cognitive coin, *i.e.*, language 1/ cognition. Without language 1 there is no cognition; without cognition, L1 is not possible. The representational description of L2 is as a kind of pre-wired blue print which acts as a sieve to allow basic operations described in the computational mode to adapt to social needs and uses. L3, finally, is widely represented as the other side (*i.e.*, the coded side) of the communicative coin, yet again. I will try to refine this representation by imagining it as convenient linguistic tool that helps creating accurate assumptions (by coding them) which may be used in communicative acts as premises to derive relevant conclusions. That is to say: even here, as far as my representation is concerned, L3 is not visualized as *the* other side of the communicative coin. It is only *a part* (a very important part, but a part, however) of this other side of the communicative coin.
- 3) *The implementational line.* Here we should be able to describe how humans at large or given human societies have tried to implement the elements concerned in order to expand and fix them in one way or another. This sort of description is seldom attempted by researchers and, when it is, it is treated as a lateral consideration. That is why, my proposals will be far from explicit in some cases, as I have no background considerations which I might use or criticize. How, for instance, has humanity implemented L1? As it has practically never been pointed to explicitly, there is hardly any implementation available. There are no conscious or mechanical ways to make concept formation easier to achieve. These processes follow a natural path, and the possible efforts to allow for better (or truer) results which some philosophers have tried to design, seem to be overrun by the relevance seeking mechanisms which are pre-wired in our minds (cf. Sperber, Cara & Girotto, 1995). The same difficulty applies to the L2 concept. As it is an inherited trait, it is difficult to impinge on it by implementing it in one way or another. Only L3 has been widely

implemented in human societies: writing, in the first place, allows it to be fixed and more easily observed. The results of these observations, called natural grammars, help people to consciously learn and use it in an effective way. And the expansion it has achieved through the media is enormous. So, the implementational scarcity we have found in L1 and L2 is compensated by the vast array of implementations L3 has developed – a clear consequence of the almost exclusive identification of language with communication.

We have yet to attain the last level of adequacy, the explanatory one. What are the reasons why those three elements have arisen? To answer this question is the goal of this paper. We will try to do that after first making further observations on some of the other concepts involved.

## 2.2. Evolution and/or history?

Let me organize the following text schematically for greater clarity.

- A) **OBSERVATIONAL** level of adequacy: when I talk about “evolution” I am pointing to a biological process of change, based on genetic material and resulting in new biological elements. When I talk about “history” I am pointing to social processes, in which the features of a given element are apt to be “infected” (cf. Sperber, 2000) by some of the characteristics of the social environment and made to change in order to cope with them. The biological changes that transformed a kind of ape into a proto-human, and this proto-human into a human being indeed form part of an evolutionary process. The processes which turned, say, Iberians into Hispano-Romans and these into Hispano-Arabs, before finally arriving at the Spaniards who live today in the peninsula are, then, historical processes.
  
- B) **DESCRIPTIVE** level of adequacy:
  - 1) Computational description:

According to Darwin, evolution happens by natural selection, a process guided by successful genetic mutations which adapt themselves to solve specific problems encountered in the environment. As this idea destroyed the previously accepted notion of a purposeful divine plan, it had to struggle for acceptance so vehemently that, nowadays, when it is finally widely accepted as *the* true story, any other account suggesting different operations may have led to evolution is immediately treated with suspicion and normally rejected. However, natural selection is not a dogma, but a scientific account which may be refined or complemented with other scientific accounts. The one I have in mind has been proposed by Lynn Alexander (1967) who signed it under the name of her then famous husband, Carl Sagan, in which she proposed,

[...] a theory of the origin of eukaryotic cells (“higher” cells which divide by classical mitosis) is presented. By hypothesis, three fundamental organelles: the mitochondria, the photosynthetic plastids and the (9+2) basal bodies of flagella were themselves once free-living (prokaryotic) cells (Sagan, 1967, p 255).

It may not be apparent to everybody, but what the author is suggesting is that sometimes evolution uses a different mechanism to the one proposed by Darwin. Evolution in this case happens when existing entities find it useful to join others to improve their living possibilities, creating a new more efficient entity which may, then, enter the natural selection process. This theory is known as *Serial Endosymbiosis Theory* (*SET*, for short). I am proposing it here as a complement to natural selection operations, for it may help us understand some of the conundrums that scientific research has faced in our topic.

We will now describe history computationally: its functioning is really very different to both the natural selection processes and the endosymbiotic operations. History, as I said earlier, describes the development and change of social events according to adjustment operations which take into account the representational character of these events and the surrounding features of the social environment. Those operations are metaphorically closer to epidemiology computations than to evolutionary operations. In epidemiology, in effect,

the success of a contagion depends on, say, the strength of some agents (*i.e.*, bacilli or viruses) and the weak defenses a set of environmental features has at any given moment. If social elements are viewed as representations, the changes that history may produce in some of them will be due to the weakness they show in the face of new and strong representations. So, in our computational description of history we must find an explicit relationship between those two forces at work and establish their respective import in ensuing changes.

2) Representational description:

Despite its many critics, Darwin's theory has proven to be very accurate and helpful to explain the diversity of species in the world. Its success has been so evident, that many metaphorical extensions have been proposed in order to explain facts which don't belong to the changing genetic world. One of the first metaphorical expansions of this biological concept was *social Darwinism*. Its main assumption is that the biological process of natural selection, *i.e.*, the survival of the fittest, could (and should!) also be used when talking about social and political entities. It thus became a so-called *scientific* alibi to defend racist (and sexist) political stances, responsible for pogroms and holocausts. A closer examination of the metaphor, however, shows that it really was an ideological tergiversation of Darwin's idea, based on Herbert Spencer's notions. Social Darwinism is now widely considered an abomination and has, thus, disappeared from the serious scientific landscape.

More successful has been Edward O. Wilson's metaphor known as *sociobiology*. He claimed that Darwin's idea of biological natural selection could be applied to the evolution of social behavior, which, in his view, could thus become, genetically prewired. Aspects, such as aggressiveness, group bonding, religious feelings and human speech capacity are natively determined although they may show different cultural features. This approach has also been strongly criticized, for similar reasons to those above.

Eventually, when the cognitive paradigm overturned behaviorism, what seemed a new modern metaphorical extension of the evolutionary idea, known as *evolutionary psychology*, offered a new representation of the

evolutionary notion. Instead of concentrating on the evolution of behavior, it has tried to describe how certain human psychological faculties are indeed the effect of adapting the mind to vital problems. Let me quote the creators of this trend directly:

[...] human psychological architecture contains many evolved mechanisms that are specialized for solving evolutionary long-enduring adaptive problems and [...] these mechanisms have content-specialized representational formats, procedures, cues and so on (Tooby & Cosmides, 1992, p. 64)

Although I am not an expert in biological evolution<sup>3</sup>, it seems reasonable to believe that the real goal that moves the evolutionary process is the attainment of an adaptive function; once a vital problem has been solved by a given adaptive function, it may be housed in a biological organ and not vice-versa –*i.e.*, an organic mutation with no given function would not have any reason to become permanent. Therefore, I think this apparent metaphorical expansion of the evolution concept really represents a basic complement to the Darwinian notion of evolution, provided we agree on the fact that the term, “mind” does not refer to a different organ to the term “brain”, but only to a different descriptive perspective, one physical, the other cognitive.

[...] cognitive descriptions and physicalist ones are not equivalent, but complementary. They cannot be reduced to each other. For this reason, the information-processing descriptions of cognitive science are not merely metaphors in which brains are compared to computers. Their status as an independent level of psychological explanation can be established by considering the fact that the same information-processing relationships can be embodied in many different physical arrangements (Tooby & Cosmides, pp. 65-66)

As has been already mentioned, the representational description of historical changes can be viewed as the infectious power of some representations that impinge on existing socially shared representations, thereby forcing them to become different. Thus, in this paper, I will try

<sup>3</sup> However, «[...] anyone of average intelligence should, given goodwill and a little effort, be able to master enough of the literature in all relevant disciplines to avoid making gross errors» (Bickerton, 2003, p. 79).

not to represent these changes as driven by evolutionary mechanisms<sup>4</sup> for the reasons given above, although there are moments when both points of view seem to coincide.

### 3) Implementational description

I am not really conversant with the possible implementations that the process of evolution may have had in our world. I suppose, however, that there are computer programs that may simulate it accurately<sup>5</sup>. Perhaps the same is true for historical changes, and it would be interesting to compare them and see how differently they are structurally organized. However, at the present state of our musing, this is really a secondary goal which, therefore, will be left untouched for the time being.

## 2.3 Human communication

This is the last term that we are going to analyze along the levels of adequacy before trying to explain how I propose to relate them in my origin of language hypothesis. Let us also do it in a schematic way:

### A) OBSERVATIONAL level of adequacy:

When I use this term, I try to point to the process by which people make their private representations public (so that others may share them) by some sort of intentional behavior. If there is no intention, I will not talk about communication, although some information may be available. For instance, the nasal hue of Jane's discourse can inform Peter that she has a heavy cold, but in fact she is not communicating it –unless she simulates this in order to mystify him. Similarly, smoke does not communicate that there is fire somewhere; it just informs us that this is so.

### B) DESCRIPTIVE level of adequacy

#### 1) Computational description

<sup>4</sup> As it is widely done by some researchers, like Dawkins (1976), who uses the metaphorical notion of “memes” to do the same work in the changes of our representational world that genes do in the physical one, overlooking the huge differences that distinguish biological mutations, almost unnoticeable in particular cases, from representational mutations which happen in every instance of their transmission.

<sup>5</sup> In this respect, the work of S. Kirby and colleagues at the University of Edinburgh has been pointed out to me, especially, Scott-Phillips, Thomas C. & Simon Kirby (2010).

The essential operations that are, thus, needed for communication to work are, in a sense, a chain of embedded representations of the following sort: an observable behavior shows an intention to inform about something, or, more schematically:

{communicative intention [informative intention (that X is the case)]}(Sperber & Wilson, 1985/96)

The first representation (*i.e.*, the communicative intention) is processed as a *fact*, or, in other words, a representation that is processed and kept directly in the mental box of representations, as it were. All the other representations are, as we may see, embedded inside the factual one, and thus, they are considered assumptions (Sperber, 1994). Thus, if Peter starts talking, this verbal behavior is a fact directly processed by Tony. We could render it like this: “it is a fact that Peter is talking”, or, shortly: “Peter says”. If Peter wants Tony to know that he is leaving, the computations involved would look like this:

{Peter says [and by saying he wants to inform Tony that (he is leaving)]}(Curço, 1995)

## 2) Representational description

As noted before, a very wide social representation of human communication is closely linked to the linguistic processes. Human communication is thus represented as consisting chiefly in the ability to code and decode messages in the form established by our L3. Although it would be foolish to deny that L3 plays an extraordinary (and probably species-specific) part in making human messages quite accurate, my representation of its importance is somewhat different to the one which is almost universally held. Therefore, it needs some previous descriptive effort on my part.

Living beings can be described as devices that must cope and adapt themselves to their environment. The processes which allow for this goal to achieve some kind of success may be very roughly described in the following way.

(1) Those processes may be triggered automatically by given stimuli. In the human case, they work in a non-conscious way, even in cases where we have conscious knowledge about its uselessness. For instance, we close our eyes automatically when somebody waves a finger near them, although, at some given occasions, we may positively know that this movement does not represent any danger whatsoever.

(2) A very close (if not the same) process consists in reacting to coded signals which we decode also in an unconscious manner. Such codes may be due to (a) genetic pre-wiring, to (b) imprinting (*i.e.*, formatting it along the lines of a prewired sieve), and to (c) learning it in a conscious manner. Examples of these may be found in animal species such as the dancing of the bees (a), the almost immediate knowledge that ducklings acquire as to who their mother is (b), the song of some birds (c); in humans, we have inherited codes which allow us to interpret our co-specific's faces (a), we acquire L3 by (b), and we learn how to transform it into writing (c). The structurally coded stores of data are known in cognitive science as "modules", for they are encapsulated independently of each other and work only according to their own rules: our identification of human faces, say, does not impinge in any way on our ability to speak our L3 or vice versa.

(3) However, we are able to relate the output results of these modules in one way or another. One canonical theory has it that there is a sort of mental general mixer which does just that (it is thus known as the *central* processor). Another point of view establishes that at a certain point, the modularized processes become prone to partake in other final modular processes (Fodor, 1983; Sperber, 1994; Carruthers, 2003a, 2003b, 2003c). Be it as it may, the fact seems to be that there is a third way for humans to cope and adapt themselves to their environment: by *interpreting* it, using a kind of logical operations, known as *inferences* – *i.e.*, deriving conclusions from mentally represented premises. It would then seem to be a steady development of adaptive behavior, from the automatic physical responses, through the (de)coded ones, to end up with interpretative processes that may be used in totally new circumstances with quite a reasonable ratio of adaptive success. Human communication is best representationally described as this sort of interpretation which

must be strongly constrained by the principle of optimal relevance (a realization of the principle of minimum effort) (Sperber & Wilson, 1986/1995).

However, it is true that a continued interpretation of a given behavior might finally be coded to ease its functioning. Let me give you a personal example. When I bought my last car, I went into an underground parking place to leave it there. When I went out, the car started to whistle “Oh Susanna”. I needed almost a minute to realize that I had left my lights on. Thereafter, however, as soon as the car started whistling the first bars, I immediately knew what the matter was, and acted quickly. The whistling that had started as a prompt to be interpreted, had become codified in my mind. I am sure that in our primeval communications, some of our groans, shouts, etc., became so coded. No doubt about that. What I am arguing is that this faculty, as soon as symbiosis was achieved, began to point to internal psychological states available by our L1 cognitive nature. And as psychological states may never be proven to be the same in every mind, humans went further and, on top of decoding, managed to retain the interpretative faculty to be able to interchange human messages<sup>6</sup>.

### 3) Implementational description

Our world seems to be massively dedicated to implement the possibilities of communication. It started by inventing external representations (pictures or other signals which some early societies used for this purpose, cf. Lewis-Williams, 2002), adapting it to messages couched in L3 written format. This last move was so successful that it is one of the main reasons for the wide representation of L3 as the other side of the human communication coin. One tends to view linguistic discourses and, even more, linguistic texts<sup>7</sup>, as the *only* perceivable reality of communicative acts. But as I hope it is clear, now, there is a lot more to decoding oral speech or to reading linguistic texts in order to successfully achieve some sort of communicative act. We will insist on this in the following section. Nowadays, the extension of communication through space is so huge that it surely surpasses the limits of this paper.

<sup>6</sup> A somewhat similar, but more detailed account, without the language triadic distinction, is made by Origgi & Sperber (2000)

<sup>7</sup> Texts are here considered to be recorded instances of communicative discourses.

### 3. Explaining the constrained concepts we have presented so far

The last level of scientific adequacy is precisely the purpose of our present musings. Why and how did some of the above “elements” appear among humans and develop into their present condition. How are, language (or, rather, L1, L2 and L3) and communication related when considered within the evolutionary process? Did one depend on the other, or were they different evolutionary adaptations? What were the reasons for their origin in our species? How did it come about?

I am aware that my musings represent an interpretation of the “objects” I have tried to observe and describe in a most explicit way. As I have no real hard evidence to back my musings, the only way to give a scientific character to them is to constrain them sharply and consider how they offer a likely biological / psychological advantage to our species and thus have become a permanent feature of our species-specific endowment. Although much will remain vaguely explained or even unexplained, I do think that further research along the present lines may bring a novel outlook which may become a source for further productive hypotheses.

If we rely on the natural selection mechanism to describe evolution, the sudden rise of a given element must be considered a side effect with no adaptive function which was subsequently acquired (Gould, 1993). I believe that this suddenness must be due, rather, to our lack of evidence –it is a “hole” in our thinking, rather than in biology. Therefore, as long as I try to explain the origin of a given human trait through the natural selection mechanism, I will presume it to be a gradual development from other species. Damasio (1994) is clear in that respect:

Many simple organisms, even those with only a single cell and no brain, perform actions spontaneously or in response to stimuli in the environment; that is, they produce behaviour. Some of these actions are contained in the organisms themselves, and can be either hidden to observers (for instance, a contraction in an interior organ), or externally observable (a twitch, or the extension of a limb). Other actions (crawling, walking, holding an object) are directed at the environment. *But in some simple organisms and in all complex organisms, actions, whether spontaneous or reactive, are caused by commands from a brain.* (Organisms with a body and no brain, but capable of movement, it should be noted, preceded and then coexisted with organisms that

have both body and brain.)

Not all actions commanded by a brain are caused by deliberation.

On the contrary, it is a fair assumption that most so-called brain caused actions being taken at this very moment in the world are not deliberated at all. They are simple responses of which a reflex is an example: a stimulus conveyed by one neuron leading another neuron to act.

As organisms acquired greater complexity, “brain-caused” actions required more intermediate processing. Other neurons were interpolated between the stimulus neuron and the response neuron, and varied parallel circuits were thus set up, but it did not follow that the organism with that more complicated brain necessarily had a mind. *Brains can have many intervening steps in the circuits mediating between stimulus and response, and still have no mind, if they do not meet an essential condition: the ability to display images internally and to order those images in a process called thought.* (The images are not solely visual; there are also “sound images,” “olfactory images,” and so on.) My statement about *behaving organisms* can now be completed by saying that *not all have minds, that is, not all have mental phenomena (which is the same as saying that not all have cognition or cognitive processes).* *Some organisms have both behavior and cognition. Some have intelligent actions but no mind. No organism seems to have mind but no action.* (Damasio, 1994, pp. 89-90, the emphases are mine).

As L1 is the other side of the cognition coin, it is clear that some sort of L1 exists in other species which do display images internally and retrieve them when needed. Some may even be able to embed representations into other representations. These species are those that may be trained by humans for, at least, they are able to represent and store in their memory something like this:

[Award (for doing X)]

[Punishment (for doing Y)]

Human beings have developed this embedding faculty to attain almost infinite levels of recursion which today seem to account for our sense of humor (Curço, 1995) and other exclusive human thought processes (Guijarro, 2009). This, I submit, is the main characteristic of our L1 (Corballis, 2011). Through it, we are able to organize spatially and temporarily given mental representations in such a way that we consider them items of more abstract

general representations, or inversely, we may find similar elements in them which are the basis to represent time and space relationships among them. This mental capacity of our species developed into organizing abilities which helped our species to order our surroundings in convenient ways. My (very tentative) hunch is that this was mainly achieved by the female individuals of our species, who had the biological instinct to make homes in which to care for their offspring.

Men, on the other hand, went out to forage for food. They were more successful when hunting in groups, needing to communicate their joint actions to the other participants. They did it in the same way some other species do, by making visual signals to those that were near, and by oral noises when the others were out of sight. In this respect, human communication at that time was not different to animal communication.

Both processes, the creation of a mental organizing tool, and the use of visual and oral behavior to communicate joint action, thus, started to solve different adaptive problems, and it is even possible that some of their elements became codified in altogether different modules.

However, at some point, both modules got entangled in one way or another. This would amount to what Daniel Dennett (1995) calls a “good trick”; one that helped to actualize the mental tool perceptually and, at the same time, made the communicative power of the species increase exponentially. It was, metaphorically speaking, a sort of mental “endosymbiotic” process in which a totally new entity was created at one go; a new entity which worked with elements of both modules, but mixing them in a novel and productive way: human linguistic communication.

There should be then no problem to admit that this new symbiotic product appeared suddenly when its two components met and went on functioning thereafter<sup>8</sup>. If we adopt this hypothesis, however, our efforts should be now directed to imagine what advantages were gained through this permanent bonding.

<sup>8</sup> The metaphorical expansion of the theory was suggested by Margulis herself in her lecture at the University of Valencia in 2002 (my emphasis): “... la historia natural, ecología, genética y metabolismo de organismos macroscópicos debe de ser suplementada con un conocimiento preciso del metabolismo y del comportamiento de los microorganismos. La fisiología y la ecología microbianas son esenciales para la comprensión del proceso evolutivo. El comportamiento de los microorganismos dentro de sus propias poblaciones y en sus interacciones con otros determinó el curso de la evolución de la vida. El mundo vivo subdivisible en último término es el *fundamento del comportamiento, desarrollo, ecología y evolución del mundo visible del cual formamos parte y con el cual evolucionamos*”.

First of all, the already mentioned “ordering” origin of L1, became modularized (and, thus, prewired in our species) in L2, with important consequences.

In effect, Chomskyans propose that the basic principles prewired in L2 are MERGE and EMBED (Chomsky, 1995). They are, thus, universal characteristics of all the human natural languages (or L3). This seems to be a unique feature of our species and it helps us to relate, in two complementary ways, a lot of seemingly independent material, establishing all sorts of meaningful links. The merging feature is the cause of the existing multiple levels in human linguistic structure. Thus, any given element,  $X$ , at level  $n$ , may rise to a higher level,  $m$ , by itself, or by merging it to another element,  $Y$  (*i.e.*,  $X_n (+Y) = X_m$ ). However, the embedding faculty allows any given element  $X$  to remain at its level although it may have been merged to another element (*i.e.*,  $X_n \rightarrow X_n+Y$ )<sup>9</sup>. What is rather amazing is that this last faculty seems to be the linguistic realization of the way we store and process mental representations inside other such mental representations, as I mentioned before. It is, thus, an evident human trait which may be responsible of much of the seemingly “spiritual” character that has been traditionally attributed to our species (Guijarro, 2009) and, as such, it seems to be a crucial feature to account for “humanity” (Sperber, 1997).

But let’s leave that idea open for the time being. What interests us at this moment is to analyze the effects that resulted from the symbiotic union between L1 and L2 and the communicative faculty of humans.

I don’t see any reason to doubt that human communication, before the symbiotic event, was in any way different to that of other species. Intentionally ostensive acts would be the gist of that sort of behavior, in which, either by gestures or by making some kind of sounds, individuals could point to certain objects and events which would then become manifest to other individuals. It may even be possible that some gestures and/or sounds became codified and used in similar way by members of a given social group.

However, when symbiosis did occur, the human communicative process began to differentiate itself from that of other species. The direct (codified or not) signaling to external objects and events shifted somehow, making it possible to point to mental states which represented external objects and events. These mental states had been already structured by L1, probably in an

<sup>9</sup> This is the gist of the so-called X-bar model in syntax (cf. Radford, 1988)

image-like (Rivière Gómez, 1982) format. Maybe, then, the symbiotic gain was to allow those primitive direct codings to push the abstracting processes a bit further, changing the image-like format into a propositional one (Levelt, 1989) couched in the coding actualizations that each group had assimilated – *i.e.*, in the L3. However, if it all had stopped there, the human way to communicate would not be so radically different to the one other (even close) species have. It would basically be a question of coding messages and decoding them, which is the general mechanism of animal communication in our world. We also code and decode, of course; but we do not convey our messages by this process. Our coding mechanism allows us to create very specific indices of our internal mental states which are thus presented as premises, along with many others that are evident in our surroundings. We, then, perform complicated but almost mandatory logical operations (Sperber & Wilson, 1986/1995), *inferences*, whose results are astonishingly accurate, for it looks as if we have read the minds of other individuals (Grice, 1969).

#### 4. Conclusions

Once we have presented a tentative explanation, it seems reasonable to end this paper by putting forward the following conclusions:

- 1) The origin of L1 is parallel to (if not the same as) the evolution of our cognitive device<sup>10</sup>. It rests mainly in the growing ability to embed representations into other representations, allowing thus to order and classify objects which are considered to form part of a given more general concept.
- 2) The origin of L2 is the result of fixing some L1 abilities into a modular sieve which permits humans to acquire (*i.e.*, to imprint) their mother language (or L3) and use it thereafter to facilitate its ordering original purpose.

<sup>10</sup> This idea is gaining weight silently but seriously: see, Donald (2011), Sperber & Wilson (1986/1995), etc., among others. However, none, so far, have been able to distinguish the three types of language. Thus, there is no real way to ascertain clearly *which* is the one concerned.

- 3) The social changes of L3 which have resulted in many different natural languages need not to be considered metaphorically as evolutionary processes. This only muddles the issue and offers no real help in understanding them. A better metaphor to use in this particular case is the epidemiological one which does not have to force the natural selection device, or the symbiotic one, to account for the changes. It may explain them by considering some changing elements to be very salient and thus contagious, provided the environment offers enough reasons for it.

Moreover, the *historical* changing of the human languages (L3) is due to the fact that they do not communicate solely by the coding-decoding process. The reason is well schematized in Sperber & Origgi (2012, p. 337):

(...) a more advanced language faculty, which leads those who possess it to internalize a richer [*i.e.*, different] code than the one present in the community, may emerge and evolve. In a coded based system, every departure of the common grammar will be disadvantageous, or at best neutral: it will never be advantageous.

This, by the way, and as the authors reasonably claim, is also the reason for the huge expansion of our linguistic codes, while other animal codes must remain restricted in order to be functionally successful.

- 4) The human communicative ability evolved in the same way for very many species: the intentional ostensive (at times codified) behavior of one individual was interpreted by others as pointing directly to objects, events, or desires which were thus made manifest. After the evolutionary process of symbiosis took place, humans added a further source of premises to allow for this sort of interpretation: the mental states. It is not that some mental states of other species are not used in their interpretation of messages. It is, rather, that, after symbiosis, humans have been able to construct very accurate premises of these states by using their linguistic abilities and, thus, use them in almost the same way as those they extract from their environment.<sup>11</sup>

<sup>11</sup> Although, when they enter into conflict, the premises of the environment usually are deemed more important than those couched in linguistic form. You may thus say “how nice he is!” and show a gesture which points to the fact that you are being sarcastic and therefore you mean exactly the contrary than what you say.

After all this clearing effort has been attempted, a lot more needs to be done to refine the hunches I have presented here. My purpose, though, was to start a novel way to look at the evolution of language and communication as two separate processes that, due to a symbiotic union, have become a species-specific human trait with unique characteristics. In other words, I have tried to move to the dark place where Nasrudin's key is to be found. Once I got hold of it, I used it in the lighted place in order to find the keyhole that opened a door to a new perspective.

## REFERENCES

- Bickerton, D. (2003). Symbol and structure: a comprehensive framework for language evolution. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution* Oxford: Oxford University Press, 77-93.
- Bickerton, D. (2009). *Adam's Tongue*. New York: Hill and Wang.
- Burge, T. (2010). A real science of mind. *New York Times* (19-XII-2010): <http://opinionator.blogs.nytimes.com/2010/12/19/a-real-science-of-mind/>
- Carruthers, P. (2003a). Fodor's Problem. *Mind and Language*, 18, 502-523.
- Carruthers, P. (2003b). Is the Mind a System of Modules Shaped by Natural Selection?. In Hitchcock, C. (Ed.), *Contemporary Debates in the Philosophy of Science*. Oxford: B. Blackwell, 293-311.
- Carruthers, P. (2003c). Moderately Massive Modularity. In A. O'Hear (Ed) (2003). *Mind and person*. Cambridge: Cambridge University Press.
- Chomsky, N. (1957). *Syntactic Structures*. Mouton: The Hague.
- Chomsky, N. (1981). Principles and parameters in Syntactic theory. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in Linguistics. The Logical Problem of Language Acquisition*. London & New York: Longman. 32-75.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge: The MIT Press.
- Christiansen, M.H., & Kirby, S. (Eds.) (2003). *Language Evolution*, Oxford: Oxford University Press.

- Corballis, M. C. (2011). *The Recursive Mind. The origins of Human Language, Thought and Civilization*. Princeton: Princeton University Press.
- Curçó, C. (1995). Some observations on the pragmatics of humorous interpretations: a relevance theoretic approach. *UCL Working Papers on Linguistics* 7, 27-47.
- Damasio, A. (1994). *Descartes' Error. Emotion, Reason and the Human Brain*. New York: Avon Books.
- Dawkins, R. (1976 /1989). *The Selfish Gene*. Oxford: Oxford University Press.
- Dennett, D. (1995). *Darwin's Dangerous Idea*. London: Penguin Press.
- Donald, M. (2011). The first hybrid minds on earth. In A. W. Geertz, & S. J. Jeppe (Eds.), *Religious Narrative, Cognition and Culture. Image and Word in the Mind of Narrative*. Sheffield: Equinox Publishing Ltd. 67-96.
- Fitch, W.T., Hauser, M., & Chomsky, N. (2005). The evolution of the language faculty: clarifications and implications. *Cognition*, 97, 179-210.
- Fodor, J. (1983). *Modularity of Mind*. Cambridge: MIT Press.
- Fodor, J. (1985). Précis of modularity of mind. *Behavioural and Brain Sciences*, 8, 1-12
- Gould, S. J. (1993). Fulfilling the spandrels of world and mind. In J. Selzer (Ed.), *Understanding scientific prose*. Madison: University of Wisconsin Press, 310 - 325
- Guijarro, J. L. (1998). Idioma, lengua y lenguaje y otras cosas del mensaje. In J. Martín, F. Velázquez & J. Bustamante (Eds.). *Estudios de la Universidad de Cádiz ofrecidos a la memoria del Profesor Braulio Justel Calabozo*. Cádiz: S.P.U.Ca.
- Guijarro, J. L. (2009). A possible way to naturalize the human soul. *Pragmalingüística*, 17, 40-51.
- Grice, H. P. (1969). Utterer's meaning and intentions". *Philosophical Review*, 78, 147-177.
- Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Hauser, M., Chomsky, N., & Fitch W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?. *Science*, 298, 1569-1579.
- Hirschfeld, L., Gelman, S., (Eds.) (1994). *Mapping the Mind: Domain-Specificity in Cognition and Culture*. Cambridge: Cambridge University Press.

- Levelt, W. J.M. (1989). *Speaking: from Intention to Articulation*. Cambridge: The MIT Press.
- Lewis-Williams, D. (2002). *The Mind in the Cave*. London: Thames & Hudson.
- Margulis, L. (2002). *Una revolución en la evolución*, Valencia: Col·lecció Honoris Causa. Universitat de Valencia.
- Marr, D. (1982). *Vision: A computational investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.
- Mayr, E. (1984). What is Darwinism Today?. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2, 145
- Mora Millán, M. L., (Ed.) (2008). *Cognición y Lenguaje. Estudios en homenaje a José Luis Guijarro Morales*. Cádiz: Servicio de publicaciones de la universidad de Cádiz.
- Origg, G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In Carruthers, P. & Chamberlain, A. (Eds.). *Evolution and the Human Mind: Language, Modularity and Social Cognition*. Cambridge: Cambridge University Press. 140-169.
- Pinker, S. (1994). *The Language Instinct*. London: Penguin Books
- Pinker, S., & Jackendoff, R. (2005). The Faculty of Language: What's Special about it? *Cognition*, 95(2), 201-236.
- Radford, A. (1988). *Transformational Grammar. A first Course*. Cambridge: Cambridge University Press.
- Rivière Gómez, Á. (1982). *Razonamiento y representación*. Madrid: Siglo XXI de España, Editores.
- Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, 14 (3), 225-IN6.
- Scott-Phillips, T., & Kirby, S. (2010). Language Evolution in the Laboratory. *Trends in Cognitive Science*, 14(9), 411-417.
- Selzer, J., (Ed.) (1993). *Understanding Scientific Prose*. Madison, Wisc: University of Wisconsin Press.
- Sperber, D. (1994). The modularity of thought and the epidemiology of representations. L. Hirschfeld & S. Gelman (Eds.), *Mapping the Mind: Domain-Specificity in Cognition and Culture*. Cambridge: Cambridge University Press, 39-67.

- Sperber, D. (1997). Intuitive and reflective beliefs. *Mind & Language*, 12 (1), 11-24.
- Sperber, D. (2000). *An objection to the memetic approach to culture*. In R. Auger (Ed.), *Darwinizing Culture: The Status of Memetics as a Science*. Oxford: Oxford University, 163-174.
- Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition* 57, 31-95.
- Sperber, D., & Origgi, G. (2012). A pragmatic perspective on the origin of language. In D. Wilson & D. Sperber (Eds.), *Meaning and Relevance*. Cambridge: CUP
- Sperber, D., & Wilson, D. (1986/95). *Relevance. Communication and Cognition*. Oxford: Basil Blackwell.
- Tooby, J., & Cosmides, L. (1992). The psychological foundation of culture. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The Adapted Mind*. Oxford (UK): Oxford University Press, 19-136.
- Wilson, D. (2008). Relevance theory and lexical pragmatics. In Mora Millán, M. L., (Ed.), *Cognición y Lenguaje. Estudios en homenaje a José Luis Guijarro Morales*. Cádiz: Servicio de publicaciones de la universidad de Cádiz, 233-247.
- Wilson, D. & Dan Sperber (2012): *Meaning and Relevance*. Cambridge: CUP

# What is Wrong, and What is Right, about Current Theories of Language, in the Light of Evolution?

*James R. Hurford*<sup>†</sup>  
jim@ling.ed.ac.uk

## ABSTRACT

Two extreme and contrasting positions held currently by various researchers in language evolution are compared. Each position comprises five ideas which contradict the corresponding ideas in the other position. In Extreme Position A, there was a single biological mutation, creating a new unique cognitive domain, Language, immediately enabling unlimited command of complex structures via Merge, used primarily for advanced private thought, and only derivatively for public communication (internalism), not promoted by natural selection. By contrast, in Extreme Position B, there were many cumulative biological mutations, allowing expanded interaction of pre-existing cognitive domains – no new domain was created, gradually enabling command of successively more complex structures, used primarily for public communication, and derivatively for advanced private thought (externalism), promoted by natural selection. These extreme positions are not hypothetical ‘straw men’, insofar as prominent researchers exist who adopt each of them. At the end of this paper I will present a ‘scorecard’ summarizing which parts of the two extreme positions are justified by available evidence.

Keywords: nativism, domain specificity, modularity, genes, communication, thought gradualism, continuity.

## Introduction

I will compare two extreme and contrasting positions held currently by various researchers in language evolution. Each position comprises five ideas which contradict the corresponding ideas in the other position.

<sup>†</sup> University of Edinburgh, UK.

In Extreme Position A, there was

1. a single biological mutation,
2. creating a new unique cognitive domain, Language,
3. immediately enabling unlimited command of complex structures via Merge,
4. used primarily for advanced private thought, and only derivatively for public communication (internalism),
5. not promoted by natural selection.

By contrast, in Extreme Position B, there were

1. many cumulative biological mutations,
2. allowing expanded interaction of pre-existing cognitive domains – no new domain was created,
3. gradually enabling command of successively more complex structures,
4. used primarily for public communication, and derivatively for advanced private thought (externalism),
5. promoted by natural selection.

These extreme positions are not hypothetical ‘straw men’, insofar as prominent researchers exist who adopt each of them. At the end of this paper I will present a ‘scorecard’ summarizing which parts of the two extreme positions are justified by available evidence. The paper thus covers a number of issues that have been at the centre of theorizing in linguistics for many decades, issues such as nativism, domain specificity, modularity, and function. Underpinning the argument is the premise that any theory of what language is like must take into account the question of how it could possibly have evolved to be that way.

### 1. Natural Selection

It is convenient here to start with the fifth issue, that of natural selection. Natural selection here is about biological selection of the DNA that ultimately encodes the human capacity for complex language, often called ‘UG’. UG is theorized to determine what languages human infants could possibly acquire, given suitable input experience. It is also often assumed that UG is the main

determinant of what languages are like. For several decades, it was claimed that UG is a richly structured set of principles. It was the job of linguistics to discover these principles.

An old argument, no longer valid, involves arbitrary abstract properties of grammars, which I will illustrate with the example of Subjacency.

The Subjacency principle, allegedly innate, is what prevents children learning English, for example, ever experimenting with strings like the following:

*\* Who do you like the man that saw?*

with a meaning like the emphatically incredulous, and acceptable, echo question

*You like the man that saw WHO?*

The Subjacency principle was one of a handful of similarly abstract principles postulated to govern the organization of grammars in languages. David Lightfoot argued amusingly, and correctly, that such principles, though taken at the time to be innate, cannot be the result of natural selection.

The Subjacency Condition has many virtues, but I am not sure that it could have increased the chances of having fruitful sex. (Lightfoot, 1991)

The programme of searching for such abstract principles of grammar gave way toward the end of the 20th century to a biologically more plausible view, dubbed ‘Minimalism’ (Chomsky, 1993, 1995). Under Minimalism, there are no such arbitrary properties. So the set of properties of language not plausibly attributed to natural selection, i.e. apparently arbitrary properties, disappears. In Minimalism, what remains is the recursive Merge operation, whose application (whether to concepts or to linguistic forms) is universally available to (non-pathological) humans, and occurs in all languages. An ability to merge concepts in ones head to form more complex conceptual representations is obviously adaptive in private thought. An ability to merge meaning-form pairs, Signs, to express more complex meanings in public communication is also adaptive – more on this later.

The ‘no natural selection’ hypothesis would

- either predict that humans who are incapable of Merge (whether privately or publicly) are not pathological cases, not evolutionarily disadvantaged. This is false.
- or claim that the Merge capacity (whether private or public) went to fixation in humans by a population-level statistical accident. This is very unlikely.

So the human capacity to merge (either private concepts or public Signs) is promoted by natural selection. This settles issue 5 in our list.

## 2. Immediate Infinite Capacity

The context is set by the following statement:

there is no possibility of an “intermediate” language between a non-combinatorial syntax and full natural language syntax -- one either has Merge in all its generative glory, or one has effectively no combinatorial syntax at all. (Berwick, 1997, p. 248)<sup>1</sup>

Temporarily, for the sake of argument, assume that such a cognitive capacity, Merge, could result from a single biological mutation, immediately enabling unlimited command of complex structures via Merge. This is impossible, as working memory limitations must have been present from the start. The putative unlimited command can only be theoretical, not real or practical. Of course, no human has literally unlimited command of an infinite range of expressions. Such a set, to be infinite, would necessarily contain an infinite number of strings too long for a brain to process. Rather than there being an infinite set of expressions that a person can manage, there is in fact an infinite set that no one can, or ever could, manage.

This raises the familiar distinction between competence and performance. Competence determines the regular behaviour of individuals. Competence is also thought of as knowledge of unbounded possibilities; practical bounds are held not to apply to a theory of competence. This reflects an early idealistic and naive love affair by linguists with infinite sets.

<sup>1</sup> See Hurford (2011, p. 585-587) for a fuller critique of the formal problems with this bald assertion.

Though conceived and applied almost exclusively in the context of human behaviour, the notion of competence can also usefully be applied to complex birdsong. Complex birdsong is very regular behaviour, based, like language, on an innate template and environmental exemplars. The regularity in birdsong includes clear numerical upper and lower bounds on the number of repetitions of phrases and the overall duration, in phrases, of songs. For example, a chaffinch song includes a repeated phrase of several notes, but the number of repetitions is limited to a range of between 4 and 11. A competence/performance distinction is appropriate for complex birdsong, but the boundary should be shifted to include numerical features of competence, what I have called ‘competence-plus’ (Hurford, 2011).

Performance is associated with two kinds of factor: (1) accidental and temporary factors, e.g. distraction, drunkenness, sudden death, and (2) permanent limiting factors, e.g. processing capacity, storage capacity, short-term memory in conditions of alertness. The latter belong in an integrated component of an individual’s acquired language, competence-plus, which has built-in, rubbery, numerical constraints.

UG, the innate capacity for language, is what, given suitable experience, scaffolds the growth of adult competence in a language. Just as an augmented concept of competence, namely competence-plus, is necessary to account for adult behaviour, so an augmented concept of UG, which I call ‘UG+’, is necessary to account for language acquisition.

Formal UG would not have evolved independently of memory and processing power. What would be the use of innate information about the form of language without a capacity for processing it? What would be the use of power to process language without the prospect of acquiring something to process? Memory and processing power are inherently numerically bounded. I propose a numerically bounded initial state of the language faculty, UG+, a package of formal and numerical information. Complex behavioural dispositions have co-evolved along with a complex cognitive computational capacity to manage them. A modern child is born with UG+ and, on enough exposure to a language, acquires competence-plus in that language.

Practice can, to some extent, extend numerical memory and processing limits. Even with quite restricted working memory, it is plausible that early hominin mutants, in a non-communicative version of events, had an adaptive advantage. Realistically, there must have been some working memory limitations on intuitive judgments and internal thought processes. And of

course there still are. Unlimited, infinite, command of any capacity, cognitive or otherwise, cannot exist in nature. There are always performance limitations.

So we can envisage a mutant with a new, but performance-limited, capacity to combine (i.e. Merge) conceptual units, in an environment where this is adaptive. She would prosper, and the mutation would spread, giving rise to a strain of *Homo* with superior private reasoning power, better tools, better hunting techniques, better shelter, better long-term planning, etc.

### 3. Private Thought versus Public Communication

I now discuss issue 4 of the two extreme positions being compared. How likely is it that increased cognitive computational power was for purely private individual thought, not communicated to other members of the group? Could early *Homo* have been a strain of clever social isolates, each person good at planning his future moves and privately solving environmental challenges as an individual?

The social isolate scenario would have favoured individual selection, but not social group selection. Was social group selection, in addition to individual selection, a factor in the rise of humans? Social group selection is not a directly biological process, but a social process, in which groups compete with each other. Members of culturally more successful groups have greater chances of biological reproduction. Were early hominins more like modern orangutans than like modern chimpanzees and bonobos? Our closest primate relatives live in social groups, form alliances and cooperate to a limited degree. There is inter-group competition in many primate species, usually motivated by food resources. Inter-group competition is positively correlated with group size, hence (indirectly) with neocortex size and levels of tactical deception.

Humans have the largest group size of any primate, the largest neocortex, and the greatest capacity for tactical deception. The humans who spread around the globe probably lived in highly cooperative groups, in competition with other groups. Such in-group cooperation, and inter-group competition, would have fostered public exchange of successively more complex information within groups.

Further, group-specific codes (akin to different languages) are compatible with inter-group competition.

This introduces externalization of (going public with) the previously private conceptual units. And, so far, we have not considered whether the capacity to merge items operated solely on concepts for private thought, or arose as an operation on public symbols. That is, a question of timing arises. There are two simple possibilities, and a more complex one, which I will introduce a bit later. First, the simplest possibilities.

1. Public externalization preceded the capacity to Merge. Even the simplest conceptual units were externalized from an early stage. This is the familiar Bickertonian Protolanguage scenario. In this view, Merge, from the start, involved public signals, i.e. was an operation on meaning/form pairs.
2. The Merge capacity preceded public externalization. The capacity to Merge conceptual units for advanced thought preceded externalization and advanced communication.

In normal humans, complex thought and complex language go together, but in pathology, they can be dissociated. There is an overall correlation between verbal and non-verbal IQ. In human children, even learning simple public labels modifies thought. Bilinguals perform better on certain nonlinguistic cognitive tasks. When we think in words, we use the words of particular public languages, as an aid to thought.

These facts point to a more complex timing possibility, namely that there was co-evolutionary spiral of successively more complex external language and successively more complex private thought, both always bounded by working memory capacity. In this spiral, complexification of public language is the evolutionary driver. The spiral builds on the asymmetry between production and comprehension. Access to more complex thoughts comes through comprehension of sporadically produced complex public expressions.

Socio-historico-cultural processes, such as grammaticalization, led to successively more complex languages. In tandem, such capacities as short-term working memory, long term storage of thousands of Signs, and fast vocal/auditory production and interpretation evolved.

The conclusion is that complexity in private thought and in public communication co-evolved. This is a nuanced solution to issue 4, the relative

timing of going public with language and the capacity for combinatorial language.

#### 4. Was there a Single Mutation?

This is the least controversial issue. The language faculty is not monolithic. Even within the areas traditionally regarded as the core of linguistic structure, namely phonology, morphology and syntax, different organizing principles apply. For example, phonotactic rules determining combinations of phonemes have no obvious counterpart in syntax. Further, competence in these core areas could not be achieved without support from so-called peripheral systems (including storage, working memory, vocal or manual skills, pragmatic skills, etc.). Capacity in all these had to evolve in partnership. True, pleiotropy (one gene, many traits) is possible, but the traits governed by a pleiotropic gene are seldom so functionally coordinated.

Known language-related genes, e.g. FOXP2, do not do the whole job of creating a complete faculty for language. There are very few phenotypic traits that can be attributed to a single gene. And there is no prospect of ever discovering a single gene that accounts for the whole human capacity for language.

In conclusion, the least controversial answer of all these issues – a single mutation underlying language is not plausible.

#### 5. Was a New Domain, Language, Unique to Humans, Created?

The evolved capacity for language has built on pre-existing hierarchical organization of behaviour, semantic memory for facts (storage), and fast routinization of useful procedures, to mention only several pre-existing factors. Seeds of these pre-existing features can be found in rudimentary form in non-human animal behaviour. But in language, each of these is now special to language in some way, and not found in other cognitive or behavioural capacities.

The working memory used in language processing is different from mere digit span usually identified with (non-linguistic) working memory. The numerical constraints inherent in UC+, hence in competence-plus, are

sensitive to different types of grammatical structure, e.g. right-branching vs. centre-embedding. Below is a classic example of right-branching

*This is the farmer sowing his corn  
 that kept the cock  
 that crowed in the morn  
 that waked the priest all shaven and shorn  
 that married the man all tattered and torn  
 that kissed the maiden all forlorn  
 that milked the cow with the crumpled horn  
 that tossed the dog  
 that worried the cat  
 that killed the rat  
 that ate the malt  
 that lay in the house  
 that Jack built!*

Despite the length of this sentence, it poses no parsing problem. It is long enough to need memorizing, but it makes natural sense and its meaning can be understood without undue effort. No nonlinguistic task, in human or non-human life, is comparable in such effortless coping with complexity.

By contrast, as is well known, centre-embedding structures, even quite short ones, are hard to parse, actually impossible beyond a very low limit, as illustrated below.

*This is the malt the rat ate.  
 This is the malt the rat the cat killed ate.  
 This is the malt the rat the cat the dog worried killed ate.  
 This is the malt the rat the cat the dog the cow tossed worried killed ate.*

The difference between right branching and centre-embedding is a matter of specifically linguistic structure, analogues of which cannot be found outside language.

An adult native speaker of a modern language has memorized literally tens of thousands of constructions. Many of these are single words, while others are more complex constructions. A common noun is a relatively simple construction, often requiring only the information that it is a common noun,

plus the accompanying semantic specification of its meaning and the phonological specification of its pronunciation. Pragmatic information, such as degree of formality (register), or taboo status, may also be stored. A transitive verb is a more complex construction, with specifications of the type of object it can take, whether it can be passivized, and so on. Function words, such as auxiliaries and determiners, are also conveniently viewed as constructions. The Merge operation combines constructions into sentences, yielding the spectacular expressive power of human languages. The vast expressive power of languages derives just as much from the massive store of items as from the combinatory power of the Merge operation. If we had only two or three items in our memory store, even a very free capacity to combine them would be extremely impoverished by comparison with a real language. The human capacity to rapidly acquire and fluently use such massive stores of linguistic items is special, and has no clear analogue outside the language domain.

The conclusion is that complex language has evolved to be a unique cognitive domain.

## 6. Summary and Scorecard

Neither of the extreme positions set out at the beginning of this paper is totally correct. The table below gives the ‘score’, as this paper has argued it, for the relative strengths and weaknesses of the two extreme positions.

	Position A	Position B
Natural selection	0	1
Immediate infinite potential	0	1
Private v public	½	½
Single mutation	0	1
New unique domain	1	0

## REFERENCES

Berwick, B. (1997). Syntax facit saltum: computation and the genotype and phenotype of language. *Journal of Neurolinguistics*, 10(2/3), 231-249.

- Chomsky, N. (1993). A minimalist program for linguistic theory. In Hale, Kenneth L. & S. Jay Keyser (Eds.) *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*. Cambridge, MA: MIT Press, 1-52.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass: MIT Press.
- Hurford, J. R. (2011). *The Origins of Grammar: Language in the Light of Evolution*. Oxford: Oxford University Press.
- Lightfoot, D. (1991). Subjacency and sex. *Language and Communication*, 11(1/2), 67-9.



# Communicative Signaling, Lateralization and Brain Substrate in Nonhuman Primates: Toward a Gestural or a Multimodal Origin of Language?

*Adrien Meguerditchian*<sup>\*†</sup>  
adrien.meguerditchian@univ-amu.fr

*Jacques Vauclair*<sup>‡</sup>  
jacques.vauclair@univ-amu.fr

## ABSTRACT

Language is a complex intentional, syntactical and referential system involving a left-hemispheric specialization of the brain in which some cerebral regions such as Broca's and Wernicke's areas play a key-role. Because nonhuman primates are phylogenetically close to humans, research on our primate cousins might help providing clues for reconstructing the features of our ancestral communicative systems. In the present paper, after emphasising the tight relation between gestures and language in humans, we underlie the specific significance of communicative gestures and of the progressive control of the oro-facial system and the vocal tract in the course of the language evolution. For this purpose, we will then review the findings related to the features, the lateralization and brain correlates of both vocal and gestural systems in nonhuman primates.

Keywords: Communication, Gesture, Hemispheric specialization, Primate, Language

## 1. Introduction

Given the phylogenetical proximity between human and nonhuman primates, the researches on the communicative, motor and cognitive systems of our

<sup>\*</sup> Laboratory of Cognitive Psychology UMR7290, CNRS / Aix-Marseille University, Marseille, France.

<sup>†</sup> Station de Primatologie UPS846, CNRS, Rousset, France.

<sup>‡</sup> Centre de Recherche en Psychologie de la Connaissance, du Langage et de l'Émotion, Aix-Marseille University, Aix-en-Provence, France.

primate cousins could help us determining the continuities/discontinuities between human language and animal communication. Such a comparative approach among primate species might thus have some significant implications for reconstructing the features of our ancestral communicative system that have been inherited from our common ancestor and for evaluating the evolutionary prerequisites of language. Most of the studies have focused naturally on the vocal modality and some researchers have suggested that language resulted from the evolution of the vocal system in our ancestors (e.g., Ghazanfar & Hauser, 1999; Snowdon, 2001; Zuberbühler, 2005; Lemasson, 2011). This theory is now challenged by a growing number of authors supporting the “gestural origins” view that gestural communication may be the first phylogenetic precursor of human language (e.g., Arbib et al., 2008; Corballis, 2002, 2003; Kendon, 1991; Kimura, 1993; Vauclair, 2004). Such an alternative gestural theory finds support in the considerable evidence of strong and tight links between gestures and language in humans including infants, adults and deaf people using sign language. It is well known that infants start using gestures for communication before being able to speak. Indeed, before about 12 months of age, gesture is the first mode of intentional and referential communication in infants (Bates, 1976). In other words, infants can not only voluntarily express by gestures a wish or an intention (that they cannot yet be expressed verbally) to a specific social partner and expect a response from the recipient (i.e. intentional properties), but also to direct the adult’s attention toward external objects or events (i.e., referential properties) by pointing gestures (Butterworth & Morissette, 1996). Such gestural productions have been shown to play an active role in the development of linguistic skills. For instance it has been shown that the size of gestures repertoire in infants predict vocabulary development (e.g., Iverson & Goldin-Meadow, 2005). Moreover, it is also very well described that we speak using systematically synchronized expressive manual gestures (co-speech gestures, see McNeill, 2005). Based on several reports showing how the two modalities interact with each other in both production (e.g., Bernardis & Gentilucci, 2006) and perception (Willems et al., 2007), it has been proposed that speech and gesture might share the same integrated communication system (Gentilucci & Dalla Volta, 2008). It has also been well documented that human signed languages share the same “phonological”, morphological and syntactical properties than spoken languages (for reviews: Bellugi, 1991; Emmorey, 2002) as well as some similar left-cortical lateralization and key-

cerebral areas such as Broca's areas (Corina et al., 2003; Emmorey et al., 2007). Given the existence of such strong links between gestures and speech in humans, one can question the implication of the gestural system in the evolution of language.

In the present paper, after reviewing briefly the recent findings on the features of gestural and vocal communication in nonhuman primates and their potential relations to some language properties, we will present research on the behavioral lateralization of these communicative systems and their cortical correlates. We will thus discuss their continuities and discontinuities with the brain hemispheric specialization of language and their implications regarding the origins of language.

## 2. Human language and gestures in nonhuman primates

The gestural origin theory finds also supports with the recent investigation of the properties of the gestural communicative system in nonhuman primates. Indeed, it is well documented that great apes, particularly chimpanzees, and in a lower degree some monkey species, use manual gestures and body movements to communicate with conspecifics in various social contexts and with different goals such as play, threat (see Figure 1), aggression, greeting, invitation for grooming, in case of shared excitement, of reassurance-seeking after stress, and for food begging (e.g., Goodall, 1986; Pika et al., 2005; Call & Tomasello, 2007 for review; see also Cartmill & Byrne, 2010; Hobaiter & Byrne, 2014). Interestingly, this communicative system has revealed some potential continuities with several key properties of human language, such as flexibility of morphology, intentional and referential properties (see for reviews, Meguerditchian & Vauclair, 2008; Pika, 2008). Indeed, although some of species-specific gestures - such as chest-beat in gorillas - seem quite stereotyped and might likely be genetically determined (e.g., Genty et al., 2009), the gestural system has been shown to be very flexible, depending of the individual social experience, eliciting variations of the composition, the morphology and the size of the gestural repertoire between individuals and also between different populations (Pika et al., 2005; Call & Tomasello, 2007 for reviews). In fact, among individuals, different gestures may be produced for the same goal and, conversely, similar gestural signals may be used for divergent goals (Tomasello et al., 1985; Tomasello et al., 1989). Those findings are

consistent with the previous research projects aimed at trying to teach American sign language to apes (Gardner & Gardner, 1969; see also Terrace, 1979 ; Patterson, 1978; Miles, 1990), revealing the remarkable ability of apes to learn and to use novel manual signs (more than one hundred) to communicate with humans rather than novel vocalizations.

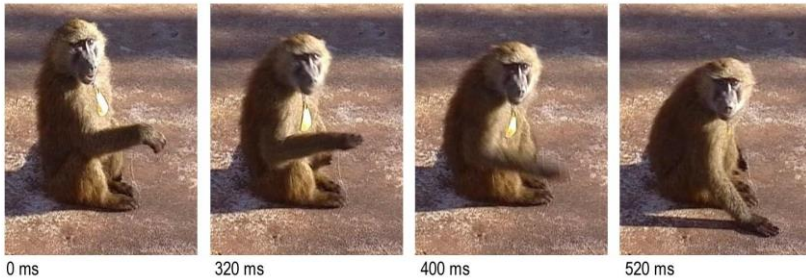


Figure 1. Manual gesture performed by a male baboon. A young baboon intimidates a human observer by quickly slapping his right-hand on the ground. Time is indicated in milliseconds (ms). *Picture: © Adrien Meguerditchian*

In addition, it has been described that gestures can be referential by the use of imperative pointing for getting food from a human (Call & Tomasello, 1994; Leavens & Hopkins, 1999; Leavens et al., 2004) or to inform about the location of a tool (Zimmermann et al., 2009). Similar observations of referential gestures have also been made in chimpanzees and bonobos within intraspecific interactions for indicating to a conspecific a particular spatial location or an object in the environment (Inoue-Nakamura & Matsuzawa, 1997; Veà & Sabater-Pi, 1998; Hobaiter et al., 2014; Genty et al., 2014) or an area of the body in order to be groomed (Pika & Mitani, 2006). Finally, we know that the production of gestures in apes and monkeys is under intentional control (e.g., Bard, 1992; Leavens et al., 1996; Tomasello et al., 1994; Genty et al., 2009; Bourjade et al., 2014). Without a doubt, in contrast to vocalisations, the production of gestures involves systematically a social partner (a conspecific or a human), within a dyadic interaction or even triadic interaction when it is referring to external objects (reviewed in Leavens, 2004 and in Pika et al., 2005). Moreover, when the recipient is not responding or is not attending, gestures can persist or be adjusted to the attentional state of the recipient until the goal is reached (Leavens et al., 2005; Cartmill & Byrne,

2007; Tomasello, 2003). For instance, in this latter case, nonhuman primates can change their position to face the recipient before producing visual gestures (e.g., Liebal et al., 2004) or elaborate more auditory or tactical gestures in order to get the attention of the recipient (Tomasello et al., 1994; 1997; Leavens et al., 2010; Bourjade et al., 2014). In short, those collective findings offer solid evidence of intentional communication when involving the manual or body motor system.

### 3. Human language and vocalizations in nonhuman primates

Regarding those latter collective findings, there is some debate whether or not the properties of the vocal system in nonhuman primates is less convincing than the gestural system as the best prerequisite for the emergence of speech (Vilain et al., 2011; Liebal et al., 2013). Recent findings have shown relative flexibility of the vocal system. For instance, audience effects have been shown to affect differentially vocal production depending on the type of social partner in both chimpanzees and monkeys (e.g., Mitani & Nishida, 1993; Wich & de Vries, 2006; Schel et al., 2013a, 2013b; Zuberbühler et al., 2011), thus questioning the potential existence of intentional properties in call productions. Some degree of plasticity in the acoustical structure of calls has also been described between or within social groups in relation to social, environmental and contextual changes in the group (reviewed in Lemasson, 2011), indicating a probable influence of a learning component during the individual's lifetime as well as some control of the production of vocal signals (e.g., for reviews: Roian-Egnor & Hauser, 2004; Meguerditchian & Vauclair, 2008). Nevertheless, although such a flexibility, in comparison to gestures' repertoire, there is very little variation of the content and the size of the vocal repertoire in a given species across individuals and different groups. There is poor evidence that monkeys and apes are able to generate new vocal signals, suggesting that the bases of the vocal features in the repertoire of nonhuman primates might be mostly genetically determined, although they can be influenced on top by a learning component from the social environment (see Roian-Egnor & Hauser, 2004). Moreover, the production of vocalizations is much context-dependent than gestures (Pollick & De Wall, 2007) in being systematically related to a specific reaction to events and likely to a specific internal emotional state (Goodall, 1986). This suggests rather poor intentional

properties. According to this hypothesis, it does not mean that nonhuman primates cannot have some degree of voluntary control on their vocal production in relation to other conspecifics. But the voluntary modulation, exaggeration or inhibition of their calls in relation to specific contexts and audience can still be interpreted as the results of an internal emotional state. From this point of view, this kind of production looks then very similar to the production of any human emotional vocalisations (e.g., cry, laugh, scream, etc.) which include also some degree of voluntary control and which can vary according to the type of the social partner (but also be expressed without any audience). Indeed, in contrast to human language and gestural communication, there is still no evidence that the production of those species-specific vocalizations could be dissociated from an emotional state and from their determined contexts of use. It is thus not excluded that this flexibility related to species-specific vocalisations does not involve intentional properties and does not require a specific goal in mind in order to motivate the production of vocalisations.

However, recent exceptional reports may provide some new clues to the evolution of the vocal system as well as an evolutionary scenario of the emergence of language (speech). Hopkins and colleagues (2007) have thus described the use of two atypical novel “learned” sounds produced by several chimpanzees among the captive groups from the Yerkes Primate Research Center: Some chimpanzees are not only able to produce non-voiced “raspberries” or “kiss” sounds (involving only the lips with the air of the mouth) but also “extended grunts” which clearly engage the vocal tract and laryngeal sound production mechanisms. The authors showed that the production of these atypical sounds and vocalizations is often produced with pointing gestures and is used exclusively in the presence of both a human and an out-of-reach food in order to beg for food, while typical species-specific ‘food’ calls were more frequent in the presence of food alone (Hopkins et al., 2007). Such atypical productions were interpreted as signals used intentionally to capture the attention of the human. Great apes have been shown to use those acoustic signals - vocal and lips sounds, cage banging or clapping gestures - especially when the recipient is not attentive, whereas visual pointing gestures are preferentially used when the recipient is attentive (e.g., Leavens et al., 2004, 2010; see also in orangutans: Cartmill & Byrne, 2007; for a review of the literature: Hopkins et al., 2011). In other words, some chimpanzees are able to extend to the vocal system this special feature of

social cognition and their ability to intentional signaling in adjusting relevantly the modality of the signal to the attentional state of the recipient. In addition, given the inter-individual variability among chimpanzees concerning the ability to produce or not those novel sounds, it has been interpreted that, as for human speech but in contrast to species-typical vocalizations, those atypical vocal and lips sounds might be socially learnt. In fact, it has been reported that chimpanzees raised by a biological mother who were able to produce those sounds, were more likely to also be able to do so than chimpanzees raised by humans in a nursery (Tagliatela et al., 2012). Moreover, a recent study showed that, among the chimpanzees that were not able to produce these atypical vocalizations, it was not only possible to explicitly train them to do it using operant conditioning, but also that those subjects would further use these novel vocalizations in a communicative context for getting the attention of a human (Russell et al., 2013). Similarly, it has been reported that the language-trained bonobo Kanzi was able to use four additional vocalizations in a context of human communicative exchanges only (Hopkins & Savage-Rumbaugh, 1991).

#### 4. Lateralization of the production of signals

In humans, most of the language functions are under the control of the left hemisphere of the brain in both left-handed and right-handed individuals (Knecht et al., 2000) and involve complex neural networks in which Broca's and Wernicke's areas play a key role (Broca, 1865; Wernicke, 1874). Regarding the strong links between language and gesture as well as the relative independence between language lateralization and handedness (Mazoyer et al., 2014; Ocklenburg et al., 2014), it is possible that manual preferences for gestural communication may constitute a better predictor of hemispheric lateralization for language than hand preference for manipulative functions (Bellugi, 1991; Kimura, 1993). Indeed gestures including signing in deaf people, co-speech gestures or pointing gestures in children have been shown to elicit robust predominance of right-handedness (see the review of Cochet & Vauclair, 2010) that could be better related to left-hemispheric dominance for language. Does gestural communication in nonhuman primates involve a left-hemispheric dominance like human speech does? In other words, are nonhuman primates predominantly right-handed for gestural communication?

Interestingly, the only studies available on hand preference for gestures - conducted in large samples of nonhuman primates including captive chimpanzees, bonobos, gorillas and baboons - have all reported population-level right-handedness for different categories of gestures (for reviews: Hopkins et al., 2012; Meguerditchian et al., 2013), a degree that was much more pronounced than the ones reported in non-communicative motor tasks, such as bimanual manipulation (Figure 2). Those gestures included communicative clapping, intraspecific gestures (e.g., hand slap, see Figure 1) and human-directed food begging gestures in both captive chimpanzees and baboons (Hopkins & Cantero, 2003; Hopkins et al., 2005; Meguerditchian et al., 2010; Meguerditchian & Vauclair, 2006, 2009; Meguerditchian et al. 2011b). Similar, though less well documented, evidence of slighter rightward asymmetries in undistinguished types of gestures have also been reported in a sample of captive gorillas (Shafer, 1987), captive bonobos (Hopkins & Vauclair, 2012) and wild chimpanzees (Hobaiter & Byrne, 2013). Such gestural asymmetries in these two species are consistent over time, across different groups and across different categories of gestures whereas, in contrast, no correlation of individual hand preferences was found between bimanual actions and any type of communicative gestures (Meguerditchian & Vauclair, 2006, 2009; Meguerditchian et al., 2010, 2012). To sum up, in contrast to non-communicative actions, different communicative gestures in nonhuman primates showed a similar pattern of hand preferences with each other and may thus share partially the same cerebral system.

Collectively, within an evolutionary perspective, these findings support the hypothesis of a continuity between baboons, gorillas, bonobos, chimpanzees and humans concerning left hemispheric specialization for gestural communication. It might then be hypothesized that such a communicative lateralized system in nonhuman primates constitutes an ideal prerequisite of the cerebral substrate for human language in the common ancestor of these species at least 30–40 million years ago (Meguerditchian & Vauclair, 2008; Meguerditchian et al., 2011a).

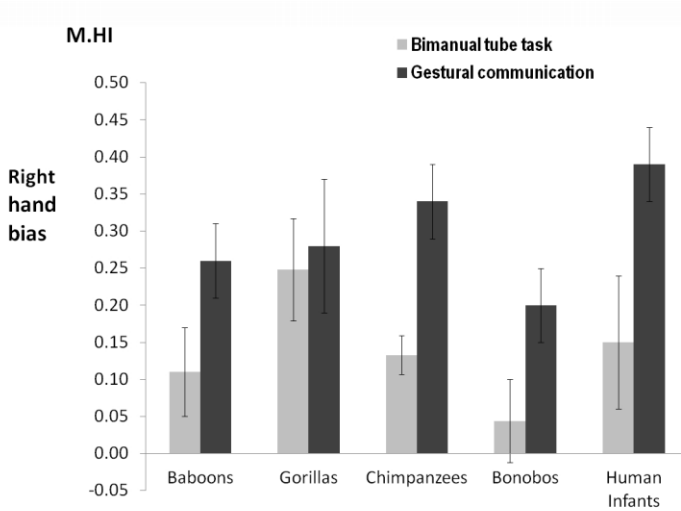


Figure 2. (Adapted from Meguerditchian et al., 2013) Degrees of population-level right-handedness (M.HI) for species-typical communicative gestures in 162 baboons (Meguerditchian et al., 2011b), in 18 gorillas (Shafer, 1987), in 70 chimpanzees (Meguerditchian et al., 2010), in 51 bonobos (Hopkins & Vauclair, 2012) and whole-hand pointing in 37 human infants (Cochet & Vauclair, 2010) compared with the bimanual coordinated task. M.HI scores  $\pm$  S.E. On the basis of total left- and right-hand responses, an individual Handedness Index (HI) was calculated for each subject and varied on a continuum from -1.0 to 1.0. The sign indicates the direction of hand preferences: positive, right-hand preference; negative, left-hand preference. The absolute values reflect the strength of hand preference. The error bar represents the S.E. around the M.HI score. Asterisks indicate that the M.HI score differed significantly from zero. \*  $P < 0.05$ . The positive M.HI values all indicates bias toward right-handedness at group level.

Concerning the vocal system, several studies have investigated hemispheric lateralization for vocal control in nonhuman primates via indirect studies of behavioural oro-facial asymmetries. In humans, the right side of the mouth opens first and wider than the left side, indicating the dominance of the left cerebral hemisphere for language control (Graves et al., 1982; Wolf & Goodale, 1987). Thus, Hook-Costigan and Rogers (1998) evaluated the asymmetries of mouth expressions during call production in marmoset monkeys and reported a larger right side of the mouth (i.e., left hemisphere

bias) in producing social contact calls (positive emotional signals) and a larger left side of the mouth (i.e., right hemisphere bias) in producing fear expressions (negative emotional signals). The authors hypothesized that the asymmetry of call production could be due to the emotional valence of the signal. However, similar studies on species-specific vocalizations in rhesus monkeys (Hauser & Akre, 2001), baboons (Wallez & Vaclair, 2011) and chimpanzees (Fernandez-Carriba et al., 2002) reported oro-facial asymmetries toward the left-side of the mouth, suggesting a right-hemispheric dominance related rather to emotional processing than to a human-like language processing.

Interestingly, in contrast to the species-typical vocal repertoire, the use of the atypical attention-getting sounds in some chimpanzees has been shown to involve an asymmetry toward the right-side of the mouth, i.e. left-hemispheric dominance (Losin et al., 2008; Wallez et al., 2012). Moreover, it turns out that these signals not only share the same communicative intent as the “food beg” gestures in captive chimpanzees but also, when produced simultaneously with these gestures, induce a stronger right-hand preference than when the gesture is produced alone (Hopkins & Cantero, 2003), indicating that the left hemisphere may be more activated when producing both gestures and these atypical vocal and lips sounds simultaneously. Thus, we might support the view that the specific left-lateralized communicative system suggested above for gesture production in baboons and chimpanzees by the reports of specific patterns of right-handedness for gestures, may be involved for both gestures and “learned” attention-getting sounds in chimpanzees. To test such hypotheses, further researches are needed on neural correlates of communicative signalling in nonhuman primates.

## 5. Neural correlates of gestural and vocal communication

As for humans, leftward neuroanatomical asymmetries (i.e., asymmetries of the volume or surface of a given cortical region between the two hemispheres) have been frequently reported in great apes concerning the Inferior Frontal Gyrus (IFG) and the Planum Temporale, (PT), two cerebral regions that are known to overlap keys cerebral regions of language in humans (i.e., Broca and Wernicke areas respectively). This result has been found according to different assessment approaches such as *post-mortem* morphological analyses for the PT

(Gannon et al., 1998), *in vivo* and *post mortem* imaging studies using traditional tracing of specific areas of interest for the IFG (Cantalupo & Hopkins, 2001) and for the PT (Cantalupo, Pilcher, & Hopkins, 2003) and voxel-based morphometry for both PT and IFG (Hopkins et al., 2008).

Interestingly, morphometric analyses of MRI (Magnetic Resonance Imaging) anatomical brain images of chimpanzees have shown that the PT and IFG neuroanatomical asymmetries were driven by the direction of hand preference (i.e., left- or right-handed) for communicative gestures exclusively, including visual food begging gestures (Tagliatalata et al., 2006; Hopkins & Nir, 2010) or auditory clapping gestures (Meguerditchian et al., 2012) but not by hand preferences for non-communicative bimanual coordinated actions (Hopkins & Cantalupo, 2004). Such neuroanatomical correlates strengthen the hypothesis suggested above by the hand preferences' data (1) of a possible neural dissociation between gestural communication and manipulative motor functions and (2) that the manual asymmetries for communicative gestures may reflect a left-lateralized specific communicative system involving, at least in chimpanzees and maybe baboons, cortical homologues of language areas and (3) may thus constitute an ideal precursor of hemispheric specialization for language.

To our knowledge, the only existing functional brain imaging studies conducted on communicative signaling in great apes are consistent with those hypotheses. Thus, the use of positron emission tomography (PET) in 3 captive chimpanzees has revealed that begging food from a human by using either gestures, atypical attention-getting sounds, or both of them simultaneously, activated a homologous region of Broca's area (IFG) predominantly in the left hemisphere (Tagliatalata et al., 2008), a pattern of activation which is enhanced in subjects who used both gestural and vocal signals simultaneously (Tagliatalata et al., 2011). Those unique neurofunctional data provide additional support to the potential existence in chimpanzees of a multimodal intentional system that not only includes gestures but can also integrate, in some individuals, oro-facial and atypical vocal sounds into the same left-lateralized lateralized system.

Further functional brain imaging studies related to the production of signals are needed but these researches are very limited. It is indeed difficult to put an awaked monkey or ape in a very noisy MRI machine and invite it to produce vocalizations or gestures on command in order to measure its brain activation... Thus, most of the functional brain imaging studies in nonhuman

primates investigate the passive *perception* of stimuli including acoustic signals. Those latter studies have used positron emission tomography (PET) and have showed that the passive listening of conspecifics' vocalizations involves cerebral areas within the superior temporal gyrus in rhesus monkeys (e.g., Gil-da-Costa et al., 2006; Petkov et al., 2008; Poremba et al., 2004) and in chimpanzees (Tagliatalata et al., 2009), that might be related to the areas that are involved in the *comprehension* of language in humans. Those findings might be interpreted as evidence of the remarkable capacities of nonhuman primates to understand and categorize the external world (Cheney & Seyfarth, 1990; Seyfarth et al., 2005) – that are at work also in the *comprehension* of human language – without having anything to do with the features of their specific vocal *production* system, and could not be thus particularly regarded as a direct precursor of the human speech *production* system (Meguerditchian & Vaclair, 2008) but rather as the precursor of the representational processes involved in the *comprehension* of language in humans (Gil-da-Costa et al., 2004; Russ et al., 2007; Zuberbühler et al., 1999).

Concerning the production of vocalizations, contrary to human language and the production of gestures in chimpanzees and baboons, vocal control in nonhuman primates seems to imply non-lateralized subcortical structures (limbic or cingulate systems) but not homologous of language areas (Aitken, 1981; reviewed in Jürgens, 2002; Ploog, 1981). In rhesus monkeys, Aitken (1981) reported that ablation of the homologue of Wernicke's and Broca's areas did not affect vocal behaviours, whereas lesions of the anterior cingulate cortex did. In squirrel monkeys, only electrical stimulations performed in limbic and subcortical structures were able to induce vocalizations, but not the stimulation of homologous language areas or other cortical regions (see Ploog, 1981). These findings strengthen the argument for the emotional control of call productions. More recently, in electrophysiological studies conducted in Rhesus macaques, the authors were successful in training some subjects to increase the rate of their calls under the control of a visual stimulus and measured their patterns of neuronal discharge in comparison to *spontaneous* vocalizations. Those studies have confirmed that *spontaneous* production of species-specific vocalizations did not involve neither the motor cortex, or homologue regions of language areas, suggesting an absence of cortical control in vocal communication (and likely an underlying emotional control). In contrast, the *trained* vocal production under visual stimuli elicited neuronal discharges within the premotor cortex (Coudé et al., 2011) and within the

prefrontal cortex in a region known to include Broca's area in humans (Hage & Nieder, 2013), reporting the first evidence of cortical control in vocal productions in monkeys. However, in this latter study, it must be noted that the neurons discharged only when the visual command was presented but not when the trained vocalization was emitted afterward. As the Broca area in humans has been shown to be involved not only for speech production but also in motor planning for non-communicative actions (e.g., Nishitani et al., 2005; Koechlin & Jubault, 2006), these findings would suggest that this homologue region of Broca's area in macaques was involved in motor planning related to the vocal training but not to vocal communication *per se* which rather showed a lack of cortical control. It has not yet been demonstrated that these trained voluntary vocalizations could be used by macaques for intentional communication with social partners as it has been reported for attention-getting "learned" atypical sounds in some chimpanzees (Hopkins et al., 2007).

## 6. Links between hand, mouth and Broca's area

These latter neurobiological studies in macaques as well as the use of attention-getting sounds in chimpanzees question the role of the oro-facial and vocal system in the origin of language and the tight motor connexions between the mouth, the hand and Broca's area. The control of the oro-facial motor system is essential in the production of articulated language. In fact, speech involves complex motor sequences in the mouth (e.g., tongues, lips), which could be considered as "internal fine gestures" of the oro-facial system (see Corballis, 2003) that have been shown to be tightly linked with co-speech manual gestures (Bernardis & Gentilucci, 2006). Interestingly, electrical stimulation of Brodmann's area 44 in rhesus monkeys induced hand and lip movements, suggesting the existence of a neural connexion between the manual and the oro-facial motor system in relation to Broca's area (Petrides et al., 2005). Studies of macaque monkeys have also demonstrated the existence of mirror-neurons in area F5 of the brain, i.e., the homologous to Broca's area (Rizzolati & Arbib, 1998), which have been considered as an ideal substrate for the emergence of imitation, theory of mind and language as well (e.g., Arbib, 2005; Rizzolati & Arbib, 1998). These neurons are activated not only when the monkey is performing a manual action, e.g., cracking nuts, but also during the observation of these actions (Gallese et al., 1996), their passive listening

(Kohler et al., 2002) as well as the observations of the use of tools (Ferrari et al., 2005) and of communicative facial actions (“lip-smacking” and lip protrusion) carried out by the experimenter standing in front of the monkey (Ferrari et al., 2003). It thus seems that, in the monkey brain, area F5 is predisposed to control and recognize visuo-gestural manual actions as well as oro-facial communication, confirming the strong links between the oro-facial motor system, the manual motor system and Broca’s area.

These collective findings support thus the hypothesis that the oro-facial system might constitute a relevant mediator between the gestural communicatory system and speech in the evolution of language. On the assumption that the basic structure of syllables derives from the succession of constrictions and mouth openings involved in chewing, sucking, swallowing and visuo-facial communicative cyclicities, such as lipsmacks, MacNeilage (1998) proposed the “frame-content” theory of speech. According to this theory, the basic components of speech – an oscillatory one (frame) and a segmental one (content) – have their source in cyclic activities of ingestion in our ancestors. Thus, it might be hypothesised that ingestive behaviours- which involved sequentially the hand to the month - were progressively ritualized in oro-facial (lipsmacking) and gestural communication in monkeys (Arbib, 2005). Gentilucci and Corballis (2006) have speculated that facial elements were gradually introduced with vocal elements into the gestural system during language evolution.

## 7. Conclusion

The investigations of the communicative and motor systems in apes and monkeys have revealed some potential behavioural and neural continuities with some features of language in humans. We believe these collective researches support the view of the existence of a left-lateralized intentional communicative system in nonhuman primates that could be referential and multimodal in chimpanzees (including both gestures and atypical vocalizations) and only gestural in Old World monkeys such as baboons. Indeed contrary to chimpanzees, there is no observation that Old World monkeys can use intentional vocalizations for communication or associate vocal and gestural signalling for transmitting the same intents. Nevertheless, regarding the existence of mirror neurons in the prefrontal cortex of macaques, it is possible

that cortical connexions already exist between the hand and the oro-facial motor system in Old World Monkeys and seem to involve the cerebral homologue of Broca's area. As a conclusion, on the bases of those combined findings, we propose the following rudimentary scenario for the origin of the intentional and left-lateralized communicative system involved in human language:

1. Precursors of a left-hemispheric cerebral substrate for language production might have emerged first with the use of communicative gestures in the common ancestor of humans, great apes and Old World monkeys at least 30-40 million years ago. This system includes neuromotor connexions between hand and the oro-facial system.
2. Then, thanks to these pre-existing motor and cortical links between the hand and mouth, this communicative system may have further turned multimodal with the progressive insertion of intentional vocalizations and oro-facial expressions into the gestural system in the course of evolution in the common ancestor of chimpanzees and humans at least 5-7 million years ago.
3. Regarding the selective evolutionary advantages of controlling intentionally vocal communication (which allow intraspecific intentional communication at night, in the dark and at long distances, e.g., Snowdon, 2001), this multimodal system and its vocal component would keep increasing in complexity. With the emergence of *Homo sapiens* at least 170 000 years ago, this system would then turn into a syntactic and generative system to finally become, as it is currently, a complex articulated language associated with co-speech gestures (see Corballis, 2003). In this view, co-speech manual gestures during speech production in humans might constitute the inherited part of our ancestral gestural and bimodal intentional communicatory system (McNeill, 1992).

This theory is consistent with the researches on gestural communication in human adults (e.g., Gentilucci, & Dalla Volta, 2008), in infants and children (e.g., Bernardis et al., 2008) which argue strongly for the view that a single integrated communication system in the left cerebral hemisphere might be in charge of both vocal and gestural communication in human language.

## ACKNOWLEDGEMENTS

This contribution is supported by the French National Research Agency (ANR): ANR-12-PDOC-0014\_1 (LangPrimate) and the ANR grant ANR-08-BLAN-0011 01.

## REFERENCES

- Aitken, P. G. (1981). Cortical control of conditioned and spontaneous vocal behavior in rhesus monkeys. *Brain and Language*, *13*, 171–184.
- Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*, 105-167.
- Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*, 1053-1063.
- Bard, K. A. (1992). Intentional behavior and intentional communication in young free-ranging orangutans. *Child Development*, *62*, 1186–97.
- Bates, E. (1976). *Language and context: The acquisition of pragmatics*. New York: Academic Press.
- Bellugi, U. (1991). The link between hand and brain: Implications from a visual language. In D. Martin (Ed.), *Advances in cognition, education and deafness*, Washington, DC: Gallaudet University Press, 11-35.
- Bernardis, P., Bello, A., Pettenati, P., Stefanini, S. & Gentilucci, M. (2008). Manual action affect vocalization of infants. *Experimental Brain Research*, *184*, 599-603.
- Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, *44*, 178-190.
- Bourjade, M., Meguerditchian, A., Maille, A., Gaunet, F., & Vauclair, J. (2014). Olive baboons (*Papio anubis*) adjust their visual and auditory intentional gestures to the visual attention of others. *Animal Behaviour*, *87*, 121-128.
- Broca, P. (1865). Sur le siège de la faculté du langage articulé. *Bulletin de la Société d'Anthropologie de Paris*, *6*, 377–393.

- Butterworth, G., & Morissette, P. (1996). Onset of pointing and the acquisition of language in infancy. *Journal of Reproductive and Infant Psychology*, *14*, 219-231.
- Call, J., & Tomasello, M. (1994). Production and comprehension of referential pointing by orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology*, *108*, 307-317.
- Call, J. & Tomasello, M. (2007). *The gestural communication of apes and monkeys*. New York: LEA.
- Cantalupo, C., & Hopkins, W. D. (2001). Asymmetrical Broca's area in great apes. *Nature*, *414*, 505.
- Cantalupo, C., Pilcher, D., & Hopkins, W. D. (2003). Are planum temporale and sylvian fissure asymmetries directly related? A MRI study in great apes. *Neuropsychologia*, *41*, 1975-1981.
- Cartmill, E. A., & Byrne, R. W. (2007). Orangutans modify their gestural signalling according to their audience's comprehension. *Current Biology*, *17*, 1345-1348.
- Cartmill, E. A., & Byrne, R. W. (2010) Semantics of orangutan gesture: determining structure and meaning through form and use. *Animal Cognition*, *13*, 793-804.
- Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world*. Chicago: University of Chicago Press.
- Cochet, H., & Vauclair, J. (2010). Pointing gestures produced by toddlers from 15 to 30 months: Different functions, hand shapes and laterality patterns. *Infant Behavior and Development*, *33*, 432-442.
- Corballis, M. C. (2002). *From hand to mouth. The origins of language*. Princeton, NJ: Princeton University Press.
- Corballis, M. C. (2003). From Mouth to Hand: Gesture, Speech, and the Evolution of Right-Handedness. *Behavioral and Brain Sciences*, *26*, 199-208.
- Corina, D. P., San Jose-Robertson, L., Guillemin, A., High, J., & Braun, A. R. (2003). Language lateralization in a bimanual language. *Journal of Cognitive Neuroscience*, *15*, 718-730.
- Coudé, G., Ferrari, P. F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., Monti, F., Rozzi, S., & Fogassi, L. (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS One*, *6*, e26822.

- Emmorey, K. (2002). *Language, cognition, and the brain: Insights from sign language research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Emmorey, K., Mehta, S., & Grabowski, T. J. (2007). The neural correlates of sign versus word production. *Neuroimage*, *36*, 202–208.
- Fernandez-Carriba, S., Locches, A., Morcillo, A., & Hopkins, W. D. (2002b). Functional asymmetry of emotions in primates: New findings in chimpanzees. *Brain Research Bulletin*, *57*, 561–564.
- Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, *17*, 1703–1714.
- Ferrari, P. F., Rozzi, S., & Fogassi, L. (2005). Mirror neurons responding to the observation of actions made with tools in the monkey ventral premotor cortex. *Journal of Cognitive Neuroscience*, *17*, 212–226.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.
- Gannon, P. J., Holloway, R. L., Broadfield, D. C., & Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's language area homolog. *Science*, *279*, 220–222.
- Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science*, *165*, 664–72.
- Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience and Biobehavioral Reviews*, *30*, 949–960
- Gentilucci, M., & Dalla Volta, R. (2008). Spoken language and arm gestures are controlled by the same motor control system. *Quarterly Journal of Experimental Psychology*, *61*, 944–957.
- Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the gorilla (*Gorilla gorilla*): Repertoire, intentionality, and possible origins. *Animal Cognition*, *12*, 527–546.
- Genty, E., & Zuberbühler, K. (2014). Spatial reference in bonobos gestures. *Current Biology*, *24*, 1601–1605.
- Ghazanfar, A. A., & Hauser, M. D. (1999). The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends in Cognitive Sciences*, *3*, 377–384.

- Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M. D., Carson, R. E., Herscovitch, P., & Martin, A. (2004). Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual and affective processing systems in the macaque. *Proceedings of the National Academy of Sciences USA*, *101*, 17516-17521.
- Gil-da-Costa, R., Martin, A., Lopes, M. A., Munoz, M., Fritz, J. B., & Braun, A. R. (2006). Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. *Nature Neurosciences*, *9*, 1064-1070.
- Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge, MA: Harvard University Press.
- Graves, R., Goodglass, H., & Landis, T. (1982). Mouth asymmetry during spontaneous speech. *Neuropsychologia*, *20*, 371-381.
- Hage, S. R., & Nieder, A. (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nature Communications*, *4*, 2409.
- Hauser, M. D., & Akre, K. (2001). Asymmetries in the timing of facial and vocal expressions in rhesus monkeys: Implications for hemispheric specialization. *Animal Behaviour*, *61*, 391-408.
- Hobaiter, C., & Byrne, R. W. (2013). Laterality in the gestural communication of wild chimpanzees. *Annals of the New York Academy of Sciences*, *1288*, 9-16.
- Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzees gestures. *Current Biology*, *24*, 1596-1600.
- Hobaiter, C., Leavens, D., & Byrne, R. W. (2014). Deictic gesturing in wild chimpanzees (*Pan troglodytes*)? Some possible cases. *Journal of Comparative Psychology*, *128*, 82-87.
- Hook-Costigan, M. A., & Rogers, L. J. (1998). Lateralized use of the mouth in production of vocalizations by marmosets. *Neuropsychologia*, *36*, 1265-1273.
- Hopkins, W. D., & Cantalupo, C. (2004). Handedness in chimpanzees is associated with asymmetries in the primary motor cortex but not with homologous language areas. *Behavioural Neuroscience*, *118*, 1176-1183.
- Hopkins, W. D., & Cantero, M. (2003). From hand to mouth in the evolution of language: The influence of vocal behaviour on lateralized hand use in manual gestures by chimpanzees. *Developmental Science*, *6*, 55-61.
- Hopkins, W. D., & Nir, T. (2010). Planum temporale surface area and grey matter asymmetries in chimpanzees (*Pan troglodytes*): The effect of handedness and

- comparison within findings in humans. *Behavioral Brain Research*, 208, 436-443.
- Hopkins, W. D., Pika, S., Liebal, K., Bania, A., Meguerditchian, A., Gardner, M., & Schapiro, S. J. (2012). Handedness for manual gestures in great apes: A meta-analysis. In S. Pika, & K. Liebal (Eds.), *Developments in Primate Gesture Research*, Amsterdam: John Benjamins, 93-112.
- Hopkins, W. D., Russell, J. L., Freeman, H., Buehler, N., Reynolds, E., & Schapiro, S. J. (2005). The distribution and development of handedness for manual gestures in captive chimpanzees (*Pan troglodytes*). *Psychological Science*, 6, 487-493.
- Hopkins, W. D., & Savage-Rumbaugh, E. S. (1991). Vocal communication as a function of differential rearing experiences in *Pan paniscus*: A preliminary report. *International Journal of Primatology*, 12, 559-583.
- Hopkins, W. D., Tagliatalata, J. P., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour*, 73, 281-286.
- Hopkins, W. D., Tagliatalata, J. P., & Leavens, D. A. (2011). Do Chimpanzees have voluntary control of their facial expressions and vocalizations? In A. Vilain, C. Abry, J. C. Schwartz, & J. Vauclair (Eds.), *Primate communication and human language: Vocalisation, gestures, imitation and deixis in humans and non-humans*, Amsterdam: John Benjamins Publishing Company, vol. 1, 71-88.
- Hopkins, W. D., Tagliatalata, J. P., Meguerditchian, A., Nir, T., Schenker, N. M., & Sherwood, C. C. (2008). Gray matter asymmetries in chimpanzees as revealed by voxel-based morphology. *Neuroimage*, 42, 491-497.
- Hopkins, W. D., & Vauclair, J. (2012). Evolution of behavioral and brain asymmetries in primates. In M. Tallerman & K. Gibson (Eds.), *Handbook of language evolution*, Oxford: Oxford University Press, 184-197.
- Inoue-Nakamura, N., & Matsuzawa, T. (1997). Development of stone tool use by wild chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 111, 159-173.
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16, 367-371.
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Review*, 26, 235-258.

- Kendon, A. (1991). Some considerations for a theory of language origins. *Man*, 26, 199-221.
- Kimura, D. (1993). *Neuromotor mechanisms in human communication*. Oxford: Oxford University Press.
- Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E.B., & Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain*, 123, 2512–2518.
- Koechlin, E., & Jubault, T. (2006). Broca’s area and the hierarchical organization of human behavior. *Neuron*, 50, 963-974.
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, 297, 846–848.
- Leavens, D. A. (2004). Manual deixis in apes and humans. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 5, 387-408.
- Leavens, D. A., & Hopkins, W. D. (1999). The whole-hand point: The structure and function of pointing from a comparative perspective. *Journal of Comparative Psychology*, 113, 417-425.
- Leavens, D. A., Hopkins, W. D., & Bard, K. A. (1996). Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 110, 346-353.
- Leavens, D. A., Hopkins, W. D., & Thomas, R. K. (2004). Referential communication by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 118, 48-57.
- Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (*Pan troglodytes*). *Child Development*, 76, 291–306.
- Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (*Pan troglodytes*). *Animal Cognition*, 13, 33-40.
- Lemasson, A. (2011). What can forest guenons “tell” us about the origin of language? In A. Vilain, C. Abry, J. C. Schwartz, & J. Vauclair (Eds.). *Primate communication and human language: Vocalisation, gestures, imitation and deixis in humans and non-humans*, Amsterdam: John Benjamins Publishing Company, vol. 1, pp. 39 – 70.

- Liebal, K., Pika, S., Call., J., & Tomasello, M. (2004). To move or not to move: how apes alter the attentional state of others. *Interaction Studies*, 5, 199-219.
- Liebal, K., Waller, B., Burrows, A. Slocombe, K. (2013). *Primate Communication. A Multimodal Approach*. Cambridge University Press.
- Losin, E. R., Russell, J. L., Freeman, H., Meguerditchian, A., & Hopkins, W. D. (2008). Left hemisphere specialization for oro-facial movements of learned vocal signals by captive chimpanzees. *PLoS One*, 3, e2529.
- MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. *The Behavioral & Brain Sciences*, 21, 499-546.
- Mazoyer, B., Zago, L., Jobard, G., Crivello, F., Joliot, M., Perchey, G., Mellet, E., Petit, L., Tzourio-Mazoyer, N. (2014). Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness. *PLoS One*, 9, e101165.
- McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Meguerditchian, A., Cochet, H., & Vauclair, J. (2011a). From gesture to language: ontogenetic and phylogenetic perspectives on gestural communication and its cerebral lateralization. In A. Vilain, J.L. Schwartz, C. Abry, & J. Vauclair (Eds.), *Primate Communication and Human Language: Vocalisation, gestures, imitation and deixis in humans and non-humans*, Amsterdam: John Benjamins, 91-118.
- Meguerditchian, A., Gardner, M. J., Schapiro, S. J., Hopkins, W. D. (2012). The sound of one hand clapping: handedness and perisylvian neural correlates of a communicative gesture in chimpanzees. *Proceeding of the Royal Society Biology*, 279, 1959-1966.
- Meguerditchian, A., Molesti, S., & Vauclair, J. (2011b). Right-handedness predominance in 162 baboons for gestural communication: Consistency across time and groups. *Behavioral Neuroscience*, 125(4), 653-660.
- Meguerditchian, A., & Vauclair, J. (2006). Baboons communicate with their right hand. *Behavioral Brain Research*, 171, 170-174.
- Meguerditchian, A., & Vauclair, J. (2008). Vocal and gestural communication in nonhuman primates and the question of the origin of language. In L. S. Roska-Hardy & E. M. Neumann-Held (Eds.), *Learning from animals? Examining the nature of human uniqueness*, London: Psychology Press, 61-85.

- Meguerditchian, A., & Vauclair, J. (2009). Contrast of hand preferences between communicative gestures and non communicative actions in baboons: implications for the origins of hemispheric specialization for language. *Brain and Language, 108*, 167–174.
- Meguerditchian, A., Vauclair, J., & Hopkins, W. D. (2010b). Captive chimpanzees use their right hand to communicate with each other: Implications for the origin of the cerebral substrate for language. *Cortex, 46*, 40–48.
- Meguerditchian, A., Vauclair, J., & Hopkins, W. D. (2013). On the origins of human handedness and language: A comparative review of hand preferences for bimanual coordinated actions and gestural communication in nonhuman primates. *Developmental Psychobiology, 55*, 637–650.
- Miles, H. L. (1990). The cognitive foundations for reference in a signing orangutan. In S.T. Parker, & K.R. Gibson (Eds.), “Language” and intelligence in monkeys and apes: *Comparative developmental perspectives*, Cambridge: Cambridge University Press, 511–539.
- Mitani, J. C., & Nishida, T. (1993). Contexts and social correlates of long-distance calling by male chimpanzees. *Animal Behaviour, 45*, 735–746.
- Nishitani, N., Schürmann, M., Amunts K., & Hari R. (2005). Broca’s Region: From Action to Language. *Physiology, 20*, 60–69.
- Oecklenburg, S., Beste, C., Arning, L., Peterburs, J., & Güntürkün, O. (2014). The ontogenesis of language lateralization and its relation to handedness. *Neuroscience & Biobehavioral Reviews, 43*, 191–198.
- Patterson, F. (1978). Linguistic capabilities of a lowland gorilla. In F. C. C. Peng (Ed.), *Sign Language and language acquisition in man and ape*, Boulder, CO: Westview Press, 161–201.
- Petkov, C. I., Kayser, C., Studel, T., Whittingstall, K., Augath, M., Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neurosciences, 11*, 367–374.
- Petrides, M., Cadoret, G., & Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca’s area. *Nature, 435*, 1235–1238.
- Pika, S. (2008). Gestures of apes and pre-linguistic human children: Similar or different? *First Language, 28*, 116–140.
- Pika, S., Liebal, K., Call, J., & Tomasello, M. (2005). The gestural communication of apes. *Gesture, 5*, 41–56.

- Pika, S., & Mitani, J. C. (2006). Referential gesturing in wild chimpanzees (*Pan troglodytes*). *Current Biology*, *16*, 191-192.
- Ploog, D. (1981). Neurobiology of primate audio-vocal behaviour. *Brain Research Reviews*, *3*, 35-61.
- Pollick, A. S., & de Waal, F. B. M. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences (USA)*, *104*, 8184-8189.
- Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P., & Mishkin, M. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature*, *427*, 448-451.
- Rizzolatti, G., & Arbib, M.A. (1998). Language within our grasp. *Trends in Neuroscience*, *21*, 188-194.
- Roian-Egnor, S. E., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, *27*, 649-654.
- Russ, B. E., Lee, Y. S., & Cohen, Y. E. (2007). Neural and behavioral correlates of auditory categorization. *Hearing Research*, *229*, 204-212.
- Russell, J. L., McIntyre, J. M., Hopkins, W. D., & Tagliatela, J. P. (2013). Vocal learning of a communicative signal in captive chimpanzees, *Pan troglodytes*. *Brain and Language*, *127*, 520-525.
- Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbuehler, K. & Slocombe, K. (2013a). Chimpanzee food calls are directed at specific individuals. *Animal Behaviour*. *86*, 955-965.
- Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbuehler, K. & Slocombe, K. E. (2013b). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS One*. *8*, e76674.
- Seyfarth, R. M., Cheney, D. L., & Bergman, T. J. (2005). Primate social cognition and the origins of language. *Trends in Cognitive Sciences*, *9*, 264-266.
- Shafer, D. D. (1987). *Patterns of Hand Preference among Captive Gorillas*. M.A. thesis, San Francisco State University.
- Snowdon, C. T. (2001). From primate communication to human language. In F.B.M. de Waal (Ed.), *Tree of Origin. What primate behavior can tell us about human social evolution*, Cambridge, MA: Harvard University Press, 193-227.
- Tagliatela, J. P., Cantalupo, C., & Hopkins, W. D. (2006). Gesture handedness predicts asymmetry in the chimpanzee inferior frontal gyrus. *NeuroReport*, *17*, 923-927.

- Tagliatalata, J. P., Reamer, L., Schapiro, S. J., & Hopkins, W. D. (2012). Social learning of a communicative signal in captive chimpanzees. *Biology Letters*, *8*, 498–50
- Tagliatalata, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2008). Communicative signaling activates “Broca’s” homologue in chimpanzees. *Current Biology*, *18*, 343–348.
- Tagliatalata, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2009). Visualizing vocal perception in the chimpanzee brain. *Cerebral Cortex*, *19*, 1151–1157
- Tagliatalata, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PLoS One*, *6*, e18852.
- Terrace, H. S. (1979). *Nim*. New York: Knopf.
- Tomasello, M. (2003). The pragmatics of primate communication. In J. Verschuere, J. O. Östman, J. Blommaert, & C. Bulcaen (Eds.) *Handbook of pragmatics*, Amsterdam: John Benjamins, 11–21.
- Tomasello, M., Call, J., Nagell, K., Olguin, R., & Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees: A trans-generational study. *Primates*, *37*, 137–154.
- Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The ontogeny of chimpanzee gestural signals: a comparison across groups and generations. *Evolution of Communication*, *1*, 223–259.
- Tomasello, M., George, B., Kruger, A., Farrar, M., & Evans, A. (1985). The development of gestural communication in young chimpanzees. *Journal of Human Evolution*, *14*, 175–186.
- Tomasello, M., Gust, D., & Frost, G. T. (1989). A longitudinal investigation of gestural communication in young chimpanzees. *Primates*, *30*, 35–50.
- Vauclair, J. (2004). Lateralization of communicative signals in nonhuman primates and the hypothesis of the gestural origin of language. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *5*, 363–384.
- Veà, J. J., & Sabater-Pi, J. (1998). Spontaneous pointing behaviour in the wild pygmy chimpanzee (*Pan paniscus*). *Folia Primatologica*, *69*, 289–290.
- Vilain, A., Abry, C., Schwartz, J. C., & Vauclair J. (Eds.) (2011). *Primate communication and human language: Vocalisation, gestures, imitation and*

- deixis in humans and non-humans*. Amsterdam: John Benjamins Publishing Company.
- Wallez, C., & Vauclair, J. (2011). Right hemisphere dominance for emotion processing in baboons. *Brain and Cognition*, *74*, 164-169.
- Wallez, C., Schaeffer, J., Meguerditchian, A., Vauclair, J., Schapiro, S. J., & Hopkins, W. D. (2012). Contrast of hemispheric lateralization for oro-facial movements between learned attention-getting sounds and species-typical vocalizations in chimpanzees: extension in a second colony. *Brain and Language*, *123*, 75-79.
- Wernicke, K. (1874). *Der aphasische Symptomencomplex on Aphasias. Eine psychologische Studie auf anatomischer Basis*. Breslau: Cohn & Weigert.
- Wich, S. A., & de Vries, H. (2006). Male monkeys remember which group members have given alarm calls. *Proceedings of the Royal Society, Biological Sciences*, *273*, 735-740.
- Willems, R., Özyürek, A. & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, *17*, 2322-2333.
- Wolf, M. E., & Goodale, M. A. (1987). Oral asymmetries during verbal and nonverbal movements of the mouth. *Neuropsychologia*, *25*, 375-396.
- Zimmermann, F., Zemke, F., Call, J., & Gomez, J. (2009). Orangutans (*Pongo pygmaeus*) and bonobos (*Pan paniscus*) point to inform a human about the location of a tool. *Animal Cognition*, *12*, 347-358.
- Zuberbühler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology*, *113*, 33-42.
- Zuberbühler, K., Arnold, K. & Slocombe, K. (2011). Living links to human language. In A. Vilain, C. Abry, J. C. Schwartz, & J. Vauclair (Eds.). *Primate communication and human language: Vocalisation, gestures, imitation and deixis in humans and non-humans*, Amsterdam: John Benjamins Publishing Company, vol. 1, pp. 13-38.
- Zuberbühler, K. (2005). The phylogenetic roots of language. Evidence from primate communication and cognition. *Psychological Science*, *14*, 126-130.

# Residuals of Intelligent Design in Contemporary Theories about Language Nature and Origins

*Antonino Pennisi*<sup>†</sup>  
anpenni@gmail.com

*Alessandra Falzone*<sup>‡</sup>  
amfalzone@unime.it

## ABSTRACT

Some contemporary theories about the origin and the nature of language resort to concepts with no bearing on Darwinian evolutionary hypothesis or evo-devo perspective which are both based on the reconstruction of species morphological structure transformation. These theories, which evoke qualitative leap, cultural evolution, structure/function coevolution as explicative principles for human evolution, in our opinion, result compatible in some points with the most recent Intelligent Design (ID) accounts. Attempting to substantiate itself as a scientific theory, the contemporary ID is ready to give up (or suspend) creationist explanation just to impeding Darwin's fundamental idea according to which it's possible to explain evolution only through a gradual material modification of structures. For comprehending a complex phenomenon as human language – according to ID – it's necessary appealing to a second substance, whatever it is. This idea seems to be at the bottom of all those theories which have rejected monistic structural explanation (modification of physiological structures) for embracing functional, psychological or cultural accounts. We consider these kinds of explanation real unresolved residuals of ID, residuals nested in the heart of the most accredited scientific theories.

Keywords: co-evolution, cultural evolution, dualism, intelligent design, language origins, language nature.

<sup>†</sup> Department of Cognitive Sciences, Educational and Cultural Studies, University of Messina, Italy.

<sup>‡</sup> Department of Cognitive Sciences, Educational and Cultural Studies, University of Messina, Italy.

## 1. The dangers of Intelligent Design 2.0<sup>1</sup>

In a useful book edited a few years ago, Barbara Forrest and Paul Gross (2004) introduced in the cultural debate the idea that not only the public education but also the scientific research could be unconsciously infected by real *Trojan Horses* of creationism. The *wedge of Intelligent Design* introduces a narrow but gradual division between the unrefined biblical-theological tradition and the more sophisticated and dangerous “theistic science” tradition. The research of a scientific acceptance – never obtained yet – has indeed induced the ID supporters to spread a vehicle of philosophical infection, an insidious vehicle because it seems reasonable and moderate: the idea that it is not obligatory to suppose a creator agent in order to state the intrinsic rationality of a biological project, but it is just enough to exclude the possibility that the latter can be only explained through resorting to complete transformations of the matter.

The ID’s current and general criticism – not only to evolutionism but also to biological science – is not that complex phenomena can’t be explained without the participation of a creator God, but rather that they can’t be entirely solved inside a radically monistic theory. In other words, they can’t be exposed to a naturalistic reduction, for example conceiving that «mental functions are by-products of physical activity in the brain, and as such are rigidly predetermined by natural law» (Bowler, 2007, p. 123). Phenomena could be also explained without the clear participation of a designer, but they can never be reduced only to structures’ transformations. To explain any (complex) phenomenon it is necessary to resort to an external substance, to the action of a *second substance*, whatever it is.

Natural selection is the actual target of this second generation creationism. That is «the interplay of undirected natural forces» (Menuge, 2007, p. 32), «a chain of black boxes; as one is opened, another is revealed» (Behe, 1996, p. 6), a theory of naturalistic evolution, which means that it absolutely rules out any miraculous or supernatural intervention at any point. Everything is conclusively presumed to have happened through purely material mechanisms that are in principle accessible to scientific investigation, whether they have yet been discovered or not (Johnson, 2001 p. 61; see Craig-Moreland, 2000). To pursue this polemic aim the new residual creationism would be also inclined to

<sup>1</sup> This paper is the outcome of a collaboration. For the specific concerns of the Italian Academy, we specify that Antonino Pennisi wrote paragraphs 1 and 2 and Alessandra Falzone wrote paragraphs 3 and 4.

pay a heavy price: admitting, for example, that the evolution can reveal itself as a sequence of related species, but never as a casual variation of structures that come in succession across a selective ecological modelling.

But what explanation can a creationistic paradigm without God (“the first thing that has to be done is to get the Bible out of the discussion” – as the vigorous Philip Johnson exhorts in an editorial on a catholic ultra conservative paper i.e. “Touchstone: A Journal of Mere Christianity” August 1999) imagine or admit? If nothing is ascribable to structural transformations generated by natural selection, but neither comes out already equipped by a demiurge’s mind, how the hell we can explain the ID foundational phenomena, or the irreducible and specific complexity in range of a carefully regulated universe? (Behe, 1996; Dembski, 2004; Dembski & Well, 2008; Poole, 2012).

No one of these new gladiators – today almost all are grouped around the *Discovery Institute’s Center for the Renewal of Science and Culture* in Seattle – has got clear ideas about this point. It prevails the currently deconstructive approach of Darwinism, more than the attempt of affirming a new universal vitalism (that could include, if necessary, also the Christian God but, to avoid a prior exclusion from scientific parameters, admits every entity that is different from the structures that it has to explain). From this unexpected pseudo-laic tension, however, it has arisen a debate about all the weak points of the new evolutionistic synthesis that can be considered an alarming attack less against the political-cultural dimension and more against the scientific one, attack that many observers have noticed (Forrest-Gross, 2004; Foster et al., 2008; Pennock, 2001; Shanks, 2004; Young-Edis, 2004; Pievani, 2006).

This attack is particularly insidious especially to some area of the actual evolutionary research since it welds to the difficult suture between social-humanistic culture and the scientific one in the field of privileged contrast of ID theorists: the complexity of human phenomena. Here some positions are arising, configuring themselves as real ID residuals in the inner of the scientific-philosophical community, even in a more advanced naturalism like that of cognitive science.

We will return in detail on these cases in the second part of the paper. At the moment it is interesting to notice how the person who can be probably considered the most qualified among the new ID exponents – the philosopher William Dembski – has taken advantage of difficulties of some areas of the new evolutionistic synthesis in giving completely naturalistic explanations, exactly

in order to hit the theories about language and human intellect that represent the weakest link in Darwinian tradition. His argument comes from the fact that, in this field, the evolutionists look at fossil evidences, at genetic proximities, at the dimension of the brain to substantiate the animal evolution not only of similar physical structures (bones, cranial capacities, DNA sequences, etc.) but also of similar cognitive-behavioural functions (Dembski & Well, 2008). Following synchronization between structures and functions it rests soon, however, the possibility to maintain untouched a strictly naturalistic methodology. Crucial points, in Dembski's opinion, seem to be two, that we can call: 1) the question of the language adaptiveness; 2) the question of the evolutionary pathways mismatch that brings to language.

The first question, well known but sometimes unexplainably refused by philosophers, linguists and biologists that deal with language origins, is that it exists an irreparable friction between the earlier state of the function taken by peripheral and central structures of language, and the ones derived by social-cultural diachronically gradual behaviours, that are inserted in the same structures. It seems that Dembski wants to keep stuck evolutionism in the Darwin's founding idea, still inalienable for biology, according to which all the intelligent processes are oriented to survival and reproduction, when considered as a product of natural selection. The secondary representational function of language, on the contrary, can't be considered an immediate advantage for survival and reproduction:

The evolutionary process, as Darwin conceived it, places no premium on accurately representing reality. The process by which our minds evolved, according to Darwin, places a premium solely on survival and reproduction. Since misrepresentations of reality could facilitate survival and reproduction better than accurate representations, there is no reason to think that our minds are adapted to know the actual state of the world. Indeed, our minds are, on standard evolutionary principles, more likely to operate at the expense of truth, preferring expedience and gratification (Dembski & Well, 2008, p. 16).

From this point of view two decisive points of evolutionary analysis are brought into question: the first one is that the origin of language has to be obligatorily connected to functions directly linked to reproductive advantage; the second one is that, from the point of view of adaptive selection, the functional components acquired with language development might not be

considered, or revealed, as an evolutionary advantage (on the contrary, they even could show themselves as counter-evolutionary features).

This second point raised up by Dembski seems to challenge the central mechanism of evolutionary reconstructions. According to Darwinian dictate, indeed, also the most complex organism derives from “numerous, successive, and slight mutations” of their own morphologic structures. The pure reconstruction of these transformations and the internal laws that rule them, is called by ID supporters a «direct Darwinian pathway» (Dembski & Well, 2008, p. 151).

This is a pathway that, for example, inspires the evo-devo, a contemporary version of evolutionary biology that adds to natural selection one factor: the limited variation possibilities of structures in relation to their own genetic restrictions (onto-genetics and phylo-genetics) and to the regulation of their expression. In evo-devo, however, the place for functions’ autonomy is minimised. The evo-devo motto might be: “tell me how structures change and I’ll tell you why the development of some functions will be impossible”. Analogously a rigorous application of natural selection’s principles doesn’t determine the functional issue of structural transformations before these are completed. We can recall a story, even intense and complicated, of morphologic changes without seeing any new function. Even though there are some hypotheses – we will talk about them soon – today we aren’t able to formulate a universal law that explains when a structural transformation (or even a series of very complex structural transformations) may cause a functional change: furthermore it is possible that we won’t ever know *ante quem*. The lack of a predictive value doesn’t prevent us from formulating advanced *post-hoc* hypothesis on complex specific phenomena. In human language, for example, a “direct Darwinian pathway” can be rebuilt taking into account the original structural constraint story (peripheral and central structures of hearing-vocal system) and the interaction between ecologic and environmental constraint (bio-geographic, for example) and social structure constraints, produced in turn by morphogenetic and cognitive constraints (from human female’s hidden ovulation, to the “sentence” of semantic and syntactic categorization caused by vocal articulation). In short an intersection between different restrictions, but all “inside” a natural perspective.

There is another practicable pathway, though, faster but, precisely, fraught with traps, shortcuts and “trojan horses” of possible intelligent designs not perceptible at first glance. In fact, one of the Dembski’s satisfaction motifs is

the lack of naturalistic patience in which a piece of contemporary evolutionism would fall down. It is obvious, for example, that «to explain irreducible complexity, Darwinists in the end always fall back on indirect Darwinian pathways» (Dembski & Well, 2008, p. 151), or reconstructions in which “not only does a structure evolve but so does its associated function”. In language theories, many examples of this kind exist: new-chomskian discontinuist hypothesis (Fitch, 2010), neo-culturalist (Tomasello, 2014), cognitive ethology (Hauser et al., 2014) or evolutionary psychology hypotheses and, especially, the infinite shapes of co-evolutive hypothesis: «the only way for Darwinism to explain irreducible complexity – concludes Dembski – is by means of an indirect Darwinian pathway in which structures and functions co-evolve» (Dembski & Well, 2008, p. 151).

“Co-evolution” has a primary meaning – derived by classical paradigm – that shows a parallel evolutionary process of different species in a same territory that interact to each other to a point that makes a selective advantage so important as to finish with influencing one another. The generalisation of this meaning, however, has brought to the meaning given by Dembski: «a form of evolution in which biological structures and functions both change so that as structures evolve they acquire new functions» (Dembski & Well, 2008, p. 310). This is exactly the meaning of many hypotheses that would be scientific on language origin and operation, and that gave up the artful separation between natural evolution and cultural evolution in order to explain appearing mismatches between the slow going forward of the structural changing and the dramatically fast appearance of ideas and language usage.

Back to Lamarck? A compromise between human-social sciences and life one? We don't think so. We quite think that these hypotheses drag themselves on unconscious residual of dualistic hypothesis compatible with *Intelligent Design*. On the other hand all the candidate hypotheses linked to the “right way” of the direct Darwinian pathway are just considered aberrant to ID that hasn't the problem of evolutionary biology to respect a totally naturalistic epistemology. The language already arises as a programmed accessory, for ID supporters, of human supremacy upon the rest of the universe and, consequently, it doesn't endeavour to understand how times and kind of action of both structures and functions are synchronized, the value has to be attributed both to a creator God and a ordering function of technologic evolution and/or cultural life. On the contrary, who wants to remain in the naturalistic field never resort to any of these pseudo-creationistic shortcuts.

For example, it always has to account for a reconstruction of the way in which morphologic changes resolve themselves in a function originally linked to primary movements of natural selection; it should avoid to assign explanatory values to so-called “cultural evolution”; it should exclude the recourse to ambiguous hypothesis as discontinuity, “co-evolution”, etc, penalty the encapsulation of ID unsolved residuals inside theories that want to call themselves scientific. In general, who wants to avoid this risk must be disposed to accept what ID stigmatizes as a “*downgrade*” of language and human intelligence (Dembski & Well, 2008, p. 15) but which should be more correctly to define as the ultimate abandonment of any anthropocentric perspective, in terms of a new naturalistic ethics of scientific research.

## 2. Hidden residuals in linguistic theories

It is undeniable the fact that until recently many explicative models, born in the field of cognitive sciences, have unconsciously adopted this dangerous dualism surreptitiously brought by ID in their attempt to explain complexity in human language without using the “right way”: an evolutionary explanation linked only to the progressive variation of morphologic structures. The first and clamorous case is Chomsky’s case in which he suddenly got rid of the question deleting all at once both the evolutionary hypothesis and the theme of morphologic correlates of language, claiming that the species-specificity of these last ones consists in an unusual, *sui generis* cognitive form: «a unique type of intellectual organization that cannot be attributed to peripheral organs or related to general intelligence» (Chomsky, 1966, pp. 4-5).

Chomskyan position, essentially unchanged for the last 50 years and confirmed in his last contribution with a meaningful title, *The mystery of language evolution* (Hauser et al., 2014), seems to be engulfed by tons of criticisms deriving from linguistics, philosophers, psychologists and biologists. However, his thesis seems to have actually success also among his own judges, under the side of language disembodiment – essential to Darwinism. Giving up to the challenge that the complexity can be explained through structural transformations alone, for example, the idea of “double evolution” (the biological and the cultural one) has spread.

Michael Tomasello, who declared to be obsessed for some time by this «ongoing dialectic between evolutionary and cultural-historical processes»

(Tomasello, 2008, p. 10) has been obliged on several occasions to change the “second substance” that could explain language complexity and human primate specificity: shared social learning, capacity of reading other’s intentionality, social cooperation (Tomasello, 1999, 2014). Because he is a compared experimental psychologist his exceptional works have demonstrated, in a Darwinian way, how difficult is not to accept the continuity between human primates and non-human primates: a good example, in this regard, his last book, in which, with a huge intellectual honesty, he admits that, over his previous works, last experiments don’t support anymore that non-human primates don’t perceive the others as intentional agents like themselves. However, moving on philosophical positions, Tomasello ends up with sacrificing Chomsky on the altar of symbolic and social faculty, the abandonment of what should be the principal aim of evolutionary “right way”: explaining the gradual transformation of acoustic-vocal and cerebral structures, and arriving to the thesis of signal-manual origin of language, already supported by other scientists (Paget, 1930; Hewes, 1973; Corballis, 2002).

Essentially, just so as not to pursue the “direct Darwinian pathway”, that should only observe and describe how sapiens talk trying to understand how auditory-vocal structures and cerebral circuits that control their articulations developed, they arrive to imagine approximately hypothetic derivation of vocal signs from manual ones. However, it must be observed that the hypotheses, even if they were true, wouldn’t solve the question given by the constitutive vocalization of actual human language. It should have been also a moment during the hominid evolutionary chain in which gestures become articulated phonic production: in that moment vocal structure of language had also to shift this passage. Then, it had to be ready to use. Here we are at the starting point. Manual gesture, or any other corporeal practice suitable to an information exchange, could be certainly used to strengthen the social cooperation (with many other natural factors), however the specific form of a body technology of language, that is today the relationship between auditory-vocal system and cerebral mapping of its usage, must record a start date, whatever it is. Unless it disclaims that actual humans talk through a developed auditory-vocal canal that is precisely and finely controlled by neuro-cerebral system and that, through this species-specific canal, they develop a cognitive form well defined. That is accepting, with Cartesian Chomsky, that human cognition is a “second substance” independent from the physiologic structure that produces it, or as

an horrified ID representative said, that «mental functions are by-products of physical activity in the brain, and as such are rigidly predetermined by natural law» (Bowler, 2007, p. 123).

The Trojan horse that allows the surreptitious spreading of the dualistic prospective – paradoxical according to whoever pursues naturalistic perspectives – is the substitution of the evolutionary reconstruction of structures with the functions, substitution that, significantly, goes collateral to dualism between natural evolution and cultural evolution. It's not by chance that these positions are especially expressed in evolutionary and compared psychology and in philosophy of mind. It is indicative the case of the thesis about co-evolution – a general trend until few years ago (see Sperber & Wilson, 1986; Origgi & Sperber, 2000). In almost all these theses talking about co-evolution between brain and language always this means co-evolution of cerebral function and language function, no one of these has anything to share with the only real object of naturalistic Darwinian perspective: the evolution of structures. Here two naturalistically insurmountable obstacles arise. The first one is that functions' evolution can become object of biological studies only when structures' transformation has passed through a whole speciation cycle. Otherwise, it's just an alteration of environmental or cultural variations certainly pertinent to sociology, cultural anthropology or any other culturalist subject, but not to natural science. The second one is that the feedback of functions on structures, a merely Lamarckian residual, becomes incompatible with phylogenesis times. There is no feedback effect of language on human brain that is, to date, structurally similar to the first *sapiens'* brain. In the infinite and continuously mutable possibilities of usage of mental processes that the story of social mutations shows us, they eventually change psychological phenomena linked to categorization, to perceptive-inferential processes, to reasoning logics, etc. Of course, as famous sociologists and mediologists claim, the use of Internet or new medial devices will change our “way of thinking”, as writing, printing and any other human cultural activity has done at the breathtaking rhythm of one every hundred years or less, but this never produce any new speciation.

### 3. From the evolution of linguistic function to the evolution of body technology of language

The temptation to adopt a dualistic perspective to understand the language development is much present not only among philosophers and psychologists but also between people who make the natural observation their own job. Language, indeed, has always been a “burden” that many scientists, also some cognitive ones, have tried to sacrifice (see the question of “linguistic negationism” in Pennisi & Falzone, 2010). However, this position has often driven to a deceitful dualism between functional and structural aspects of language.

If we analyze most of the theoretical hypotheses on language origin, indeed, the first starting point seems to be the defence of linguistic function adaptivity: scholars committed to do this job try to fight against one of the harder positions in this field (Chomskyan discontinuity) showing that language functions (or better the functions of each language component, from grammar to pragmatics) have a clear adaptive value (i.e. they are used to improve communicative aspects for *sapiens*) and therefore they have been selected by natural selection. This debate is very vigorous and calls into question compared psychologists and cognitive ethologists as well as illustrious linguists and famous philosophers infected by irresistible temptation of abolishing every kind of “speciality” from linguistic function.

To obtain this aim, however, they paradoxically don't search for the evolutionary structural antecedents of language, but for its functional precursors: they essentially try to behold what kind of core knowledge is collocated at the basis of cognition in general and language in particular, by tracing the presence in species phylogenetically close to *sapiens*.

In this way, they are missing two central aspects of “direct Darwinian pathway”: on one hand they don't consider how linguistic function is realized today (i.e. the use and the function of the sum of central and peripheral vocal structures that allow to produce speech); on the other hand it is not valued the effective usage of the function in natural and species-specific contexts. Essentially most of the scholars who study evolution of language don't ask themselves “how speech has evolved” (or better “how vocal articulatory propriety has evolved”), but “is language an adaptation?” (or better “of what use is language?”).

Answering to the latter question allows on one hand to bypass morphological and ecological (usage in natural context of a certain capacity) differences present among different animal species and on the other hand to reduce language to a simple instrument of communication between members of the same group. Answering to the former question implies, indeed, to underline all the changes that have allowed the continuous production of articulated voice (ability found in *sapiens* only) under a morphological point of view and to consider language as a complex function in which a “set” of anatomical changes allows the rooting of functional “innovations” that differentiate us from non-human primates.

It is maybe the fear of falling into supposed anthropocentric traps that drives theorists of language evolution to a functional analysis of evolution that loses sight of the only evolutionarily valuable aspect: forms’ evolution and functional possibilities that these forms admit. However, it remains an irreducible data, accepted by all the scientists of language evolution, to take into account: if *sapiens* hadn’t had an articulatory morphology and a neuro-cerebral system to finely control it, it wouldn’t have been possible for him to produce articulated speech. Human language is constrained by a highly specialized body structure without which it wouldn’t be practicable.

Now, we start from this assumption, in our opinion the only one that is not getting involved in the quarrel between who thinks that language is a trivial by-product of structure and who believes that it is the result of a “more complex” readjustment of a series of basilar cognitive capacities: *language is a body technology, as a coordinate collection of morphological constraints that allow to learn, produce and understand speech* (Pennisi, 2013).

The concept of morphological constraint, adopted from evo-devo perspective, assumes a decisive role for language evolutionary hypothesis: indeed, in evo-devo perspective the connection between structure and function is not problematic. Morphology, also the complex one, is driven by inner expression laws and mutual influence of genes (Breuker et al., 2006; Klingenberg, 2010; Albertson et al., 2005). Several studies have explored how genetic bases of different morphological traits, that often join the *Bauplan* development of different species, influence the functional meaning of these structures in a decisive way (Dalziel et al., 2009; Barrett & Hoekstra, 2011). Evo-Devo approach tries not to complicate the relationship between morphology and function: there is more interest to understand how a structure has evolved rather than to explain how it is possible that a function is adaptive.

Functions show themselves because they are essentially constrained by the possibilities offered by structures (Irschick et al., in press).

From this assumption, we support here the idea that, at least for language evolution, to explain the merely transformation of functional aspects not only doesn't describe the real nature of language but can even result epistemologically misleading (it means that it can give residual dualistic explanations and not a naturalistic ones). This is the reason why we propose a theory of language evolution that aims to explain not what language has allowed us to do, but how structures that permit language have evolved.

#### 4. A denied natural history: the evolution of voice's forms

One of the risk that occurs when one tries to value adaptivity of the single components of a cognitive function is to offer adaptationist explanations: evolutionary psychology has often been accused of adaptationism, that is the tendency to consider every single aspect of a specific behaviour or a certain cognitive capacity as adapt to survival (see Cosmides & Tooby, 2013). From this point of view, the whole human cognition is constituted by a series of functional adaptation stretched to obtain more reproductive success. Then every actual cognitive capacity should be the result of selection – happened in an “ancestral environment” during Pleistocene – of more advantageous behaviours, as if natural selection worked to optimize our cognition. The paradox to which adaptationism leads is to explain, under the evolution lens, those behaviours that are counter-adaptive and that are realized by *sapiens* (for example the choice of not having children or the homosexual marriages, see Boyd & Richerson, 2004; Pennisi, 2014) or to use “embarrassing evolutionary explanation” (Pievani, 2014) to account for those behaviours that are not directed by natural selection certainly, as the mainly masculine inclination to political activities or the feminine one to establish sentimental relationship.

On the other hand, it's just the exaggerate use of adaptive explanation that has reduced many methodologically careful scientists to give up, if not to the entire evolutionary explanation (Chomsky, 1972), at least to the identification of causal relationships between structure and function (Hauser et al., 2014) or to the use of the term “adaptation”: even Fitch (2012) declares himself

sceptical against the applicability of the concept of adaptation to language evolution in the name of the “scientific respect” for this term!

Actually, from an evolutionary point of view, there are more plausible explanations that try to integrate various language components in a single evolutionary scenario. For example Miyagawa et al. (2013) have proposed an *Integration Hypothesis* of human language evolution according to which there are two principal components of language, the expressive (prosodic-vocal) one and the lexical (referential) one, both present separately in other species, but present in an integrated way uniquely in *Homo sapiens* (Miyagawa et al. 2014).

According to our hypothesis, this evolutionary scenario is that in which central and peripheral anatomical structures of language have been selected not necessarily for linguistic purposes. Obviously this rooting is determined by microscopical genetic variations which have produced new relationships in DNA (see Carrol, 2005). From those variations, a kind of morphological conformation is arisen and this conformation has offered a modality of vocal articulated production for the first time accessible in a constant way. Natural history of voice is a story of mediation among microscopic, macroscopic, functional, environmental and ecological levels.

In this respect, it really seems contradictory that many hypotheses about language origin that have supported the politically correct cause of evolutionary continuity of linguistic function didn't polarize their own research to that anatomic-functional element which most characterizes human language: the vocal production. For decades vocal production has been considered by linguists as the fortress of human language uniqueness. At the same time, paleo-anthropological studies (see Lieberman, 2007) and comparative studies (Goodall, 1986; Pollick & de Waal, 2007) have converged on a data: non-human primates don't have that autonomy on vocal tract as to allow them to use voice for communication, except for referential signals, linked to the context (Seyfarth & Cheney, 1980) or for concrete needs (food and reproduction, Hauser, 1996). This convergence has advanced the false idea that primate world has to be considered just a mute world (see Falk, 2009). Maybe because of this prejudice, scientists of language origin didn't search for the evolutionary precedents of linguistic function in primate vocalization. Because of this supposed impossibility, for non-human primates, to produce similar-linguistic vocal sounds, and the epistemological necessity of tracing a “precursor system” for linguistic function, many scientists have identified in gestural communication the evolutionary continuity between animal and

human communication. Indeed, a similar connection is not understandable apart from the perspective of dualistic residuals which we have discussed above.

As seen in section 2, even if we admit the truthfulness of this hypothesis, it doesn't solve at all the question of explaining the origin of constitutive vocality of actual human language. To not recognize to primates the capacity of using manual signs in a "symbolic" way seems to be quite questionable (Tomasello, 2014). What we cannot ignore in any way – because it would entirely deny an essential part of natural history that has conduced to *sapiens* and it would establish the "direct Darwinian pathway" for language evolution – is that many non-human primates use voice to communicate not only when they want to express concrete needs, but also to recognize conspecifics, to identify the social role (Geissman, 1993), to connote the membership to a specific group and the dominance (Goodall, 1986), to delimit and to defend their own territory, to recall the partner for sexual purposes and even to communicate with individuals of other species (Gamba et al., 2012). If we look nearer, going back to darwinian evolutionary coral, it is possible to see how lots of species among fishes, amphibians, birds and mammals use sounds, vocalizations and even singings that are functional in their ecological-social contest (Bass & Chagnaud, 2012). This implies that the vocal communication system is used by many animal species to communicate both with conspecifics (mating, territory defence, social roles: see Catchpole & Slater, 1955) and with other animal species (predation/defence, cooperation: see Zahavi, 2003).

Furthermore, different recent studies have underlined that this vast use of vocal communication in animal kingdom depends on the presence of neural patterns of vocal-acoustical signalling that mediate social behaviours among all the vertebrates. The philogenetic presence of a vocal compartment that presents a shared specific organization in the motor nuclei, then, could suggest that there has been a common development that is conserved and evolved among genetic pathways, including combination that brings to the expression of the omeotic gene Hox in mesoderm, and naturally through the development of romboencephalon and spinal cord. Studies have underlined similarity with mammals and primates, highlighting that they have also the same structure designed for the control of sounds production. Bass et al. (2008) suggested, in other words, that the premotor circuits that has given the start to acoustic behavior have been originated in fishes. No a "jump" anymore!

The identification of this functional evolutionary continuity is not yet linked by the exam of morphological possibilities: animal species that use sounds to communicate are characterized by the presence, in the central and peripheral level, of morphology adapt to vocal production and decoding. It is undeniable, indeed, that to succeed to show an adaptive behaviour as vocal communication it is necessary that species present a vocal form adequate to sound production.

Supporting an evolutionary hypothesis of language that considers the structural aspects that allow the production of language as constraints to its exhibition permits to determinate an evolutionary continuity with previous species and to delete any form of “speciality” bounden to the idea of language as abstract producer of symbols. Unlike the hypotheses that assign to gestural communication the role of guarantee of evolutionary continuity, that aren't able to give an unproblematic explanation of the passage from primate gesture to human voice (Corballis, 2002), the analysis of “voice forms”, i.e. of morphological structures that constitute the constraints to vocal production possibility, seems to constitute both the most direct and Darwinian approach and the less anthropocentric one.

If there is a positive aspect that this kind of analysis is, indeed, just to delete every possible residual of anthropocentrism: every animal species produce more or less complex vocal communicative forms, using their own species-specific structures (for example, just remember Indri singings: Gamba et al., 2011; or gibbons complex vocal duet: Geissmann et al., 2006). The approach we propose, indeed, considers central and peripheral morphology that allows language as an anatomical constraint, with components that can be separately observed also in other animal species, not always philogenetically close to *Homo sapiens*. Just think, for example, to the case of supra-laryngeal vocal tract, that for decades was considered the emblem of speciality of human language production. Thanks to comparative studies (Fitch, 2000), we know that vocal tract conformation used by *Homo sapiens* – in which the horizontal portion of oral cavity has similar dimension to the vertical one of larynx cavity – it is reached by many animal species at least starting from crocodiles, thanks to muscular efforts of larynx lowering (Fitch, 2010).

Fitch thinks that the forced lowering of the larynx muscular is present in many animal species because it permits to pretend to have bigger corporal dimensions than the effective ones, producing lower and more defined sounds: this is a characteristic of members with a bigger size, that result more appetizing for reproduction. In *Homo sapiens*, permanent lowered larynx

should be selected because it would offer the possibility to operate this pretence without the continuous muscular effort that the other animal species are obliged to do. The immediate advantage associated to larynx lowering has not, then, directly to do with language, but with a fitness increasing. Language should be installed later (exaptation) and only when the human brain has also become “speech-ready”, i.e. ready to organize voluntary articulator movements. Elsewhere, we have defined “auditory-vocal technology applied to symbolic needs” the group of all the articulator-auditory possibilities in which our individual and social cognition is “condemned” to reach the fulfilment of the own purposes (see Pennisi & Falzone, 2010). This auditory-vocal technology – with defined morphological (peripheral-central) and social (as body forms bind social organization) correlates (Pennisi, 2014) is evolved during a very long time among a huge quantity of mutations that have acted directly or not in the formation of structures, that became suitable to articulated speech at a certain point of their development. The two criteria (direct and indirect) aren’t dissociable and today we can study the effect of these direct mutations through experimentation in cognitive science and reconstructive indirect explanation in evolutionism.

Direct Darwinian pathway goes through the analysis of body morphology: organisms narrate an evolutionary story made by phylogenetic heredity and species-specific changes. The functions that every organism show depend by constraints given by its body shape and by the interaction with the habitat he lives. Human language, as any cognitive function, showed itself only when *sapiens*’ morphology reached a “usable minimum threshold”, a discreet ergonomic target and a system of neural control that make possible compositional segmentation (Wray, 2002) and constant articulation of vocal sounds. In this way, it is possible to explain the presence of a communicative and representative complex function as human language without “intelligent” residuals, without necessarily using an external substance, and without spasmodically researching the adaptivity of every linguistic component to demonstrate evolutionary continuity.

## REFERENCES

- Albertson R.C., Strelman J.T., Kocher T.D., & Yelick P.C. (2005). Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. *PNAS USA*, 102, 16287–16292.
- Barret R.D.H., & Hoekstra H.E. (2011). Molecular spandrels: test of adaptation at the genetic level. *Nature Reviews Genetics*, 12, 767–779.
- Bass A.H., Gilland E.H., & Baker R., (2008). Evolutionary Origins for social vocalization in a vertebrate hindbrain – Spinal Compartment. *Science*, 32, 417–421.
- Bass, A. H, & Chagnaud B. (2012). Shared developmental and evolutionary origins of neural basis of vocal-acoustic and pectoral-gestural signaling. *PNAS, USA*, 09 Suppl 1, 10677–84.
- Behe, M.J. (1996). *Darwin's Black Box. The Biochemical Challenge to Evolution*, New York: Free Press.
- Boyd, R., & Richerson, B. (2004). *Not by genes alone*. Chicago: University of Chicago Press.
- Bowler, P.J. (2007) *Monkey Trials and Gorilla Sermons. Evolution and Christianity from Darwin to Intelligent Design*. Cambridge: Harvard University Press.
- Breuker, C.J., Debat, V., & Klingenberg, C.P. (2006). Functional evo-devo. *Trends in Ecology and Evolution*, 21, 488–492.
- Carroll, S.B. (2005). Evolution at Two Levels: On Genes and Form. *PLoS Biology*, 3, 7, 1159–1166.
- Catchpole, C.K., & Slater, P.J.B. (1995). *Bird Song: Themes and Variations*. Cambridge: Cambridge University Press
- Chomsky, N. (1966). *Cartesian linguistic*. New York: Arper and Row.
- Chomsky, N. (1972). *Language and Mind*, New York: Harcourt, Brace, Jovanovich.
- Corballis, M. (2002). *From Hand to Mouth: The Origins of Language*. Princeton: Princeton University Press.
- Cosmides, L., & Tooby, J. (2013) Evolutionary psychology: New perspectives on cognition and motivation. *Annual Review of Psychology*, 64, 201–229.
- Craig, W.L., Moreland, J.P. (2000). *Naturalism. A Critical Analysis*. London: Routledge.

- Dalziel, A., Rogers, S.M., & Schulte, P.M. (2009). Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. *Molecular Ecology*, 18, 4997–5007.
- Dembski, W.A. (2004). *The Design Revolution. Answering the Toughest Questions About Intelligent Design*, IVP Books.
- Dembski, W.A., & Well, J. (2008) *The Design of Life. Discovering Signs of Intelligence in Biological Systems*. Dallas: Foundation for Thought and Ethics.
- Falk, D. (2009). *Finding our tongues. Mothers, Infants, and the Origins of Language*. New York: Basic Books.
- Fitch, W.T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Science*, 4, 258-267.
- Fitch, W.T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.
- Fitch, W.T. (2012). Evolutionary Developmental Biology and Human Language Evolution: Constraints on Adaptation. *Evolutionary Biology*, 39, 613–637.
- Forrest, B., & Gross, P.R. (2004). *Creationism's Trojan Horse. The Wedge of Intelligent Design*. Oxford: Oxford University Press.
- Foster, J.B., Clark, B., & York, R. (2008). *Critique of Intelligent Design. Materialism versus Creationism from Antiquity to the Present*. New York: Monthly Review Press.
- Gamba, M., Favaro, L., Torti, V., Sorrentino, V., & Giacomini, C. (2011). Vocal tract flexibility and variation in the vocal output in wild Indris. *Bioacoustics*, 20, 251-266.
- Gamba, M., Colombo, C., Giacomini, C. (2012). Acoustic cues to caller identity in lemurs: a case study. *Journal of Ethology*, 30, 1, 191-196.
- Geissmann, T. (1993) *Evolution of Communication in Gibbons (Hylobatidae)*, Doctoral dissertation, University of Zurich, Switzerland.
- Geissmann, T., & Nijman, V. (2006). Calling in wild silvery gibbons (*Hylobates moloch*) in Java (Indonesia): behavior, phylogeny and conservation. *American Journal of Primatology*, 68, 1–19
- Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge: The Belknap Press of Harvard University Press.
- Hauser, M.D. (1996). *The Evolution of Communication*. Cambridge: MIT Press.

- Hauser, M.D., Yang, C., Berwick, R.C., Tattersall, I., Ryan, M.J., Watumull, J., Chomsky, N., & Lewontin, R.C. (2014). The mystery of language evolution. *Frontiers in Psychology*, 5, 401, 1-12.
- Hewes, G.W. (1973). Primate Communication and the Gestural Origin of Language. *Current Anthropology*, 14, 5-24.
- Irschick, D.J., Albertson, R.C., Brennan, P., Podos, J., Johnson, N.A., Patek, S., & Dumont, E. (in press) Evo-devo beyond morphology: from genes to resource use. *Trends in Ecology & Evolution*.
- Johnson, P. (2001). Evolution as Dogma: The Establishment of Naturalism. In R.T. Pennock (Ed.), *Intelligent Design Creationism and Its Critics. Philosophical, Theological, and Scientific Perspectives*. Cambridge: MIT Press, 59-76.
- Klingenberg, C.P. (2010). Evolution and development of shape: integrating quantitative approaches. *Nature Reviews Genetics*, 11, 623-635.
- Lieberman, P. (2007). The Evolution of Human Speech. Its Anatomical and Neural Bases. *Current Anthropology*, 48, 1, 39-66.
- Menuge, A. (2007). Who's Afraid of ID? A Survey of the Intelligent Design Movement. In W.A., Demski & M. Ruse (Eds.), *Debating Design*. Cambridge: Cambridge University Press, 32-51.
- Miyagawa S., Ojima S., Berwick R.C., & Okanoya K. (2014). The integration hypothesis of human language evolution and the nature of contemporary languages, *Frontiers in Psychology*, 5, 564, 1-6
- Miyagawa, S., Berwick, R.C., & Okanoya K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology*, 4, 71.
- Origg G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers, & A. Chamberlain (Eds.), *Evolution and the Human Mind*. Cambridge: Cambridge University Press, 140-169.
- Paget, R.A.S. (1930). *Human Speech*. London: Routledge & Kegan.
- Pennisi, A. (2014). *L'errore di Platone. Biopolitica, linguaggio e diritti civili in tempo di crisi*. Bologna: Il Mulino.
- Pennisi, A. (2013). Origine e funzioni della tecnologia corporea del linguaggio, *Scienza e filosofia*, 10, 11-23.
- Pennisi, A., & Falzone, A., (2010). *Il prezzo del linguaggio. Evoluzione ed estinzione nelle scienze cognitive*. Bologna: Il Mulino.

- Pennock, R.T. (2001). *Intelligent Design Creationism and Its Critics. Philosophical, Theological, and Scientific Perspectives*. Cambridge: The MIT Press.
- Pievani, T. (2006). *Creazione senza Dio*. Torino: Einaudi.
- Pievani, T. (2014). *Evoluti e abbandonati*. Torino: Einaudi.
- Pollick, A.S., & de Waal, F.B.M. (2007). Ape gestures and language evolution. *PNAS, USA*, 104, 8184-8189.
- Poole, V. (2012). *A Comprehensive Introduction to Intelligent Design*, White Word Publications.
- Seyfarth, R.M., Cheney, D.L., & Marler P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, 210, 447, 801-803.
- Shanks, N. (2004). *God, the Devil, and Darwin. A Critique of Intelligent Design Theory*. Oxford: Oxford University Press.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Cambridge: Harvard University Press.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.
- Tomasello, M. (2008). *Evolution of human communication*. Cambridge: MIT Press
- Tomasello, M. (2014). *A natural History of Human Thinking*. Cambridge: Harvard University Press.
- Wray, A. (2002) (Ed.). *The transition to language. Studies in the evolution of language*, Oxford: Oxford University Press.
- Young, M., & Edis, T. (2004). *Why Intelligent Design Fails. A Scientific Critique of the New Creationism*. Piscataway, NJ: Rutgers University Press.
- Zahavi, A. (2003). Indirect selection and individual selection in sociobiology: my personal views on theories of social behavior. *Animal Behaviour*, 54, 859-863.

# Language as a Critical Factor in the Emergence of Human Cognition

*Ian Tattersall*<sup>†</sup>  
iant@amnh.org

## ABSTRACT

Modern human beings are most sharply distinguished from all other organisms alive today by their possession of symbolic reasoning, the cognitive capacity that makes possible the mental construction of alternative versions of the world. Scrutiny of the human fossil and archaeological records reveals that, while brain sizes expanded independently in several hominid lineages over the course of the Pleistocene, this qualitatively distinctive symbolic faculty only emerged in our own. What is more, this acquisition was made remarkably recently: well within the 200,000-year tenure on Earth of our anatomically distinctive species *Homo sapiens*. The earliest anatomical *Homo sapiens* appear to have behaved in much the same manner as their non-symbolic contemporaries, although it is highly likely that they had acquired the neural wiring necessary for symbolic thought in the same event of developmental reorganization that gave *Homo sapiens* its strikingly derived bony morphology. Only subsequent to about 100,000 years ago do archaeological traces suggest that our forebears had actually begun to think symbolically. This implies that the new capacity was released by a purely cultural stimulus (after all, the biology was necessarily already in place). I suggest that cultural trigger involved was the spontaneous invention of language by members of a small population isolate of *Homo sapiens* in Africa, at some time after about 100,000 years ago. Structured, rule-bound language is intricately intertwined with symbolic thought as we experience it today; and it is possible to conceive at least in principle how each could have fed back into the other to create a new dynamic.

Keywords: origins of language, human evolution, symbolic cognition, evolutionary pattern.

<sup>†</sup> Division of Anthropology, American Museum of Natural History, New York, NY, USA.

Language has long been recognized as a core peculiarity of modern human beings. But views have varied wildly as to how and when this unusually structured means of communication, and the unprecedented cognitive style that underpins and permits it, were acquired by ancient members of our lineage. Evolutionary psychologists like to suppose that hominids gradually accumulated linguistic abilities in a feedback process between brain and behavior over the entire Pleistocene epoch, roughly the past two million years during which bona fide members of our genus *Homo* have been around on the planet (Tooby & Cosmides, 2000). They are supported by certain paleoanthropologists, who have favored the appearance of some form of language early in the history of *Homo* (Tobias, 1991; Holloway et al., 2004), as well as by neurobiologists such as Terrence Deacon (1997). On the archaeological side, McBrearty & Brooks (2000) have favored a gradual emergence of “modern” behaviors over most of the length of the Middle and Late sections of the Pleistocene.

At the other end of the spectrum of possibilities, some practitioners of linguistics have advocated a recent “big bang” appearance of language. Derek Bickerton (1995, p. 69), for example, has declared that “true language, via the emergence of syntax, was a catastrophic event, occurring within the first few generations of *Homo sapiens sapiens*.” Something similar has been concluded by archaeologists such as Henshilwood et al. (2002) and Marean et al. (2007), who perceive a rather abrupt appearance of “modern” behavior patterns in the Late Pleistocene. From the standpoint of genetics, Tim Crow (2002) has strongly argued that a single recent mutational event gave rise to the anatomically-distinctive species *Homo sapiens*, along with all of its cognitive peculiarities including language, theory of mind, and mental maladjustment. Crow’s conclusion, it might be noted, is strikingly similar to the one reached on archaeological grounds by Richard Klein (see Klein & Edgar, 2002).

All of the participants in this debate over human origins concur that, at some remove, modern human beings are descended from a non-linguistic, non-symbolic precursor that did not, as we do today, remake the world in its head via the reshuffling in the mind of discrete mental symbols. What divides them is the issue of process. Essentially, one group believes that modern cognition was acquired steadily and gradually over the last two million years during which human brain:body size ratios have markedly increased (in multiple independent lineages). This gradual process of cerebral and cognitive increase would more or less exclusively have involved natural selection at the

individual level. Smarter individuals behaved more appropriately to circumstances, and thus reproduced more successfully, generation by generation, than less intelligent ones did. Under this view, we human beings have been honed by Nature to be intelligent and linguistic, and thus more generally to be the very unusual kind of creatures we are. Nature, in other words, has contrived to engineer us to a specific human condition.

The alternative view has profoundly different implications for the kind of organism we happen to be. For, if the unique human way of mentally processing and communicating information was in fact a recent and relatively short-term attainment in our lineage, then it is highly likely that a significant element of chance was involved in this acquisition. And if such was the case, then our behavioral repertoire was clearly not fine-tuned by selection to suit the now-vanished “environment of evolutionary adaptation” that is blamed by evolutionary psychologists and their intellectual fellow-travelers for our many inappropriate behaviors.

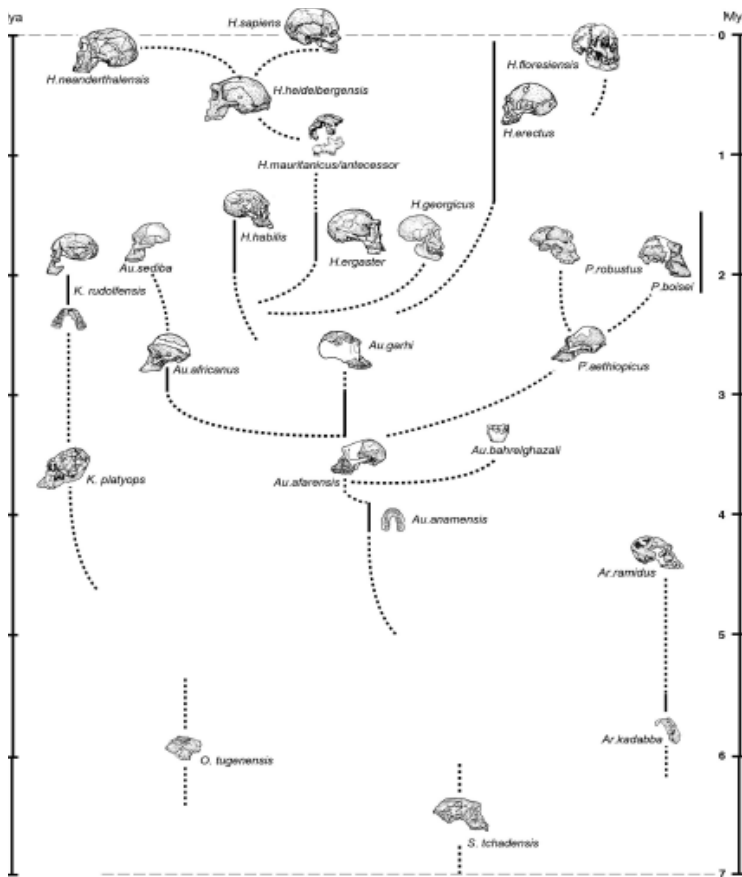
One of the reasons for these radical divergences in perspective on human evolution is that language and other cognitive attributes do not imprint themselves directly upon the preserved record. Archaeological sites from the Pleistocene consist for the most part of temporary hunter-gatherer camps that contain the debris of occupation: mainly animal bones representing food remains, and the stone artifacts used in butchery and other activities. None of these materials can be used as direct proxies for the cognitive states of the hominids responsible for them, so that in most cases we are obliged to make very indirect inferences about all behaviors except for explicitly technological ones. And I would argue that, while many Paleolithic stone-working techniques are certainly witnesses to very sophisticated cognitive states, few if any can be used alone to infer a specifically modern human symbolic cognitive style, let alone the possession by their makers of language (Tattersall, 2008, 2012). Indeed, apart from multi-stage technological sequences requiring extensive planning and recursive inputs, it seems reasonable to insist that only explicitly symbolic artifacts can confidently be used as proxies for symbolic thought processes on the part of their makers. Of course, recognizing a symbolic artifact is not always easy, or unequivocal. Can a roughly-altered lump of stone that looks to a modern person like something vaguely familiar be considered symbolic? Was a pierced gastropod shell necessarily part of a symbolic ornamentation system? Does grinding ochre in itself imply symbolic bodily decoration? There will always be tricky cases like these. But most

archaeologists can probably agree that the engraving of a geometrical design on a flat plaque pretty certainly implies that the engraver was thinking in a modern way, and that an artist producing realistic images or repetitive elements of a notational system clearly was.

The relationship of such activities as these last ones to language is probably a relatively straightforward one. Or at least, the two are intricately and intimately intertwined. Modern symbolic thinking is basically inconceivable in the absence of modern language, while language itself depends on the combining and recombining of mental symbols in exactly the way that thought does. In other words, we can be pretty confident that an ancient human who made what we can recognize as art or symbolic objects had language as we recognize it today. On the other hand, we are most prudent to conclude that hominids who did not routinely indulge in such behaviors did not; that they were basically doing business in the old way. Of course, this is not to imply that language arose out of nowhere, or that earlier humans who lacked symbolic information processing did not possess high levels of intuitive intelligence, or lacked sophisticated means of gestural and vocal communication. Certainly it did not, and they did. But it may also be fair to claim that, prior to the advent of symbolic mental manipulation, gesture and vocalization, however sophisticated, were exclusively about expression. These behaviors were produced by the pre-existing intuitive cognitive states that they expressed. Structured, rule-bound language, with grammar and syntax as well as a vocabulary, added an extra dimension to the cognitive process: in an intricate feedback, a mode of expression became a portal to symbolic thought.

How could this astonishing transformation, this transition from an ancestral non-symbolic and non-linguistic cognitive condition, to a descendant symbolic and linguistic one, have come about? In order to understand this, we have to return to the fossil and archaeological records to determine the exact patten of change in human evolution. Was this pattern one of gradual change, as evolutionary psychologists and their allies claim? Or was change abrupt, as many others are beginning to conclude? In deciding between these two alternatives, the first task is to examine the topography of the hominid genealogical tree. And purely on the basis of the tree shown in Figure 1, we have to reject the notion that *Homo sapiens* is the most recent phase of a steady, long-term refinement of a central lineage. For no central lineage is discernible in this tree. Instead, the history of the hominids has quite evidently been one of vigorous experimentation, as one new species after another was thrown out on to the environmental stage to compete for ecological space with other species, both closely and distantly related – and ultimately, as

likely as not, to become extinct. This is a rather routine pattern for any successful mammalian group, and it applied from the very earliest days of the hominid family's existence. At times, as the figure shows, there were as many as eight hominid lineages coexisting – and that is just in the known fossil record, which represents only a fraction of past diversity. On an intensely local level, in the period around two million years (myr) ago, we have evidence for at least four different hominid species inhabiting the landscape around Lake Turkana in northern Kenya alone.



**Figure 1.** Highly tentative phylogeny of the hominid family, showing the diversity of species currently known within the group, and indicating some possible lines of descent. Multiple hominid lineages have typically existed in parallel. Artwork by Jennifer Steffey; ©Ian Tattersall.

What is not evident from the figure is that, also from the very beginning, innovation in the anatomical and technological realms were out of phase. New species did not bring new technologies along with them. The very first hominids were a diverse but mostly poorly known assortment of (where known) relatively small-brained seven- to four- million-year old African hominoids that are believed, for one reason for another, to have adopted upright bipedal locomotion when they were on the ground. Better known are the “australopiths,” a quite diverse radiation of hominids from between about 4 and 1.5 myr ago. These were definitely bipedal when moving on the expanding woodlands and open grasslands of Africa; but they were still quite small-statured, retained substantial climbing abilities, and possessed ape-like small brains and protruding faces. It was among one species of australopith that the first use of simple sharp stone flakes to butcher carcasses may have begun as long as 3.4 myr ago (McPherron et al., 2010); and by about 2.5 myr ago (Semaw et al., 1997) the first such implements were being deliberately produced by fracturing stone through intentional percussion.

Crude but effective stone flakes of this kind continued to be made even after the first well-defined species of our own genus, the tall, slender, long-legged *Homo ergaster*, showed up in the fossil record at around 1.9 myr ago (Wood & Collard, 1999). The bifacially-flaked handaxe, the first major refinement in stone tool technology, only effectively appeared at about 1.5 myr ago (see Klein, 2009), long after *Homo ergaster*'s entrance on the scene. And then the handaxe, too, remained unrefined in concept until about 0.3 myr ago, when “prepared-core” tools were introduced: implements shaped carefully on both sides until a single blow would detach an effectively finished tool with a continuous cutting edge all around it (see Klein, 2009). This highly episodic history of innovation, happening over a period marked by epic environmental fluctuation, suggests that it was typical for hominids to greet changing circumstances by adapting old technologies to new uses, rather than by inventing new ones as we do.

Meanwhile, new hominid species were appearing and vanishing from the record. Best-known of these is *Homo heidelbergensis*, a modestly large-brained form that showed up in both Africa and Europe at about 600 kyr ago (Clark et al., 1984; Wagner et al., 2010). *Homo heidelbergensis* is particularly interesting because it is within its tenure across the Old World that many significant technological innovations were made. These included the hafting of stone tools, the building of the first complex shelters, and also the

carving of carefully-shaped wooden throwing spears (de Lumley & Boone, 1982; Thieme, 1997). It was also evidently in the time of *Homo heidelbergensis* that the domestic use of fire became an entrenched part of hominid life.

But significantly, members of this species, like their predecessors, did not make anything that we can confidently interpret as a symbolic object. Smart and resourceful these hominids undoubtedly were; but they were evidently not processing information about the environment using any version of our mental algorithm. The same can even be said of *Homo neanderthalensis*, an endemic European and Western Asian species that flourished between about 200 and 30 kyr ago. Most famous for having brains larger on average than ours today (though not than those of Pleistocene *Homo sapiens*), the Neanderthals are incomparably better-known than any other extinct hominid species. But although they were excellent craftsmen in stone, showed considerable curiosity (picking up fossils and carrying them home, for example), occasionally buried their dead, hunted some fearsomely large animals, and flourished in sometimes very severe climatic conditions, they did not show the spark of creativity that distinguishes *Homo sapiens*. Smart they undoubtedly were, but they were not smart in our way; they were evidently not symbolic thinkers. It is probably as a consequence of this cognitive difference that (along with cognitively archaic hominid species in other regions) the Neanderthals promptly disappeared when their heartland was invaded by cognitively modern *Homo sapiens*.

As for *Homo sapiens* itself, fossils of this very anatomically distinctive species are first found in Ethiopia, in deposits ranging between about 195 kyr and 160 kyr old (White et al., 2003; McDougall et al., 2005). Yet the archaeological contexts in which these fossils occur are notably archaic, and indeed they include the very last handaxes found in Africa (Clark et al., 2003; see also Klein, 2009). The earliest *Homo sapiens* fossils known outside Africa are from the adjacent Levant. Around 100 thousand years old, they occur in archaeological contexts that are once more indistinguishable from those of penecontemporaneous hominids, in this case the Neanderthals (Bar-Yosef, 1988). In other words, there was no replacement of Neanderthals by anatomically modern *Homo sapiens* for as long as the two species were functioning on similar cognitive levels.

In waiting for some considerable time after the appearance of anatomically modern *Homo sapiens* to find any evidence of modern behavior/cognitive patterns, we once more find a discontinuity between the arrival of a new

species and the appearance of a new technology. The lapse was apparently in the order of 100 kyr, since at about 100 kyr pierced marine shell beads and ochre deposits begin to show up in deposits around the Mediterranean and in South Africa (Bouzouggar et al., 2007; Vanhaeren et al., 2006). But even if this kind of proxy for modern cognitive processes seems a little tenuous, we do not have long to wait. At about 77 kyr, Middle Stone Age (MSA) levels at Blombos Cave on the southern African coast have yielded not only ground ochre and shell beads, but a smoothed ochre plaque on which a geometric design was engraved (Henshilwood et al., 2002). This basic design is found repeated at a slightly later South African MSA site (Texier et al., 2010), supporting its identity as a symbolic motif with social meaning. At around 72 kyr, the nearby MSA caves of Pinnacle Point have additionally yielded evidence of a complex multi-stage heating technology for converting the abundant soil derivative known as silcrete from a poor tool-making material into an excellent one (Brown et al., 2009). Especially in concert with the early appearance of microliths (Brown et al., 2010), this is a technology sufficiently elaborate to allow the fairly confident inference that it was produced by symbolic minds.

There are no diagnostic hominid fossils associated with these early expressions of behavioral modernity in South Africa. But that they were the work of members of our species is very firmly implied by early *Homo sapiens* occurrences at other sites of comparable age in Kenya and Sudan and elsewhere in Africa as well as in South Africa itself (Schwartz & Tattersall, 2005). It thus seems quite clear that *Homo sapiens* had begun to acquire its modern cognitive processes somewhere in Africa well before 60 kyr ago, the date by around which molecular evidence from modern populations (Templeton, 2005) suggests that our species definitively emerged from the continent of its birth. In contrast to its earlier failed and apparently nonsymbolic foray into the Levant, once out of Africa the newly symbolic *Homo sapiens* rapidly took over the world, in the process displacing those hominid species that were already resident in Europe and across Asia. By 40 kyr ago (Pike et al., 2012) the inauguration of the dazzling tradition of European cave art had already begun to leave the most eloquent expression possible of a fully-formed modern sensibility. And, possibly most significantly of all, it is in the African MSA that we find the beginnings of that restless appetite for technological and presumably other change that has been so fundamental a component of the human psyche ever since.

The evidence to hand thus indicates that, after some seven million years of hominid evolution, something happened in Africa that revolutionized hominid life and experience. Up to this point, significant change, both anatomical and behavioral, had been both rare in hominid history, and highly sporadic. Over the eons hominid lifestyles, and the beings themselves, had undeniably become more complex. But they had done so in an incremental manner rather than along a smooth trajectory; and for vast periods change had been the exception, rather than the rule. What is more, it seems fair to say that, except at the point of appearance of the physically novel genus *Homo*, successful new entrants on the hominid scene had been improvements on their successors, rather than radically new entities. And then, at around 100 kyr ago something happened to upset this pattern once again, this time in the cognitive realm. Members of one particular hominid species began to process information in an entirely new and unprecedented way. Significantly, this change happened well *within* the tenure on Earth of our distinctive species *Homo sapiens*. In other words, the extraordinary and unprecedented transition from non-symbolic to symbolic cognition was a cultural event rather than a biological one. Intuitively, this might seem rather odd. But in fact it is hardly surprising, because in order for this change to occur, the enabling biology must necessarily already have been in place. Indeed, as we have seen, changes in the biological and cultural domains were typically out of step throughout human evolution.

This formative event in human evolution took place within a context of extreme climatic instability. The hominid populations of Africa would have been regularly battered by the climatic vagaries of the late Pleistocene, a period during which frequent climatic oscillations must at times have caused dramatic decreases in hominid numbers, the desolation of large swaths of territory, and the isolation of small hominid populations in local refugia with relatively kind environmental conditions. Tiny isolates restricted to such refugia would have provided exactly the demographic circumstances in which novelties, both behavioral and genetic, would have been most likely to become “fixed” as population attributes. It was such demographic conditions that presumably accounted for the initial origin of anatomically recognizable *Homo sapiens*, and, later, for the emergence of its behaviorally modern descendant populations.

Anatomical *Homo sapiens* is hugely derived skeletally, and this distinctive species almost certainly appeared in a single event involving radical developmental reorganization that occurred at around 200 kyr ago. It is quite

plausible that this change (large in its effects, but very probably structurally minor at the genomic level) also had ramifications in other organ systems of the body. Affected systems would likely have also included the already large brain; and the changes involved would almost certainly have involved augmented anatomical connections among the cerebral structures physically permitting the associations that underpinning symbolic thinking. It has been energetically debated exactly what those fateful neural changes were (see, e.g., Lieberman, 2007; Coolidge & Wynn, 2009); but while this question has yet to be satisfactorily resolved, the radically new capacity was evidently not co-opted immediately by its possessor. It needed to be “discovered,” just as the ancestral birds learned that they could use their feathers to fly only millions of years after acquiring these unusual features.

So what was the evidently cultural agent of this discovery? By far the most plausible candidate for this role is the invention of structured language (Tattersall, 1998), something we know can spontaneously happen among modern human populations (Kegl et al., 1999). After all, language is the human faculty that is most intricately intertwined with the processes of human thought. Both of these fundamental human possessions are inherently symbolic, and each is literally unimaginable to us today in the absence of the other. What is more, it is quite easy at least in principle to imagine how the addition of structured sequencing, to a complex pre-existing system of vocal communication based on utterances as symbols of particular emotional or intuitive states, could have created a pattern of associations that initially mimicked, and then constituted, what we now experience as thought. Such is not self-evidently the case with such other putative drivers of symbolic cognition as theory of mind (e.g., Dunbar, 1998), which is by its nature internalized. And, as an externalized group property rather than an internalized one, language would have had the advantage over theory of mind in being readily transmitted within a population that happened, exaptively, to be already biologically enabled for it. What is more, if language was invented by hominids that possessed the highly specific anatomical attributes of *Homo sapiens*, notably the retracted face and its sequelae (see Lieberman, 2011), then the peripheral vocal apparatus needed to express this qualitatively new behavioral proclivity was also already in place, having initially been acquired in some other functional context entirely, or perhaps purely as a matter of chance.

In this scenario language, in its modern structured, articulate form, was an essential catalyst in the amazingly recent attainment of fully behaviorally

modern humanity. Originating as a means of communication (even possibly among children at play: Tattersall, 2012), language acted simultaneously as a portal to internal thought, providing a framework within which the brain could form and shuffle symbols to create new and alternative visions of the social and physical worlds. Significantly, although they were based on the fruits of many million years of vertebrate evolution, language and symbolic thought were not merely extrapolations of what had preceded them. A final addition to an ancient structure that had been accreted in a complex manner had produced an organ with an entirely unprecedented and emergent potential. But this potential evidently had to be released by a behavioral innovation: a requirement that neatly explains the disconnect between the earlier acquisition of the underlying potential, and its later behavioral expression.

The product of the stimulus provided by the invention of structured articulate language was, and is, without doubt an altogether remarkable one. Indeed, it is not too much to say that in cognitive terms we modern *Homo sapiens* are qualitatively entirely discontinuous with anything that preceded us. The large-brained Neanderthals were complex and resourceful beings, of that there is no doubt. But, as among all of our other known extinct relatives, among whom change of all kinds seems to have been both rare and highly sporadic, there is scant reason to believe that the Neanderthals ever processed information symbolically. The resulting contrast between their lives and those of the neophilic modern humans who replaced them could scarcely be greater. On the other hand, purely in terms of evolutionary process, nothing unusual had happened at all.

#### ACKNOWLEDGMENTS

I thank Francesco Ferretti and Ines Adornetti for kindly inviting me to contribute these reflections to this fascinating Special Issue of *Humana.Mente*.

#### REFERENCES

- Bar-Yosef, Y. (1998). The chronology of the Middle Paleolithic of the Levant. In T. Akazawa, K. Aoki, O. Bar-Yosef (Eds), *Neandertals and Modern Humans in Western Asia*. New York: Plenum Press, 39-56.

- Bickerton, D. (1995). *Language and Human Behavior*. Seattle, W: University of Washington Press.
- Bouzzougar, A., Barton, N., Vanhaeren, M., d'Errico, F., Collcutt, S., Highham, T., Hodge, E., Parfitt, S., Rhodes, E., Schwenninger, J. L., Stringer, C., Turner, E., Ward, S., Moutmir, A., & Stambouli, A. (2007). 82,000-year-old shell beads from North Africa and implications for the origins of modern human behavior. *Proceedings of the National Academy of Sciences, USA*, *104*, 9964-9969.
- Brown, K. S., Marean, C. W., Heries, A. J. R., Jacobs, Z., Tribolo, C., Braun, D., Roberts, D. L., Meyer, M. C., & Bernatchez, J. (2009). Fire as an engineering tool of early modern humans. *Science*, *325*, 859-862.
- Brown, K. S., Marean, C. W., Jacobs, Z., Schoville, B. J., Oestmo, S., Fisher, E. C., Bernatchez, J., Karkanas, P., & Matthews, T. (2102). An early and enduring advanced technology originating 71,000 years ago in South Africa. *Nature*, doi:10.1038/nature11660.
- Clark, J. D., Asfaw, B., Assefa, C., Harris, J. W. K. Kurashina, H. R., Walter, R. C., White, T. D., & Williams M. A. J. (1984). Palaeoanthropological discoveries in the Middle Awash Valley, Ethiopia. *Nature*, *307*: 423-28.
- Clark, J. D., Beyene, Y., WoldeGabriel, G., Hart, W. K., Renne, P. R., Gilbert, H., Defleur, A., Suwa, G., Katoh, S., Ludwig, K. R., Boisserie, J-R., Asfaw, B., & White, T. D. (2003). Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature*, *423*, 747-752.
- Coolidge, F. L., & T. Wynn. (2009). *The Rise of Homo sapiens: The Evolution of Modern Thinking*. New York: Wiley-Blackwell.
- Crow, T. (2002). Sexual selection, timing and an X-Y homologous gene: Did *Homo sapiens* speciate on the Y-chromosome? In T. Crowe (Ed.), *The Speciation of modern Homo sapiens*. Oxford, UK: Oxford University Press, 197-216.
- de Lumley H., & Boone, Y. (1976). Les structures d'habitat au Paléolithique inférieur. In H. de Lumley (Ed.), *La Préhistoire française vol. 1*. Paris: CNRS, 635-643.
- Deacon, T. W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: W. W. Norton.
- Dunbar, R. I. M. (1998). Theory of mind and the evolution of language. In Hurford, J. R., Studdert-Kennedy, M., Knight, C. (Eds), *Approaches to the Evolution of Language: Social and Cognitive Bases*, eds. Cambridge, UK: Cambridge University Press, 92-110.

- Henshilwood, C., d'Errico, F., Yates, R., Jacobs, Z., Tribolo, C., Duller, G. A.T., Mercier, N., Sealy, J. C., Valladas, H., Watts, I., & Wintle, A. G. (2002). Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science*, *295*, 1278-1280.
- Henshilwood, C., d'Errico, F., Vanhaeren, M., van Niekerk, K., & Jacobs, Z. (2004). Middle Stone Age shell beads from South Africa. *Science*, *304*: 404.
- Holloway, R. L., Broadfield, D. L., & Yuan, M. S. (2004). *The Human Fossil Record, Vol 3. Brain Endocasts: The Paleoneurological Evidence*. Hoboken, NJ: Wiley-Liss.
- Kegl, J., Senghas, A., & Coppola, M. (1999). Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. deGraaf, (Ed.), *Comparative Grammatical Change: The Intersection of Language Acquisition, Creole Genesis and Diachronic Syntax*. Cambridge, MA: MIT Press, 179-237.
- Klein, R. G. (2009). *The Human Career*, 3<sup>rd</sup> ed. Chicago: University of Chicago Press.
- Klein, R. G., & Edgar, B. (2002). *The Dawn of Human Culture*. New York: Wiley.
- Lieberman, D. E. (2011). *The Evolution of the Human Head*. Cambridge, MA: Harvard University Press.
- Lieberman, P. (2007). The evolution of human speech: its anatomical and neural bases. *Current Anthropology*, *48*, 39-66.
- Marean, C. W., Bar-Matthews, M., Bernatchez, J., Fisher, E., Goldberg, P., Herries, A. I. R., Jacobs, Z., Tribolo, C., Duller, G. A. T., Mercier, N., Sealy, J. C., Valladas, H., Watts, I., & Wintle, A. G. (2007). Early use of marine resources and pigment in South Africa during the Middle Pleistocene. *Nature*, *449*, 905-908.
- McBrearty, S., & Brooks, A. S. (2000). The revolution that wasn't: A new interpretation of the origin of modern human behavior. *Journal of Human Evolution* *39*, 453-563.
- McDougall I., Brown, F. H., & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature*, *433*, 733-736.
- McPherron, S., Alemseged, Z., Marean, C. W., Wynne, J. G., Reed, D., Geraads, D., Bobe R., & Béarat, H. A. (2010). Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. *Nature*, *466*: 857-860.
- Pike, A. W. G., Hoffmann, D. L., Garcia-Diez, M., Pettit, P. B., Alcolea, J., De Balbin, R., Gonzalez-Sainz, C., de las Heras, C., Lasheras, J. A., Montes, R., & Zilhao,

- J. (2012). U-series dating of Paleolithic art in 11 caves in Spain. *Science*, *336*, 1409-1413.
- Semaw, S., Renne, P., Harris, J. W. K., Feibel, C. S., R. L. Bernor, R. L., Fesseha, N., & Mowbray, K. (1997). 2.5 million-year-old stone tools from Gona, Ethiopia. *Nature*, *385*, 333-336.
- Schwartz, J. H., & Tattersall, I. (2005). *The Human Fossil Record, Vol. 3: Genera Australopithecus, Paranthropus, Orrorin, and Overview*. New York: Wiley-Liss.
- Tattersall, I. (1998). *The Origin of the Human Capacity*. New York: American Museum of Natural History James Arthur Lectures.
- Tattersall, I. (2004). What happened in the origin of human consciousness? *Anatomical Record (New Anatomist)*, *267B*, 19-26.
- Tattersall, I. (2008). An evolutionary framework for the acquisition of symbolic cognition by *Homo sapiens*. *Comparative Cognition and Behavior Reviews*, *3*, 99-114.
- Tattersall, I. (2012). *Masters of the Planet: The Search for Our Human Origins*. New York: Palgrave Macmillan.
- Tattersall, I., & Schwartz, J. H. (2009). Evolution of the genus *Homo*. *Annual Reviews of Earth and Planetary Science*, *37*, 67-92.
- Templeton, A. R. (2005). Haplotype trees and modern human origins. *Yearbook of Physical Anthropology*, *48*, 33-59.
- Texier P. J., Porraz, G., Parkington, J., Rigaud, J.-P., Poggenpoel, C., Miller, C., Tribolo, C., Cartwright, C., Coudenneau, A., Klein, R., Steele, T., & Verna, C. (2010). A Howiesons Poort tradition of engraving ostrich eggshell containers dated to 60,000 years ago at Diepkloof Rock Shelter, South Africa. *Proceedings of the National Academy of Sciences, USA*, *107*: 6180–6185.
- Thieme, H. (1997). Lower Palaeolithic hunting spears from Germany. *Nature*, *385*, 807-810.
- Tobias, P. V. (1991). *Olduvai Gorge, Vol. 4: The Skulls, Endocasts and Teeth of Homo habilis*. Cambridge: Cambridge University Press.
- Tooby, J., & Cosmides, L. (2000). Toward mapping the evolved functional organization of mind and brain. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press, 1167-1178.
- Vanhaeren, M., D'Errico, F., Stringer, C., James, S. L., Todd, J. A., & Mienis, H. K.

(2006). Middle Paleolithic shell beads in Israel and Algeria. *Science*, *312*, 1785-1788.

Wagner, G. A., Krbetschek, M., Degering, D., Bahain, J.-J., Shao, Q., Falguères, C., Voinchet, P., Dolo, J.-M., Garcia, T., & Rightmire, G. P. (2010). Radiometric dating of the type-site for *Homo heidelbergensis* at Mauer, Germany. *Proceedings of the National Academy of Sciences, USA*, doi/10.1073/pnas.1012722107.

White, T. D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G. D., Suwa, G., & Howell, F. C. (2003). Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature*, *423*, 742-747.

Wood, B., M. Collard. (1999). The human genus. *Science*, *284*, 65-71.



# Human Uniqueness, Bodily Mimesis and the Evolution of Language

*Jordan Zlatev*<sup>†</sup>  
jordan.zlatev@semiotik.lu.se

## ABSTRACT

I argue that an evolutionary adaptation for *bodily mimesis*, the volitional use of the body as a representational devise, is the “small difference” that gave rise to unique and yet pre-linguistic features of humanity such as (over)imitation, pedagogy, intentional communication and the possibility of a cumulative, representational culture. Furthermore, it is this that made the evolution of language possible. In support for the thesis that speech evolved atop bodily mimesis and a transitional multimodal protolanguage, I review evidence for the extensive presence of sound-symbolism in modern languages, for its psychological reality in adults, and for its contribution to language acquisition in children. On a meta-level, the argument is that dividing human cognitive-semiotic evolution into *a sequence of stages* is crucial for resolving classical dichotomies concerning human nature and language, which are both natural and cultural, both continuous with and discontinuous from those of (other) animals.

Keywords: conventionality, cross-modality, iconicity, representation, sound symbolism

## 1. Introduction: what makes us human?

It is commonly assumed that it is language that has made us unique in the animal world (e.g. Christiansen & Kirby, 2003). Indeed, the representational and combinatorial powers of language place it on a level of semiotic complexity that is qualitatively distinct from that of the communicative systems of animals (Zlatev, 2009). However, this does not imply either that language evolved through any kind of sudden jump or “saltation”, or that it was the evolution of

<sup>†</sup> Department of Cognitive Semiotics, Lund University, Sweden.

language *per se* that set our species on a separate trajectory compared to that of all other living creatures on our planet. Rather, the thesis put forth in this paper is that human nature – characterized by a consciousness that is uniquely social and representational – rests on a specific pre-linguistic adaptation: *bodily mimesis* (Donald, 1991, 1998, 2001, 2012; Zlatev, 2007, 2008a, 2008b, 2014a). It is this ancient, nearly two million old adaptation that unleashed unprecedented capacities for representation, skill and imagination, making us who we are today over a prolonged process of bio-cultural evolution. To put it somewhat bluntly: we are fundamentally mimetic creatures, and only as a result of this, and secondarily, linguistic creatures.

In Section 2, I spell out the thesis, showing how “the bodily mimesis hypothesis” (cf. Brown, 2012), in particular as formulated within the conceptual-empirical model known as the *Mimesis Hierarchy*, helps to bridge the apparent gulf between animal and human cognition, including language, within a generally continuous, Darwinian framework. The empirical evidence for the thesis has been presented elsewhere (Donald, 1991; Zlatev, 2008a, 2008b, 2014a). Hence I focus on the theoretical/conceptual questions of how a, relatively speaking, “small difference” – the evolution of enhanced motor control – could make a huge difference in terms of cognitive-semiotic evolution. But I also highlight the need to distinguish between at least two different levels of bodily mimesis, as well as precursors on the one hand, and post-mimetic competencies on the other.

But how could bodily mimesis lead to language, which according to many has “design features” such as arbitrariness and dominance of the spoken-auditory channel (Hockett, 1960)? In Section 3, I argue that the thesis of *the arbitrariness of the linguistic sign* (Saussure, 1916/1983) has been both overextended and misconstrued, in part due to the fundamental ambiguity of the term ‘arbitrary’ latent in one of the canonical texts of linguistics. Once the two major senses: *conventional* (= socially shared) and *unmotivated* (= lacking any iconicity or indexicality between expression and content) are de-coupled, it becomes straightforward to see speech as overwhelmingly conventional, but not unmotivated. Indeed, I will review recent work showing that *sound symbolism*, i.e. non-arbitrariness in the second sense, should be regarded as a universal feature of language. At the same time, given the whole-body, multimodal nature of bodily mimesis (Zlatev, Donald & Sonesson, 2010), it is possible to predict that with increased vocabularies (emerging with increasingly complex cultures), a higher role would have been given to the

communicative channel that is *relatively* less iconic, i.e. on vocalization rather than on bodily movements (Brown, 2012; Zlatev, 2014a).

Hence, the main argument is that bodily mimesis is an essential part of what made us human, while language, important as it for all current human cultures and for our existence as individuals, is essentially “post-mimetic”. The secondary argument is that the “transition” to speech occurred gradually and partially, as shown by multimodality and sound-symbolism. A third, and somewhat implicit in the presentation, argument is that dividing human cognitive-semiotic evolution into *a sequence of stages* is crucial for resolving the dual character of human nature: both natural and cultural, both continuous and discontinuous from that of (other) animals, and thus for understanding the complex bio-cultural evolution of the «half-art and half-instinct of language» (Darwin, 1874, p.194).

## 2. Bodily mimesis as the “missing link” and the Mimesis Hierarchy

Etymologically stemming from the Greek verb *mīmēsthai* (‘to imitate’) the concept of *mimesis* encompasses the imitation of actions but goes considerably beyond it, involving «an embodied, analogue, and primordial mode of representation» (Donald, 2012, p.180). It is Donald’s version of the concept that is most relevant in the context of human cognitive evolution. But it should be acknowledged that the ancient Greeks and especially Aristotle attributed to mimesis a central place in human nature: «... man’s natural propensity, from childhood onwards, to engage in mimetic activity (and this distinguishes man from other creatures, that he is thoroughly mimetic, and through mimesis takes his first steps in understanding» (Aristotle, 1987, p. 34). Prefiguring modern cognitive theories with two millennia, Aristotle also thought that it is through a corresponding act of mental mimesis (re-enactment) that we respond to the observed acting, and are capable of empathizing with the characters represented in the play.

Donald’s original contribution was to describe how a “naturalized” version of the mimesis concept could be sufficient to account for the conjectured lifestyles of «the first universally accepted member of our own genus» (Fitch, 2010, p. 265), *Homo erectus* of 1.8-0.5 MYA. The specific hypothesis is that of «a unified neuro-cognitive adaptation that formed the early foundation of a distinctly human mind-sharing culture» (Donald, 2012,

p.181). The archeological record of such mimetic culture includes fairly complex (Achulean) tool-manufacture, campfires, long-distance migration, endurance running and basically modern human anatomy – and yet no evidence for any of the fossilizing markers of vocal language. Combining this with evidence from neuroscience, psychology, anthropology and primatology (cf. Zlatev, 2014a) it appears likely that an adaptation for enhanced voluntary control of the body (“a mimetic controller”) served as the key to a «cultural style that can still be recognized as typically human» (Donald, 2001, p.261). A strong feature of the hypothesis is its parsimony: while the original adaptive function of bodily mimesis could have been tool production, it would have naturally been “exapted” and extended for much else: «...pantomime, imitation, gesturing, shared attention, ritualized behaviors, and many games. It is also the basis of skill rehearsal, in which a previous act is mimed, over and over, to improve it» (Donald, 2001, p. 240).

We may single out the following social-cognitive domains, in which bodily mimesis has contributed to uniquely human capacities, or at least to uniquely high levels within these domains.

- *Skills*. Many motor skills, especially those under strong genetic control such as species-specific patterns of locomotion, do not require anything corresponding to a mimetic controller. However, the kinds of complex skills necessary for bipolar axe production, or for precision throwing, require systematic rehearsal and the ability to «compare, in imagination, the performed act with the intended one» (Donald, 2012, p.182). Mimesis also implies the ability to «shift attention from the external world, and redirect it to [our] own bodies and actions» (Donald, 1998, p.45), and to align the performed and observed movements. It thus clearly brings about an expansion of the scope and flexibility of consciousness, and serves as an important prerequisite for representational thought.
- *Social learning*. More simple forms of learning with the help of others such as goal emulation, response facilitation and stimulus enhancement, are available to many primate species. However, true imitation in which a novel act is observed, modelled and eventually added to the repertoire, is much more restricted (cf. Tomasello, 1999), and attested only to a degree in chimpanzees, apart from our own species. Importantly, only children have been found to reproduce an observed action with high fidelity even when some of the steps are clearly not functional to achieving the goal, i.e. what is now known as “over-imitation” (Horner & Whiten, 2005).

From the other side, helping the novice by overtly demonstrating, guiding and when necessary correcting is also a universal, apparently human-specific trait (Gergely & Csibra, 2006). What these features jointly demonstrate is that bodily mimesis should not be seen as a purely motor-cognitive adaptation, but as a *social*-cognitive one, co-evolving with aspects of intersubjectivity such as trust and altruism (Zlatev, 2008a; 2014b; Hutto, 2008).

- *Memory and planning.* The ability to (consciously) remember some event experienced in the past is characteristic of episodic memory, allowing mental access to «a particular experience (witnessed, or felt, or thought something) in a particular place at a particular time» (Tulving, 2005, p.15). Such memory is also important for planning and guiding of future actions (known as “foresight” or “prospective memory”). There has been accumulating evidence that at least some episodic memory is not limited to human cognition (cf. Hurford, 2007). In fact, Donald (1991) referred to the minds and cultures of chimpanzees (and analogously to those of the last common ancestor) as “episodic”, acknowledging that it is not only human consciousness that goes beyond the here and now of direct perception.<sup>1</sup> The mimetic controller adds to this the ability to explicitly re-enact a past event through bodily motion, and perhaps more importantly, to go through the steps of a future act. This allows making the act more than a private “visualization” (Thompson, 2007), into a fully-fledged public representation, and thus much more accessible for oneself and for others (cf. Sonesson, 2007).
- *Rites and rituals.* Moving further into the social domain and combining the functions discussed above – skill rehearsal, re-enactment and (over-) imitation – provides the bases for another universal of human cultures: rituals. These involve more or less “formalized”, invariant and stylized bodily performances, loaded with “symbolic” (in the sense of non-utilitarian) meaning, and serving social bonding (Bell, 1997). Donald writes of “reciprocal mimesis” (Donald, 1991, p.6) as the means for establishing such forms of “group mentality”. But we need to be careful

<sup>1</sup> Admittedly, Donald (1991) displays somewhat unusual usage of the term “episodic”. On the one hand, «[b]oth episodic and procedural memory systems seem to be present in a variety of animals» (ibid, p.151), but on some occasions it is said that apes live their lives entirely in the present, in sequences of “concrete episodes”. In other cases, however, he seems to deny this, as when he describes the performances of language-taught apes, «using episodic memory to remember how to use the sign; the best they can manage is a virtual “flashback” of previous performances» (ibid, p.153).

here, since the highly normative and symbolic character of many rituals (such as those of religious character), appear to transcend the borders of mimesis proper, and intermix with the subsequent “mythic” stage (Donald, 1991), characterized by narrative and language.

- *Mime and gesture.* A re-enacted hunting dance is clearly representational, in the sense that expression and content are clearly differentiated for both the performer and audience. In a general sense, it is also communicative. But rituals are generally *performative* rather than informative, and lack the full sense of (Gricean) intentional communication, in which there is both an intention to inform the audience of something new, and a higher-order intention for the audience to understand this (Sperber & Wilson, 1995). Hence, it is not really the case that, as Donald (2012, p.182) proposes: «mime and non-linguistic gesture come for free with skill, because the neuro-cognitive mechanism and computational logic is the same». The cooperative motivations and cognitive capacities for the use of communicative intentions are necessary as well (Zlatev, et al., 2013), and need to be seen as an extension of the motoric aspects of mimesis.

As shown in this summary, the concept of bodily mimesis is both specific and rich in relations and extensions: from the motoric skill to social cognition and human-specific culture. Hence, it has been useful to both constrain it, and provide it with a hierarchical structure, distinguishing simpler from more elaborated forms (e.g. Zlatev 2008b).

Adapting somewhat a definition provided in the context of ontogenetic development (Zlatev, 2013, p.51), an actual or imagined act of cognition or communication is an act of bodily mimesis if: (1) it involves a *cross-modal mapping* between exteroception (e.g. vision) and proprioception (e.g. kinesthesia); (2) it is *under conscious control* and *is perceived by the subject to be similar* to some other action, object or event, (3) the subject *intends* the act to stand for some action, object or event *for an addressee*, and for the addressee to recognize this intention; (4) it is *not fully conventional and normative*, and (5) it does *not divide (semi)compositionally* into meaningful sub-acts that *systematically relate* to other similar acts, as in grammar.

The Mimesis Hierarchy follows from this definition by assuming that the features (1-5) build incrementally atop one another, so that only possessing (1) yields *proto-mimesis*, (1) and (2) together give *dyadic mimesis*, while adding (3) leads to full *triadic mimesis*. With the last two, negative criteria in the

definition follow the two “post-mimetic” stages: (4) *protolanguage*, with signs following criteria for correctness, but with very little systematicity and (5) *language*, with sufficient systematicity to allow the construction of discourse and narratives. Table 1, also adapted from earlier work (Zlatev, 2008b, p.139) shows the five stages of the Mimesis Hierarchy, alongside corresponding social-communicative skills. Reviews of comparative psychological and social neuroscience research (Zlatev, 2008a, 2008b) have revealed abundant evidence for proto-mimesis in non-human primates, and some for dyadic mimesis in non-human apes, and especially chimpanzees. But without extensive human enculturation, triadic mimesis skills are inaccessible, and even the most successful enculturants such as Kanzi do not appear to master them fully, which can explain their inability to acquire more than proto-linguistic skills. The conclusion is thus that it is the lack of bodily mimesis, rather than any “language acquisition device” or such that prevents non-human creatures from evolving both cumulative culture and language.

Stage		Feature	Communicative skills
#5	Language	Semiotic systematicity	- Grammar (= conventional symbolic system) - Narrative
#4	Protolanguage	Conventionality/ normativity	- Two-word utterances - Multimodal constructions
#3	Triadic mimesis	Communicative intention	- Declarative pointing - Iconic gestures - Joint attention
#2	Dyadic mimesis	Volitional re-enactment	- (Over) imitation - Imperative pointing - Shared attention
#1	Proto-mimesis	Mapping exteroception and proprioception	- Emotional contagion - Attentional contagion - Neonatal mirroring - Mutual gaze

Table 1. The five stages of the Mimesis Hierarchy, with incremental features and corresponding cognitive-communicative skills

While there are occasional reports of great-ape performances on the three highest levels of the hierarchy (#3-5), these have been either anecdotic, or else the performances were brought about through extensive human enculturation. A similar conclusion is reached by Vaesen (2012), who documents evidence for «striking differences between humans and great apes» (ibid, p.203) in seven pre-linguistic domains: (a) one motoric: *hand-eye coordination*, (b) three social-cognitive: *imitation*, *teaching* and *social reasoning*; and (c) three general cognitive capacities: *causal reasoning*, *function-based categorization* (e.g. related to tools) and *executive control* (e.g. related to planning). Consequently, Vaesen concludes that «no individual cognitive trait can be singled out as the key trait differentiating humans from other animals» (ibid, p.203). However, considering the functions of bodily mimesis outlined earlier (i.e. skill, planning, imitation, rites, gesture), we may notice considerable overlap with the domains outlined by Vaesen: from the most specific (a), to the intersubjective consequences of bodily mimesis in (b) and the most general ones in (c), the latter developing in tandem with, or as consequences of the extended social-cultural mind (cf. Tomasello, 1999; Zlatev, 2008a). In other words, given the *poly-functional* nature of bodily mimesis, we may agree with Vaesen that the bets need not be placed on a single “cognitive trait”, while at the same time acknowledge the internal coherence among the features that distinguish human and animal cognition, apart from language.

The overlap discussed in the previous paragraph may be in part due to the fact that Vaesen (2012), similarly to Donald (1991), highlights the production and use of *tools* as a crucial evolutionary factor. But the payoffs of bodily mimesis go far beyond this. The five functions emphasized in this section can be argued to provide the essential ingredients of a *uniquely human, yet non-linguistic culture*: based on shared skills, (simple) rites, (public) representations, and non-linguistic communicative signs. An essential property of human cultures (as opposed to other behavioural traditions) is that they are *cumulative* (Richerson & Boyd, 2005). My proposal is that the level of imitation that comes with bodily mimesis is what makes this possible since «only imitation gives rise to cumulative cultural evolution of complex behaviors and artifacts» (Richerson & Boyd, 2005, p.108). But there is another no less important characteristic of human cultures that mimesis potentiates even more clearly: they are the only cultures building extensively on *representational practices and artifacts*. While language clearly builds on this (as argued in Section 3) and makes new important representational formats possible such as

narrative (cf. Table 1), it is a mistake to consider representations, or fully-fledged *signs* (Sonesson, 2007), as dependent on language.

A third universal feature of human cultures is morality, shared norms of “right and wrong”, and socially accepted forms of punishment against transgressors. The normative aspects and functions of rituals discussed earlier are clearly relevant in this respect. However, as already stated, precisely because of the strongly normative, rule-like character of moral systems, there seems to be a strong relationship between them and language (cf. Zlatev, 2014b). Indeed, a co-dependence between the evolution of moral systems and language was proposed by Deacon (1997). Hence, while Donald (1991, p.175) states that «[l]anguage is not necessary for the development of complex social roles and rules, but mimesis is essential», this seems to be only partially true: some “social roles and rules” and communication systems may be mimetic, but with conventionalization follow the defining characteristics of language as a «conventional–normative semiotic system for communication and thought» (cf. Zlatev, 2008b, p.137). Here we encounter the need for a transitional stage between bodily mimesis with its “embodied, analogue” representations on the one hand, and language with hierarchical, narrative and normative structures on the other. Indeed, Stage 4 of the Mimesis Hierarchy fulfills exactly this role. In the next section, we will explore the nature of the “protolanguage” of this stage, and the gradual transition to modern-like spoken language.

### 3. Language “atop” bodily mimesis: multimodality and iconicity

An important implication of the bodily mimesis theory of language origins is that the type of “protolanguage” used by our ancestors of, say, 0.5 MYA would have been neither only vocal (Fitch, 2010) nor only gestural (Corballis, 2002), but *multimodal* (cf. Imai & Kita, 2014). It would have been characterized by combinations of facial-manual-vocal expressions, with high degrees of iconicity (i.e. similarity) and indexicality (i.e. spatio-temporal contiguity) in relation to their meanings. The speech and gestures of children toward the end of the second year of life (e.g. Bates et al., 1979; Andrén, 2010; Lock & Zukow-Goldring, 2010) can be seen as a reflection of such a communicative system, with many «multimodal patterns involving the coordination of specific gestures and vocalizations» (Murillo & Belinchón, 2012, p.31). As well-known, the

close connection between the bodily and the vocal modes of expression is also fully present in adults, to the extent that some psychologists consider speech and gesture to be part of a single cognitive-communicative system (McNeill, 2005). Hence, the evolution of speech would have implied not a “switch”, but a gradual shifting of the communicative load toward the vocal channel – in the case of spoken languages (Collins, 2013; Zlatev 2014b). What is less known is that the close connection between mouth and hands is also displayed in signed languages (Sandler, 2012), though naturally the balance between the communicative channels is reversed in that case.

One has wondered: why is this so? In other words: why are languages predominantly spoken (Kendon, 2009)? A possible explanation goes as follows. Despite what was until recently the received wisdom (see below), both spoken and signed languages display extensive iconicity, a “general property of language” (Perniss, Thompson, & Vigliocco, 2010). Still, despite the strong presence of sound symbolism in speech, the vocal medium has less potential for representing meaning on the basis of similarity than the manual-bodily medium (Fay, Arbib, & Garrod, 2013), and there is general agreement that there is more iconicity in signed than in spoken languages (though so far it has been difficult to provide a quantitative support for this claim). Experimental studies and computational models have suggested a trade-off: with smaller vocabularies it is more efficient with a highly iconic code (i.e. system for expression-meaning mapping), but with larger vocabularies, iconic coding leads to ambiguity, and hence a more “arbitrary” mapping is preferred (Monaghan, Mattock, & Walker, 2012). Together, this interplay between multimodality, iconicity and complexity provides a plausible basis for explaining the gradual and partial transition from bodily mimesis to “symbolic communication” (Brown, 2012; Zlatev 2014a).

It is a *partial* transition, since as we saw above, even spoken languages like English are not only “spoken” – when learned, and when used in communication – and even less so are other fully “natural” and “verbal” languages like American Sign Language (ASL). It is also partial in the sense that the highly iconic and indexical nature of bodily mimesis (and presumably of protolanguage) is also found in the linguistic sign as well, as pointed out above.

The latter claim is only recently beginning to gain acceptance, since it appears to contradict «the fundamental principle of the arbitrary nature of the linguistic sign» (Saussure, 1916/1983, p.130). In fact, the thesis of

“arbitrariness” has been radicalized in structuralist semiotics (and even more so, in post-structuralism) to involve the sign and its relation to reality, as well as relations between signs (i.e. grammar):

[...] the arbitrariness principle can be applied not only to the individual sign, but the whole sign system. ...The arbitrariness of the sign is a radical concept because it establishes the autonomy of language in relation to reality (Chandler, 2007, p.22, 25).

However, there is little license for such interpretations. First, Saussure himself warned against extending the arbitrariness of single signs to that of a whole linguistic system: «A language is not completely arbitrary, for the system has a certain rationality» (Saussure, 1916/1983, p.130). But even concerning individual signs, Saussure warns:

The fundamental principle of the arbitrary nature of the linguistic sign does not prevent us from distinguishing in any language between what is intrinsically arbitrary – this is, unmotivated – from what only relatively arbitrary. Not all signs are absolutely arbitrary. In some cases, there are factors which allow us to recognize different degrees of arbitrariness .... The sign may be motivated to a certain extent. (ibid, p. 130)

This passage is not taken from the section where Saussure acknowledges onomatopoeic expressions such as *bow-wow*, only to discard them as “never organic elements of a linguistic system” and as “far fewer than is generally believed” (cf. Ahlner & Zlatev, 2010, p. 303f), but where he struggles with notions such as “relative arbitrariness”, and “partial motivation”. These are difficult to comprehend alongside with the “fundamental” nature of the arbitrariness principle, and to reconcile with it. The only straightforward conceptual resolution is to conclude that *Cours de Linguistique Générale*, composed as well-known by Saussure’s students on the basis of his 1906-11 Geneva lectures, conflates two related but distinct senses of French term *arbitraire*. The first sense is *unmotivated*, as in the quotation above. Using the familiar Peircian notions (e.g. Sonesson, 2007), this means that the relation or “ground” between expression and content is neither iconic nor indexical: a purely negative definition. The second sense is positive, and suggested in many other parts of the *Cours*, where Saussure talks about the key roles of “tradition”, “society”, “collective habit” and “convention”. In other words,

the linguistic sign is *conventional*, i.e. an object of common knowledge (cf. Itkonen 2008). If now this second sense is taken as central, and as the true “design feature” of language, then it is fully possible to combine it with various forms of non-arbitrariness in the first sense, i.e. signs that are based on intermixtures of iconic, indexical and symbolic grounds (cf. Jakobson, 1965; cf. Ahlner & Zlatev, 2010). Indeed, such intermixture is what unbiased description and psychological investigation of both spoken and signed languages shows (Perniss, Thompson, & Vigliocco, 2010). Based on the latter review article and a few other recent summaries, this section presents evidence for: (a) the extensive presence of *sound symbolism*, i.e. the cover term for any kind of motivated mapping between sounds and meanings across languages; (b) the fact that adults are aware of and make use of it; and (c) that children do so as well, and it is functional for learning a language.

### 3.1 Linguistic typology

Perhaps the primary reason for the downfall of the “Saussurean dogma” (Jakobson, 1965) has been the wealth of descriptive evidence of non-arbitrariness in language. It was possible in the past to downplay the relatively few expressions resembling the sounds made by animals (e.g. *meaw*) or events (e.g. *bang*) in familiar European languages. But the explosion of the typological database during the last few decades has shown that sound symbolism is far from “marginal”:

When we move outside the Indo-European language family, however, we find that iconic mappings are prevalent and are used to express sensory experiences of all kinds. Languages for which a large iconic, or sound-symbolic, lexicon has been reported include virtually all sub-Saharan African languages..., some of the Australian Aboriginal languages..., Japanese, Korean, Southeast Asian languages..., indigenous languages of South America..., and Balto-Finnic languages (Perniss, Thompson, & Vigliocco, 2010, p.3).

As stated at the onset of this quote, a central finding is that the iconicity involved in such expressions (called variously “ideophones”, “expressives”, “mimetics”, or simply “sound-symbolic forms”) is by and large *cross-modal* (cf. Ahlner & Zlatev, 2010), i.e. often in subtle ways, the sound shapes resemble the

experiences from other sensory modalities, movements and mental processes. Based on the typological evidence, Dingemanse (2012) proposes the typological implicational hierarchy shown in (1). This states that languages with unimodal (SOUND) ideophones will be found in all languages, while cross-modal mappings further along the hierarchy imply the presence of all those to the left.

- (1) SOUND < MOVEMENT < VISUAL PATTERNS < OTHER SENSORY PATTERNS < OTHER MENTAL PROCESSES

Apart from the strong tendency for cross-modality (consistent with the first feature of bodily mimesis, cf. Section 2 and Table 1), two related features characterize such conventionalized iconic mappings. First is what is sometimes called *Gestalt iconicity* with more abstract structural properties of the word matching the represented spatio-temporal structure. This can be realized through reduplication, e.g. Japanese *goro* ('heavy object rolling') and *gorogoro* ('heavy object rolling continuously'). Second, phonological contrasts (e.g. voiceless vs. voiced) mark semantic contrasts, e.g. compare the above with *koro* and *korokoro*, 'light object rolling (continuously)'. Together, these properties make sound-symbolic forms *semi-transparent*: what they represent is not as clear as in "primary iconic signs" such as realistic pictures (Figure 1a), but at the same time, the expression-meaning mappings are not as attenuated as in "secondary iconic signs" (Figure 1b), where the similarity can be seen first after having been pointed out (cf. Ahlner & Zlatev 2010, for discussion). This means that sound-symbolism can play a functional role in communication and learning (cf. below).

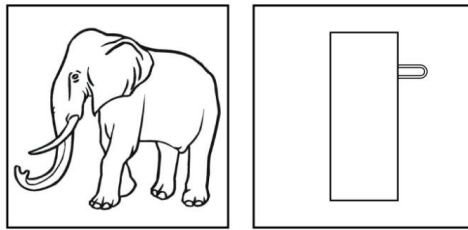


Figure 1. A primary vs. secondary iconic sign: the represented object in the first case is obvious, while in the latter it can be seen only after having been pointed out, e.g. a trombone sticking out from behind a door (from Ahlner & Zlatev, 2010, p. 306)

Furthermore, sound symbolism is not limited to ideophones, but can be found to various degrees in “ordinary” vocabulary for semantic dimensions such as SIZE, SHAPE, SPEED and DISTANCE. For example, demonstrative pronouns such as English *this* and *that* tend to code the more proximal with a front, close vowel like /i/, and the more distal with a back vowel such as /u/. Johansson & Zlatev (2013) show that in a typologically balanced sample of 101 languages (i.e. languages from all over the world, representing language families proportionally to their size) in 56% of the cases the more proximal term had a vowel of higher frequency (e.g. as in English), in 22% there was no difference (e.g. Swedish *denna/detta*), while in (only) 22% the pattern was reversed. Skewed distributions such as these can be explained by assuming that there is a bias or preference for sound-symbolic coding, which of course, can be over-ruled by historical and other contingent factors.

Another example of patterns of sound symbolism that only partially overlap across languages are so-called *phonesthemes* (Abelin, 1999): sounds occurring in words that share some semantic component more often than chance. Some English candidates are /gl-/ suggesting LIGHT (glimmer, glitter, glisten, glow...), /fl-/ suggesting MOVEMENT (*flap, flee, flicker, fling, flip, flow...*) and /-mp/ suggesting PHYSICAL CONTACT (*thump, bump, dump...*).

### 3.2 Psychological reality

But are speakers aware of such (potential) non-arbitrariness, and do they make use of it in learning and communication? Indeed, if all we have are historical “relics” or “vestiges”, not much an argument for the psychological reality of sound symbolism can be made. That is why numerous ingenious experiments have been performed along with the descriptive work. In an early study, Brown et al. (1955) showed that English speakers could match antonym pairs such as *big-small* to corresponding pairs in Chinese, Czech and Hindi significantly better than chance, sometimes by as many as 90% of the participants.

Köller (1929) introduced an experimental paradigm that has been developed and re-applied in many ways: figures of two quite different shapes, typically one sharp and the other roundish are to be matched with various fictive words. In an often quoted paper, Ramachandran and Hubbard (2001) used *bouba* and *kiki*, showing that more than 90% of participants matched *kiki* to the sharp figure and *bouba* to the round figure. Ahlner and Zlatev (2010) varied vowels and consonants independently, showing that both contributed to establishing the (cross-modal) iconic ground: /i/ was

“sharper” than /u/, but so were the voiceless stops /p/, /t/, /k/, compared to the voiced sonorants /m/, /l/, /n/. Westbury (2005) was able to show that such matching affects word recognition in a lexical decision task: words with stops were recognized easier in spiky frames, and sonorants in roundish frames, even when the visual frames were not relevant for the current task. In a context that resembled language learning, Kovic et al. (2010) used an implicit categorization task where participants “made guesses” about which fictive words referred to which depicted animals, received feedback and were later tested on these mappings: some congruent with shape sound symbolism (as described above), and others against it. The findings were that congruent mappings were easier to learn, and faster to confirm in the testing phase.

Beyond the domain of SHAPE, the domain of MOTION has been explored to some degree. Shintel et al. (2006) asked English speakers to describe the movement of dots on a monitor and found that a higher pitch was consistently used for upward movement than for downward movement. Even more interesting, the speed in which the sentences were pronounced corresponded (iconically) to speed of the movement. Furthermore, when listening to these descriptions, another set of participants could correctly judge the speed of the described event, and thus «... both speakers and listeners used speaking rate to convey/comprehend information about an event independent of the semantics of the lexical items» (Perniss, Thompson, & Vigliocco, 2010, p.7).

### 3.3 Language acquisition

Still, a skeptic could complain that such studies with adults could result from cognitive processes that are more typical for poets than “ordinary language users”. But if sound symbolism can be shown to be (first) detectable and (then) functional for language learning, this objection would lose its power. The types of sound-symbolism studied ontogenetically have again concerned above all the domains of SHAPE and MOTION.

Oztruk, Krehm, & Vouloumanos (2012) could show that infants as young as 4-months were sensitive to sound-shape mappings. Using an infant-controlled sequential preferential looking paradigm, the researchers found that infants looked longer when a shape did *not* match the label sound-symbolically than when it did, which can be interpreted as an index of effort or surprise. This is consistent with the findings of somewhat older, 11-month old Japanese infants (Imai et al., 2012). Using an Event Related Potentials (ERP) method, in a number of trials the researchers presented infants with a picture of a shape, followed by a sound-

symbolically congruent or non-congruent form. Both the timing and the topography of the signal were similar to the so-called “N400 effect”, with stronger negative deflection for the non-congruent forms at about 400ms after the stimulus onset, which is usually taken as indicating difficulty in semantic integration.

The ages of the children and the methods of these studies only allow us to infer that the children are performing adequately cross-modal mappings, i.e. perceiving the iconic ground (which is remarkable enough). But do they also use this for learning the signs themselves? In a recent publication, Imai et al. (2013) show a positive role of sound symbolism in infant word learning with 14-month old children. First, the children were repeatedly presented with two word-shape pairs, for half the children sound-symbolically congruent, and for the others not. After this habituation phase, they heard one of two fictive words (*kipi* or *moma*), and saw the two shapes side by side; they looked faster and longer at the congruent. Even if such word-referent associations do not equal lexical meaning, they are a plausible first step to it. And indeed, in another study and experimental paradigm, Maurer et al. (2006) show that 2.5 year old children perform on the classical bouba-kiki task at the same level as adults. Since this task requires understanding the referential function of linguistic signs (cf. Ahlner and Zlatev, 2010), we can see the four studies reviewed here, from that of Ozturk et al. (2012) to that of Maurer et al. (2006), as more or less tracing an ontogenetic version of the Mimesis Hierarchy (see Table 1, Section 2): from the proto-mimesis of cross-modal mappings, to the “post-mimesis” of conventional signs and the onset of a symbolic system.

Let me conclude this section by reviewing a few studies that converge with the thesis developed in this paper, both empirically and theoretically. As noted earlier, Japanese is one of the languages with an extensive inventory of sound-symbolic words, also called “mimetics”. Focusing on the domain of BODILY LOCOMOTION, Imai, Kita, Nagumo, & Okada (2008) asked if Japanese children rely on the sound symbolism of such expressions in acquiring them, or perhaps rather learn them as “arbitrary”. For that purpose they constructed novel mimetic expressions, modeled on existing ones, for example: «*batobato* = a large energetic movement, arms are swinging back and forward outstretched, whereas legs are making large leaping movement; *chokachoka* = walking quickly in very small steps with the arms swinging quickly with bent elbows» (Kantartzis, Imai, & Kita, 2011, p.578).

In a first task, when 25-month old Japanese children were presented with such a novel mimetic expression and two video clips, they were able to select the congruent video at levels above chance. In a second, more difficult task, 3-

year old Japanese children first observed an actor walking in three very different manners, and heard either congruent or non-congruent mimetic verbs, which they had to learn. Then they got to see new video clips, and to point out which one displayed the newly learned expression. Successful *generalization* occurred only with the sound-symbolic expression. Kantartzis, Imai, & Kita (2011) then performed a corresponding study with monolingual 3-year old English-speaking children, using the same “novel mimetics” as before (*batobato, chockachocka* etc.). The remarkable result was that the English-speaking children performed just as the Japanese: “with words that did not sound-symbolically match their referent actions, both Japanese and English 3-year olds failed to generalize the newly taught verb to the identical action performed by another actor. However, when the novel verb matched the action, not only the Japanese 3-year olds, also English-reared 3-year olds... were able to use this cue to generalize the verb to a new event” (Imai & Kita, 2014). Thus, the authors justifiably conclude that they are tapping onto a potentially universal capacity for sound symbolism, and then proceed to link it to an evolutionary scenario.

[...] *the sound symbolism bootstrapping hypothesis*, which states that sound-symbolism can help children single out the referent of a novel word in the complex reality, which in turn allows them to store the semantic representation in such a way that children can correctly generalize the verb to new situations. ... Universal sound-symbolism in modern languages may be the “fossils” of a sound symbolic communication system our ancestors once used (Kantartzis, Imai, & Kita, 2011, p.576, 583).

I find the developmental interpretation of the phenomena fully compelling, especially given the findings earlier reviewed of sensitivity to cross-modal mappings from early infancy. On the evolutionary side, however, I do not find the metaphors of “fossils” or “vestiges” so appropriate, since they suggest non-functional relics, despite the authors’ intentions. Also, there is hardly any reason to propose any evolutionary stage consisting only (or predominantly) of a “sound symbolic communication system”; this faces complementary problems to those of a “purely gesture-first” theory (cf. Fitch, 2010). In fact, elsewhere Imai & Kita (2014) emphasize the close links between sound symbolic forms and gesture, as well as some of the same reasons for gradually attenuating iconicity in the vocal modality as those given in the onset of this

section. Thus, with the risk of being accused of “assimilation”, I would propose that the bodily mimesis hypothesis encompasses the empirical findings and the theoretical proposals of Imai, Kita and colleagues quite nicely.

### Summary and conclusions

In this paper, appearing in a special issue devoted to the relations between human nature and language origins, I have emphasized what may at first have appeared as counter-intuitive: it is not language but rather (above all) bodily mimesis that “makes us special”. In the first half of the paper, I performed what could be seen as a valorization of mimesis: showing how it encompasses – and to some degree explains – unique human features such as tool manufacture, a high degree of intersubjectivity, over-imitation, pedagogy, cumulative culture, and last but not least: the evolution of language itself. Mimesis is a crucial prerequisite for language since, as in Donald’s original evolutionary model, it provides the basis for three of its essential features: (i) conventions (through imitation, dyadic mimesis), (ii) intentional communication (through triadic mimesis), and (iii) for bringing the two together in shared communicative, representations/signs. Donald himself puts this logical dependence in strong terms:

Language is different from mimesis, but it has mimetic roots. It is a collective product and must have evolved as a group adaptation, in the context of mimetic expressive culture. Given the conventional, collective nature of language, it could not have emerged in any other way. (Donald, 2001, p.274)

In the second half of the paper, I have attempted to further strengthen the theory by arguing that mimesis was never just a “prerequisite” to be used and then pushed away like the proverbial ladder, but that the transition to language should be conceived of as partial. The lower layers of bodily mimesis are very much alive and kicking, i.e. functional in everything from everyday communication, performance, empathy and learning – also of language itself. I can summarize the argument of Section 3 by adding the part in italics to the Donald quote from above as follows: “Given the conventional, collective nature of language, *and given its extensive multimodality and non-arbitrariness*, it could not have emerged in any other way.”

Finally, while I started somewhat provocatively by positioning myself with respect to the debate on the nature of relationship between human language and animal signals in the “discontinuity” camp, it should be clear that the theory that I have outlined is one of underlying continuity. It is just that models of language (and human) origins have to accept complex explanations with multiple causal factors and a number of intervening stages. 33 years after Donald (1991), and 15 years after having “discovered” the idea myself, I find that there is a compelling argument that bodily mimesis is one of the major “missing links” in human evolution.

#### ACKNOWLEDGEMENTS

I would like to thank the editors of this special issue, as well as Michael Ranta and Göran Sonesson, for helpful comments to an earlier version of this text.

#### REFERENCES

- Abelin, Å. (1999). *Studies in Sound Symbolism*. Göteborg: Gothenburg Monographs in Linguistics.
- Alhner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign System Studies*, 38(1/4), 298-348.
- Andrén, M. (2010). *Children's Gestures between 18 and 30 Months*. Lund: Media Tryck.
- Aristotle. (1987). *The Poetics of Aristotle*. Translation and Commentary S. Halliwell. Chapel Hill: The University of North Carolina Press.
- Bates, E., Benigni, L., Bretherton, I., Camioni, L., & Volterra, V. (1979). *The Emergence of Symbols: Cognition and Communication in Infancy*. New York: Academic Press.
- Bell, C. (1997). *Ritual: Perspectives and Dimensions*. New York: Oxford University Press.
- Brown, J. E. (2012). *The evolution of symbolic communication: An embodied perspective*. PhD Thesis: University of Edinburgh.
- Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology*, 50, 388-393.
- Chanlder, D. (2007). *Semiotics: The Basics, Second edition*. London: Routledge.

- Christiansen, M. H., & Kirby, S. (2003). Language evolution: The hardest problem in science? In M. H. Christiansen, & S. Kirby (Eds.), *Language Evolution*. Oxford: OUP, 1-15.
- Collins, C. (2013). *Paleopoetics. The Evolution of the Literary Imagination*. New York: Columbia University Press.
- Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.
- Darwin, C. (1874). *The Descent of Man*. London: Murray.
- Deacon, T. (1997). *The Symbolic Species*. New York: Norton.
- Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6(10), 654–672.
- Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Human Culture*. Cambridge, MA: Harvard University Press.
- Donald, M. (1998). Mimesis and the executive suite: Missing links in language evolution. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Biases*. Cambridge: Cambridge University Press, 44-67.
- Donald, M. (2001). *A Mind so Rare: The Evolution of Human Consciousness*. New York: Norton.
- Donald, M. (2012). The mimetic origins of language. In M. Tallerman, & K. R. Gibson (Eds.), *The Oxford Handbook of Language Evolution*. Oxford: OUP, 180-183.
- Fay, N., Arbib, M., & Garrod, D. (2013). How to bootstrap a human communication system. *Cognitive Science*, 37(7), 1356-1367.
- Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.
- Gergely, G., & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of human culture. In N. J. Enfield, & S. C. Levinson (Eds), *Roots of human sociality: Culture, cognition, and human interaction*. Oxford: Berg Publishers, 229- 255.
- Hockett, C. F. (1960). The origin of speech. *Scientific America*, 203, 89–97.
- Horner, V. K., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children. *Animal Cognition*, 8, 164-181.

- Hurford, J. R. (2007). *The Origins of Meaning: Language in the Light of Evolution*. Oxford: OUP.
- Hutto, D. (2008). First communions: Mimetic sharing without theory of mind. In J. Zlatev, T. Racine, C. Sinha, & I. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: Benjamins, 245-276.
- Imai, M., Asano, M., Miyazaki, M., Okada, H., Yeung, H., Kitanjo, K., et al. (2012). Sound symbolism helps infants' word learning. In T. Scott-Phillips, M. Tamariz, E. Cartmill, & J. R. Hurford (Eds.), *The Evolution of Language: Proceedings of Evolang 9*, Singapore: World Scientific, 456-457.
- Imai, M., Kita, & S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B*, 369 (1651).
- Imai, M., Kita, S., Nagumo, S., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109, 54-65.
- Itkonen, E. (2008). The central role of normativity in language and linguistics. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: Benjamins, 279-306.
- Jakobson, R. (1965). Quest for the essence of language. *Diogenes*, 13, 21-38.
- Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning. *Cognitive Science*, 35, 575-586.
- Kendon, A. (2009). Signs for language origins? *The public journal of semiotics*, 2(1), 2-27.
- Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, 114, 19-28.
- Köhler, W. (1929). *Gestalt Psychology*. New York: Liveright.
- Lock, A., & Zukow-Goldring, P. (2010). Preverbal communication. In J. Bremner, & T. Wachs (Eds.), *The Wiley-Blackwell Handbook of Infant Development*. Oxford: Willey-Blackwell, 394-425.
- Maurer, D., Pathman, T., Mondloch, & J. C. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, 9(3), 316-322.
- McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.
- Miyazaki, M., Hidaka, S., Imai, M., Yeung, H. H., Kantartzis, K., Okada, H., et al. (2013). The facilitatory role of sound symbolism in infant word learning. In M. Knauff, M.

- Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, Texas: Cognitive Science Society, 3080-3085.
- Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in word learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *38*, 1152-1164.
- Murillo, E., & Belinchón, M. (2012). Gestural-vocal coordination. *Gesture*, *12*(1), 16-39.
- Oztruk, O., Krehm, M., & Vouloumanos, A. (2012). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month olds. *Journal of Experimental Psychology*, *114*(2), 173-186.
- Perniss, P., Thompson, T., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 1-15.
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies*, *8*, 3-34.
- Richerson, P. J., & Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago press.
- Sandler, W. (2012). Dedicated gestures, and the emergence of sign language. *Gesture*, *12*(3), 265-307.
- Saussure, F. d. (1916/1983). *Cours de Linguistique Générale/ Course in General Linguistics*. Paris/London: Payot/Duckworth.
- Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic in speech communication. *Journal of Memory and Language*, *55*, 165-177.
- Sonesson, G. (2007). From the meaning of embodiment to the embodiment of meaning: A study in phenomenological semiotics. In T. Ziemke, J. Zlatev, & R. Frank (Eds.), *Body, Language and Mind. Vol 1: Embodiment*. Berlin: Mouton de Gruyter, 85-128.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. Malden, MA: Blackwell Publishing.
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology and the Sciences of Mind*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

- Tulving, E. (2005). Episodic memory and autoeoesis: Uniquely human? In H. S. Terrace, & J. Metcalfe (Eds.), *The Missing Link in Cognition: Origins of Self-reflective Consciousness*. Oxford: OUP, 3-56.
- Vaesen, K. (2012). The cognitive bases of human tool use. *Behavioral and Brain Sciences*, *35*, 203–262.
- Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task. *Brain and Language*, *93*, 10–19.
- Zlatev, J. (2007). Language, embodiment and mimesis. In T. Ziemke, J. Zlatev, & R. Frank (Eds.), *Body, Language, Mind. Vol 1: Embodiment*. Berlin: de Gruyter, 297-337.
- Zlatev, J. (2008a). The coevolution of intersubjectivity and bodily mimesis. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: Benjamins, 215-244.
- Zlatev, J. (2008b). From proto-mimesis to language: Evidence from primatology and social neuroscience. *Journal of Physiology – Paris*, *102*, 137-152.
- Zlatev, J. (2009). Levels of meaning, embodiment, and communication. *Cybernetics and Human Knowing*, *16*(3-4), 149-174.
- Zlatev, J. (2013). The mimesis hierarchy of semiotic development: Five stages of intersubjectivity in children. *Public Journal of Semiotics*, *4*(2), 47-70.
- Zlatev, J. (2014a). Bodily mimesis and the transition to speech. In M. Pina, & N. Gontier (Eds.), *The Evolution of Social Communication in Primates*. Berlin: Springer, 165-178.
- Zlatev, J. (2014b). The co-evolution of human intersubjectivity, morality and language. In D. K. Dor, & D. & Lewis (Eds.), *The Social Origins of Language*. Oxford: Oxford University Press, 249-266.
- Zlatev, J., Donald, M., & Sonesson, G. (2010). From body to mouth (and body). In A. Smith, M. Schouwstra, B. deBoer, & K. Smith (Eds.), *The evolution of language*. London: World Scientific, 527-528.
- Zlatev, J., Madsen, E., Lenninger, S., Persson, T., Sayehli, S., Sonesson, G., et al. (2013). Understanding communicative intentions and semiotic vehicles by children and chimpanzees. *Cognitive Development*, *28*, 312-329.



# Making Tools and Planning Discourse: the Role of Executive Functions in the Origin of Language

*Ines Adornetti*<sup>†</sup>  
ines.adornetti@uniroma3.it

## ABSTRACT

In this article we propose that executive functions play a key role in the origin of language. Our proposal is based on the methodological assumption that some of the cognitive systems involved in language functioning are also involved in its phylogenetic origin. In this regard, we demonstrate that a key property of language functioning is discourse coherence. Such property is not dependent on grammatical elements but rather is processed by cognitive systems that are not specific for language, namely the executive functions systems of action planning, control and organization. Data from cognitive archaeology on the making of stone tools show that the processes requested to produce Prehistoric tools imply action organization operations similar to those involved in the processing of coherence. Based on these considerations, we propose that executive functions represent the link between stone tool making and language origins and suggest that they allowed our ancestors to develop forms of proto-discourse governed by coherence.

Keywords: discourse coherence, executive functions, language evolution, neuropsychology, pragmatics, stone tools making.

## 1. Introduction

In this article we propose a hypothesis on the role of executive functions in language origins in reference to a precise methodological assumption: the idea that some of the systems and cognitive skills involved in language functioning are also involved in its phylogenetic origin. In this regard, we show that the proper functioning of language is tied to a key pragmatic property: discourse

<sup>†</sup> Department of Philosophy, Communication and Visual Arts, Roma Tre University, Italy; Department of Human Sciences, University of L'Aquila, Italy.

coherence. Through discussion of a collection of neuropsychological and neurolinguistic studies, we show that such property is essential for effective communication, and its processing does not rely on the grammatical systems that manage the combination of the internal constituents of the sentence, but rather on the executive functions responsible for action control and organization. On the basis of these considerations, we argue that an evolutionarily plausible model of language has to be founded on the primacy of pragmatics on grammar and that a key role in the dawn of language has been played by the development of executive functions underlying discourse coherence. Data from cognitive archeology indicate that the expansion of executive functions is closely linked to the evolution of prehistoric lithic industries and that the processes requested to produce these tools imply action organization operations similar to those involved in the processing of coherence. Based on this, we propose that a key moment in the evolution of language has been the exaptation for communication purposes of executive systems originally tied to action control and that such exaptation allowed our ancestors to develop a communicative system based on forms of proto-discourse governed by a principle of coherence. Building language up from the systems related to the control of actions and which are responsible for the crucial properties of communication—such as discursive coherence—constitutes a way to anchor the advent of human verbal faculty to the communication systems of other hominins preceding *Homo sapiens* in human phylogeny, giving rise to a model of language origin in line with the principles of gradualism and the continuism of the Darwinian naturalistic tradition.

## 2. Methodological and theoretical assumptions

The assumption at the basis of this article is that the processes underlying language comprehension and production and those which are at the basis of language origins are closely related. From this point of view, studying how language works (how it is processed; what the cognitive systems are that make possible verbal production and comprehension) is an extremely worthwhile way to make a hypothesis about how language may have originated (cf. Ferretti & Adornetti, 2012; Ferretti & Adornetti, 2014). This assumption does not apply to any model of language: it is valid only for certain cognitive systems and for specific communicative properties that such systems manage. In our view,

for example, it does not apply to theoretical models based on grammar, such as Chomsky's Universal Grammar (UG). This is for two basic reasons: the first is specific and is tied to the Chomskyan paradigm; the second is more general and is linked to the idea that the essence of language is the grammatical competence (not necessarily identified with UG).

The specific reason has to do with the incompatibility of UG with the principles of gradualism and continuism of the Darwinian tradition (cf. Corballis, 2011). Indeed, according to Chomsky the computational device at the base of syntax (the device that allows the operation "Merge") is unique to *Homo sapiens* and «has not evolved in any significant way since human ancestors left Africa, approximately 50. 000–80 000 years ago [...]. *The human language faculty emerged suddenly in evolutionary time and has not evolved since*» (Berwick et al., 2013, p. 89, our emphasis; cf. also Chomsky, 2010). Corballis (2013) defines this idea as miracoulus (see also Deacon, 2010) and writes:

The idea that language emerged suddenly is comforting to those eager to demonstrate human uniqueness and mental superiority. [...] Nevertheless, 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, p. 35)

In our opinion there is a second general reason that makes the hypothesis that the origin of language has to do with the advent of grammatical competence problematic. According to several authors (e.g. Arbib, 2005; Tomasello, 2008), in fact, grammar emerged because of a need to construct a code in order to make more efficient sophisticated forms of communication that *already* were in possession of our ancestors. In this sense, grammar is a late product in the evolution of language: if you want to give an account of the *origin* of language, grammar is not a good point to start with. When the aim is the study of language from an evolutionary perspective, it is necessary to overturn the traditional view proposed in Chomskyan linguistics. From the generativist perspective the main study of linguistics is syntax; semantic features are added when grammar is not sufficient to ensure understanding; pragmatics is essentially a container that stores all that is left out by the syntactic and semantic analysis. Nevertheless, as Gärdenfors (2004, p. 245, see also his article in this volume) points out:

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. As a consequence, the pragmatic aspects of language are the most fundamental from an evolutionary point of view. [...] when the goal is to develop a theory of the evolution of communication, the converse order—pragmatics before semantics before syntax—is more appropriate. In other words, there is much to find out about the evolution of communication before we can understand the evolution of semantics and syntax.

From such theoretical considerations, in this paper we argue that in order to propose a model of language's evolutionary origins plausible, it is necessary to move from the syntax based paradigms to those that are pragmatic based. We show that this transition is closely related to the transition from sentence (the essence of language in syntactic-centric perspective) to discourse. In particular, our proposal is that a key role in the origin of human verbal skills is played by the cognitive devices involved in the processing of a specific pragmatic property emerging at the level of discourse: coherence. In support of this proposal, in the following paragraphs we show that 1) coherence is an essential property for effective communication that emerges at the discourse level, rather than on that of individual sentences, and this does not depend on syntactic and grammatical elements; 2) that cognitive systems responsible for coherence are the executive functions involved in action planning, control and organization; 3) that these systems have evolved in the course of human phylogeny in the context of making stone tools and that the operations underlying the production of these tools are similar to those required for processing of coherence. On the basis of these considerations, we argue that the hominins preceding *Homo sapiens* during evolution were cognitively equipped for the development of a communication system based on proto-discursive forms governed by a principle of coherence.

### 3. The crucial role of coherence in human communication

Coherence generally can be defined as the conceptual organizational aspects of discourse at the suprasentential level. More specifically, coherence is the way through which arguments in a discourse are structurally organized towards a

goal, plan or a general theme (Glosser & Deser, 1990). When is a discourse coherent? The theoretical models that equate language with grammar and linguistic processing with sentence processing (cf. Pickering, Clifton & Crocker, 2001) explain discourse coherence in terms of the linear relations of cohesion between consecutive sentences (e.g. Daneš, 1974; Halliday & Hasan, 1976; Reinhart, 1980). The most influential work on cohesion is the volume *Cohesion in English* by Halliday and Hasan published in 1976; the two authors define cohesion as the relations of meaning existing within a text that «enable one part of the text to function as the context for another» (Halliday & Hasan, 1989, p. 489, quoted in Bublitz, 2011 p. 38). In a text cohesion relations are accomplished through grammatical and lexical elements. Grammatical cohesion includes elements such as reference, substitution, ellipsis and conjunctions; lexical cohesion is based on reiteration (e.g., repetition, synonymy) and collocation (co-occurrence of lexical item). What is important to emphasize for the purpose of our argument is that in this perspective cohesion is a *necessary* condition for discourse coherence. Consider in this regard the following text:

- (1) After the forming of the *sun* and the *solar system*, our star began its long existence as a so-called dwarf star. In the *dwarf phase* of its life, the energy that the *sun* gives off is generated in its core through the fusion of hydrogen into helium (Berzlánovich, 2008, p. 2).

In this text the sentences are connected through lexical cohesion: the lexical cohesive relations hold among the lexical items *sun*, *solar system*, *star*, *dwarf star* and *dwarf phase* in the text.

The model by Halliday and Hasan has been criticized over the years primarily because of its insistence on conceiving cohesion as a necessary property for the creation of unity in texts (for a discussion, see Tanskanen, 2006; Giora, 2014). Several researchers demonstrated that cohesion was not necessary at all to make a text appear a unified whole (e.g. de Beaugrande & Dressler, 1981, p. 3; Brown & Yule, 1983, p. 195; Enkvist, 1978; Giora, 1985; Sanford & Moxey, 1995). To this end, Enkvist proposed the following example:

- (2) I bought a Ford. The car in which President Wilson rode down the Champs Élysées was black. Black English has been widely discussed. The

discussions between the presidents ended last week. A week has seven days. Every days I feed my cat. Cats have four legs. The cat is on the mat. Mat has three letters (Enkvist, 1978, pp. 110-111).

In this text the sentences are connected through the mechanism of repetition. However, the set of sentences, despite the abundance of cohesive ties, is not perceived as a coherent whole. Indeed, this text is characterized only by *local coherence*, but not *global coherence* (cf. Glosser & Deser, 1990). Local coherence refers to the conceptual links between consecutive individual sentences or propositions. Instead, global coherence is the manner in which discourse is organized with respect to an overall goal, plan, theme, or topic. As the previous example has shown, cohesion (i.e., grammatical and lexical devices) is responsible for local coherence, not for that global: the cohesive bonds between adjacent sentences do not guarantee the overall coherence of the discourse. Consider, instead, the following text:

- (3) George's high pass was headed to the right. The forward shot at once without dribbling and made a goal. The referee declared the kick off-side (Enkvist, 1978, p. 111).

Unlike (2), there are not cohesive ties between adjacent sentences in (3). However, the text (3) is pragmatically appropriate because the topic under discussion is clear. These examples show that the linear concatenation of sentences based on cohesion is not a necessary condition of the overall coherence of the discourse because: a) it does not ensure the overall coherence of the discourse; b) it is possible to have coherent discourse even in the absence of cohesive ties.

However, it can be argued that Enkvist's text in (2) is simply an artificial construction that does not reflect how human beings communicate with each other. This is not wholly true. Indeed, neurolinguistics and neuropsychological research showed that in several neurological patients there is a dissociation between the abilities that underlie sentence processing (microstructure or microanalysis) and those that underlie discourse processing (macrostructure or macroanalysis) (e.g., Davis et al. 1997; Marini et al. 2008). Specifically, results of several studies support the idea that global coherence does not depend on the skills involved in the processing of individual sentences and on the skills implied in the processing of local coherence. Particularly relevant to

the purposes of this paper are the studies on the discursive capacities of individuals with traumatic brain injury (TBI). TBI subjects generally do not present serious difficulties processing individual sentences (they have no problems processing lexical items and grammatical aspects) and local coherence, but they have deficits in the organization of global discourse (e.g., Coelho, 2002; Coelho et. al., 2012; Davis & Coehlo 2004; Galetto et al., 2013; Strauss Hough & Barrow, 2003; Marini et al., 2011; McDonald 2008). Consider, for example, the following transcript of a discourse of a TBI patient (C), with a therapist (T) (Perkins 2007, p. 86):

C: I admit this government we've got is not doing a good job but the unions are trying to make them sound worse than what they are

T: mm

C: they . they . cos I'm a Tory actually but I do vote . if there's a . er . a communist bloke there I will vote communist but . it all depends what his principles are but I don't agree . with the Chinese communism . and the Russian communism

T: right

C: but I believe every . should be equal but . I'm not knocking the royal family because you need them

T: mm

C: and they they they bring people in to see take photos

Despite the local sequential links between trade unions–government, government–Tory, Tory–communist, communism–Chinese/Russian communism, communism–equality, equality–Royal Family, Royal Family–tourist attraction, C shows a form of *topic drift*: he is unable to monitor what has already been talked about or to relate each individual utterance to some overall coherent plan or goal. In general, the studies on TBI patients showed that global coherence does not rely on the grammatical and lexical skills underling the processing of the single sentences: producing a coherent discourse does not correspond to put a well-formed sentence after another.

#### 4. Discourse coherence and executive functions

Studies on TBI subjects are particularly relevant for our purposes because they shed light on the cognitive systems underlying discourse coherence processing. According to several scholars the problems of coherence of TBI patients are due to the deficit of executive functions that generally affect these patients (e.g., Biddle et al., 1996; Marini et al., 2014; McDonald, 2008; Perkins, 2007; for a discussion see Adornetti, 2013, 2014).<sup>1</sup>

The expression *executive functions* (EF) is an umbrella term that encompasses a wide range of cognitive and behavioral skills (Alvarez & Emory, 2006; Banich, 2009; Barkley, 2012; Jurado & Rosselli, 2007). From a general point of view, it is possible to characterize executive functions as the higher-order cognitive processes, mainly mediated from the areas of the prefrontal cortex (the anterior portion of the frontal lobes), that are needed to guide behavior toward a goal in non-routine contexts and in complex and conflicting situations (Banich, 2009; Gilbert & Burgess, 2008). According to Lezak (1982), executive functions allow the formulation of goals, planning, and carrying out plans effectively. Welsh and Pennington (1988, pp. 201–202) defined executive functioning as the capacity «to maintain an appropriate problem-solving set for attainment of a future goal».

Among the different skills of which EF are composed, a crucial role for the organization of behavior is played by action planning. This ability allows creating and performing goal-oriented behaviors through the identification and the appropriate organization of the elements necessary to achieve a goal. Action planning involves different stages and processes ranging from the conceptual formulation of the plan (identification of the ultimate goal; splitting the final goal into sub-goals; prediction of the consequences of the actions required to achieve the sub-goals) to its execution. The several stages of action planning take place in different areas of the frontal lobes: the more rostral (anterior) frontal regions are involved in the processing of more abstract goals and more temporally extended actions; in the most caudal (posterior) regions

<sup>1</sup> TBI patients are not an entirely homogeneous group from a cognitive point of view: the areas of the brain most frequently damaged after a brain injury are the frontal and parietal regions, but on the basis of the type and the mechanical characteristics of the trauma other areas of the brain can be injured. Therefore, the subjects may suffer from disturbance of various nature. Particularly relevant for our purposes are the neuro-behavioral disorders resulting from brain injuries in the frontal and prefrontal lobes. Such injuries, in fact, impair the executive functions and this has important consequences on the discursive abilities of TBI. In this paper we discuss the studies on TBI with prefrontal and frontal injuries.

more concrete information is processed and is more closely linked to the actual motor response (Badre, 2008; Badre & d'Esposito, 2007). Particularly important for the construction of global coherence are the phases of the conceptual formulation of the plan. Research conducted on healthy subjects through neuroimaging techniques have shown that areas most clearly involved in these planning tasks are the dorsolateral prefrontal regions (e.g., Baker et al., 1996; Fincham et al., 2002; Tanji et al., 2007). Crescentini and colleagues (2012), for example, used functional magnetic resonance imaging (fMRI) during the execution of the test of the *Tower of Hanoi* (one of neuropsychological tests used to evaluate planning skills)<sup>2</sup> on healthy subjects and showed that the dorsolateral prefrontal cortex (especially in the right hemisphere) is activated preferentially during the initial stages of planning and this activation is related to the generation and evaluation of abstract sequences of responses that have to be implemented. To successfully make plans (to perform goal-oriented behaviors), a constant monitoring of the task in progress is also needed. Monitoring can be defined as «the capacity to hold abstract coded representations of events that are expected to occur, so as to mark their occurrence or non-occurrence (i.e. monitor their relative status in relation to each other and the intended plan)» (Petrides, 2005, p.789). This capacity, of which the main neural substrate is the right lateral prefrontal cortex (see Stuss & Alexander, 2007; Vallesi & Crescentini, 2011), is needed to calibrate the effects of actions on the environment, to detect possible errors, to enable corrective action where there is a mismatch between the behavioral responses (effects) and the mental representations (goals and expectations) of those responses, and to reorganize the following steps.

Significant research has shown that TBI subjects, because of injuries in specific areas of the prefrontal regions, have difficulty in managing the conceptual aspects involved in the planning of goal-oriented behaviour and that, as a consequence, these difficulties are reflected in the execution of the corresponding actions (e.g. Duncan, 1986; Eslinger & Damasio, 1985; Zalla et al., 2003). Zalla and colleagues (2001) have shown that TBI patients with injury in the anterior and in dorsolateral regions of the prefrontal cortex have

<sup>2</sup>The *Tower of Hanoi* test is a mathematical game/ puzzle which consists of three pegs and a number of discs of different sizes which can slide onto any peg. The puzzle starts with the discs neatly stacked in order of size on one peg, smallest at the top, thus making a conical shape. The object of the game is to move the entire stack to another peg, obeying the following two rules: only one disc may be moved at a time; no disc may be placed on top of a smaller disc.

difficulty in the formulation and evaluation of a coherent and well-structured conceptual plan and that these difficulties impair the execution of the action. Similarly, Kliegel and colleagues (2004) have found that TBI patients are not able to plan and carry out actions that require the formulation and execution of complex intentions (prolonged in time).

The problems that TBI patients have at the level of action organization are the basis of the discursive pragmatic deficits of which they suffer. Biddle and colleagues (1996), for example, explicitly put in relation the disorders of global coherence of TBI with their deficits of action planning. According to the authors:

The narrative impairments of the population of adults and children with TBI in this study appeared to be the results of problems with planning, production and monitoring discourse. [...] It is possible that the disruptions evident in the narrative of the children and adults with TBI were related less to a language impairments that to difficulties with the executive processes utilized in discourse production. (Biddle et al., 1996, p. 463)

The existence of a causal relationship between executive dysfunctions and narrative deficits of TBI has also been proposed by Coelho (2002). Interesting results in this regard have also been obtained by Coelho et al. (2012) who have demonstrated the existence of a causal relationship between coherence deficits and executive dysfunctions at the level of brain areas. Indeed, the authors have shown that TBI patients with lesions on the left dorsolateral prefrontal cortex (the main neural substrate of conceptual planning skills) have difficulties in managing the global coherence of discourse.

To sum up this section: discourse production relies on the ability of the speaker to organize the verbal utterances towards of a general purpose (the global theme of discourse) by identifying the correct sequence of steps needed to reach it. During the execution of the plan, that is to say during the discursive production, a constant monitoring and control of the task is needed to avoid inserting irrelevant material into the overall previously planned theme (cf. Adornetti, 2013; Ferretti et al., 2013, p. 329-330). Then, the deficit of discourse coherence of TBI can be considered as the linguistic manifestation of a more general cognitive problem that concerns, first, the level of action organization and only secondly the level of organization of language.

## 5. Making tools and planning (proto)discourses

The basic tenet of this paper is that the discourse coherence is an essential property both of language functioning and of its genesis and evolution. By putting at the basis of the origin of language such property, we believe it is possible, in fact, to explain the transition from simple expressive systems without a grammar and a shared code between speakers and listeners (as it is legitimate to assume the primordial forms of human communication were) to gradually more complex systems. Our hypothesis is that a key moment in the evolution of language was the exaptation for communication purposes of the executive functions originally tied to action control and that such exaptation has given birth to proto-discursive forms of communication governed by coherence.



Figure 1: Oldowan chopper  
(from Johanson & Edgar, 1996).



Figure 2: Acheulean hand axe  
(from Johanson & Edgar, 1996).

A first step to validate this hypothesis is to show the existence in extinct hominins of the cognitive skills underlying coherence in modern human beings. Interesting evidence in this regard comes from cognitive archeology (cf. for a general presentation of cognitive archeology Coolidge et al., 2015; Mahaney 2014; Wynn 2002). Indeed, archeological records indicate the timing and

context of distinctively human elaborations to the executive control functions, including the gradual emergence of increasingly complex action control (Stout, 2010; Stout, 2011; Stout et al., 2008; Uomini & Meyer, 2013). A key moment of this emergence has been the development of stone tool-making industries, especially the transition from the earlier Oldowan industry (from 2,6 Mya to 1,5 Mya) to the later Acheulian industry dating from around 1.7 million years ago with *Homo ergaster* and *Homo erectus*. Using fMRI<sup>3</sup>, Stout and colleagues (2008) have shown that the making of Oldowan tools (stone tools consisting of nothing more than sharp stone flakes struck from river cobbles: see Figure 1) implies, in particular, the involvement of the posterior parietal areas (responsible for objects recognition), and ventral premotor areas (managing manual grasping). However, there is no significant activation of the prefrontal cortex. This means that Oldowan industry is characterized by a complex motor-manipulative activity, but limited executive capacities of action planning and organization. An increase in the development of the capacities of action control has been the emergence of the Acheulian industry 1.7 million years ago. The characteristic Acheulean tool is a two-sided hand axe (see Figure 2). This artefact is built by modelling a large stone on both sides until an almond-shaped symmetrical and regular stone is obtained. According to Stout (2010) this technique requires a more complex level of hierarchical control than individual flake removals, which must be subordinated to the broader goal of shaping the piece and seems to entail the involvement of the lateral prefrontal cortex. This area, in fact, allows humans to assemble the individual removal in wider coherent chunk action (Koechlin & Jubault, 2006) and to manage relations of increasing abstraction during the execution of the action (Badre & D'Esposito, 2007), namely to relate the individual flake removals with the realization of the general shape of the nucleus. Such operations, as we have seen, are also crucial for producing organized discourses.

On the basis of these considerations, our hypothesis is that the hominis who made Acheulean tools have developed systems of communication based on protodiscourses governed by a principle of coherence. The point, of course, is to clarify what is meant by “discourse” in such circumstances (in circumstances where there were no sentences because there was not a grammar that regulated their formation) (see on this point Ferretti, this volume). An interesting

<sup>3</sup> In the experiments were observed the brain activations of modern human beings who had been asked to produce stone tools through the process of chopping characteristic of prehistoric lithic industries.

characterization in this regard is provided by Linell (1998) who claims it is possible conceive discourse as a connection of sequences of communicative actions performed through verbal and nonverbal strategies. As Bowie (2008) stresses, a conception of this kind allows to give an account of the continuity from communicative exchanges using simple resources (such as single words or gestures) to those that use complex grammatical conventions: the discourse understood as connected sequences of communicative actions can be constructed exploiting any communication resource that participants are able to manage and does not depend on sentence or on a particular level of grammatical complexity. Our hypothesis is that coherence (which primarily concerns the level of action organization) has had an important function in structuring these sequences of communicative actions. From this point of view, coherence can be conceived as a functional equivalent on the level of pragmatics of syntax.

The hypothesis of discourse as a sequence of communicative actions is, in our view, compatible with the proposal made by different scholars for whom a crucial stage in the evolution of language has been that in which hominins used to communicate through forms of mimesis (Collins, 2013; Corballis, 2011; Corballis, this volume; Donald, 1991; Zlatev, 2008; Zlatev, this volume). Donald (1991, p. 168) defines mimesis as «the ability to produce conscious, self-initiated, representational acts that are intentional but not linguistic». It may assume various forms: pantomime, imitation, ritualized behaviors, etc. A mimetic act is a performance that reflects the perceived event structure of the world and has three behavioural manifestations: rehearsal of skill (the actor imagines and reproduces previous performances to improving them); re-enactive mime (patterns of actions of others are reproduced in the context of play); non linguistic gesture (an action communicate an intention through resemblance) (see Donald, 2012, p. 180). The underling cognitive process can be broken down into a standard sequence: construction of a plan of action; execution of an approximation of the action; comparison, in imagination, of the performed act to the intended one. According to Donald (e.g. 2012, p. 182), the strongest argument for the early emergence of mimesis in hominin evolution is that mime and non-linguistic gesture come free with skill, because the neuro-cognitive mechanism and the computational process are the same. Thus, the archaeological evidence of refined tools is evidence of a mimetic capacity in our ancestors. Specifically, according to Donald (1991), mimesis arose with the emergence of *Homo ergaster*, the hominin who first introduced

the Acheulean industry. This conception of mimesis fit very well with the idea of proto-discourse presented earlier. In fact, like (proto)discourse, mimetic communication also involves the capacity to organize temporally action sequences directed to the accomplishment of a specific communicative goal. Indeed, the proposal of Donald (2001, pp. 263-266) is that:

Mimesis is the result of evolving better conscious control over action. [...] Mimetic capacity was primarily the result of merging the executive brain with the action brain, when the hominid executive brain system extended its anatomical territory into the frontal and subcortical regions that control voluntary action.

The evolution of forms of proto-discourse governed by coherence may have been forced by the need of human mind to produce stories and narratives (e.g., Boyd, 2009; Gottschall, 2012). According to several scholars, narrative has played an important adaptive function in human evolution because it offers a way to simulate an experience (representing the human social, physical and mental environment) and to draw conclusions about the real world (cf. Bower & Morrow, 1990; Gottschall, 2012). As Sugiyama (2001, p. 224) states:

The interactions of story characters, for example, can be seen as models of the human social environment that enable an individual to observe local consequences of a wide variety of actions (e.g., incest, marital infidelity, homicide). These models can be used both to acquire information and to refine knowledge before putting it into actual practices.

What is important to stress is that narrative does not require a complex grammar to operate (indeed narrative does not require language at all: cf. Boyd, 2009, p. 159). But grammar, of course, makes storytelling more accurate and efficient. For this reason, it is possible to speculate that the need to share the information included in stories in a more precise way has brought grammatical structure to the communicative system. In this sense, pragmatics (in the form of coherent proto-discourses) precedes and is the condition for grammar to emerge.

## Conclusions

In this article we argued that a particularly fruitful way to account for the origin of language is to analyze the systems and cognitive skills involved in its actual functioning. We suggested that a key role in the advent of human language was played by the executive functions involved in the processing of a specific pragmatic property: discourse coherence. Through the results from cognitive archeology, we showed that these systems have evolved in the course of human phylogeny in the context of the making of stone tools and the operations requested to produce these tools overlapping with those requested to process coherence. On the basis of these considerations, we have hypothesized that our ancestors developed forms of proto-discursive communication governed by a principle of coherence. In a scenario of this kind, the origin of human verbal skills is primarily a matter of pragmatics rather than of grammar.

## REFERENCES

- Adornetti, I. (2013). *Il farsi e il disfarsi del discorso. Pragmatica del linguaggio e processi cognitivi*. Firenze: Le Lettere.
- Adornetti, I. (2014). A Neuro-Cognitive Perspective on the Production and Comprehension of Discourse Coherence. In P., Chruszczewski, J. Rickford, K. Buczek, A. Knapik, J. Mianowski, S. Wacewicz, & P. Żywicznyński (Eds.), *Ways to Protolanguage 3*. Wroclaw: Wydawnictwo WSF, 9–24.
- Alvarez, J.A., & Emory, E. (2006). Executive Function and the Frontal Lobes: A Meta-Analytic Review. *Neuropsychology Review*, 16, 17–42.
- Arbib, M. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105–167.
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal axis of the frontal lobes. *Trends in Cognitive Science*, 12, 193–200.
- Badre, D., & D'Esposito, M. (2007). fMRI evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience*, 19(12), 1–18.
- Baker, S.C., Rogers, R. D., Owen, A.M., et al. (1996). Neural systems engaged by planning: a PET study of the Tower of London task. *Neuropsychologia*, 34, 515–526.

- Banich, M.T. (2009). Executive Function: The Search for an Integrated Account. *Current Directions in Psychological Science*, 18(2), 89–94.
- Barkley, R. (2012). *Executive Functions. What They Are, How They Work and How They Evolved*. New York: The Guildford Press.
- de Beaugrande, R. & Dressler, W. (1981). *Introduction to Text Linguistics*. London & New York: Longman.
- Berwick, R., Friederici, A., Chomsky, N., & Bolhuis, J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Science*, 17(2), 89–98.
- Berzlánovich, I. (2008). *Lexical cohesion and the organization of discourse*. <http://www.rug.nl/let/onderzoek/onderzoekinstututen/clcg/berzlanovich.Pdf> (accessed 15 September 2014).
- Biddle, K.R., McCabe, A., & Bliss, L.S. (1996). Narrative skills following traumatic brain injury in children and adults. *Journal of communication disorders*, 29(6), 447–469.
- Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science*, 247(4938), 44–48.
- Bowie, J. (2008). Proto-discourse and the emergence of compositionality. *Interaction Studies*, 9(1), pp. 18–33.
- Boyd, B. (2009). *On the origin of stories. Evolution, cognition and fiction*. Cambridge: The Belknap Press of Harvard University Press.
- Brown, G. & Yule, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.
- Bublitz, W. (2011). Cohesion and Coherence. In J. Zienkowski, J. Östman, & J. Verschueren (Eds.), *Discursive Pragmatics*. Amsterdam-Philadelphia: John Benjamins Publishing Company, 37–49.
- Chomsky, N. (2010). Some simple evo-devo theses: How true might they be for language?. In R.K. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language*. Cambridge: Cambridge University Press, 45–62.
- Coelho, C. (2002). Story narratives of adults with closed head injury and non-brain injured adults: Influence of socioeconomic status, elicitation task, and executive functioning. *Journal of Speech, Language, and Hearing Research*, 45(6), 1232–1248.

- Coelho, C., Lê, K., Mozeiko, J., Krueger, F., & Grafman, J. (2012). Discourse production following injury to the dorsolateral prefrontal cortex. *Neuropsychologia*, *50*(14), 3564–3572.
- Collins, C. (2013). *Paleopoetics: The Evolution of the Preliterate Imagination*. New York: Columbia University Press.
- Coolidge, F.L., Wynn, T., Overman, K. A., & Hicks, J. M. (2015). Cognitive Archaeology and the Cognitive Sciences. In E. Bruner (Ed.), *Human Paleoneurology*. Berlin: Springer Series in Bio-/Neuroinformatics 3, 177–208.
- Corballis, M. C. (2011). *The Recursive Mind. The origins of human language, thought and civilization*. Princeton: Princeton University Press.
- Corballis, M. C. (2013). Toward a Darwinian perspective on language evolution. In Y. Coello, & A. Bartolo (Eds.), *Language and Action in Cognitive Neuroscience*. New York: Psychology Press, 33–58.
- Corballis, M. C. (2014). The gradual emergence of language. *Humana.Mente. Journal of Philosophical Studies*, *27*.
- Crescentini, C., Allaei, S., Vallesi, A., & Shallice, T. (2012). Two networks involved in producing and realizing plans. *Neuropsychologia*, *50*(7) 1521-1535.
- Daneš, F. (1974). *Functional sentence perspective and the organization of the text*. In F.Daneš (ed.), pp. 106–128.
- Davis, A. G., & Coelho, C. (2004). Referential Cohesion and Logical Coherence of Narration after Closed Head Injury. *Brain and Language* *89*, 508–523.
- Davis, G. A., O’Neil-Pirozzi, T. M., & Coon, M. (1997). Referential cohesion and logical coherence of narration after right hemisphere stroke. *Brain and Language*, *56*(2), 183–210.
- Deacon, T. (2010). A role for relaxed selection in the evolution of the language capacity. *PNAS*, *107*, Supplement 2, 9000–9006.
- Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Human Culture*. Cambridge, MA: Harvard University Press.
- Donald, M. (2001). *A Mind so Rare: The Evolution of Human Consciousness*. New York: Norton.
- Donald, M. (2012). The mimetic origins of language. In M. Tallerman & K. R. Gibson (Eds.), *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press, 180-183.

- Duncan, J. (1986). Disorganization of behavior after frontal lobe damage. *Cognitive Neuropsychology*, 3(3), 271–290.
- Enqvist, N.L. (1978). Coherence, Pseudo-Coherence, and Non-Coherence. In J. Ostman (Ed.), *Semantics and Cohesion*. Abo Akademi, 109–130.
- Eslinger, P.J., & Damasio A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology*, 35(12), 1731–1741.
- Ferretti, F. (2014). Travelling in Time and Space at the Origins of Language. *Humana.Mente – Journal of Philosophical Studies*, 27.
- Ferretti, F., & Adornetti, I. (2012). *Dalla comunicazione al linguaggio. Scimmie, ominidi e umani in una prospettiva darwiniana*. Milano: Mondadori Università.
- Ferretti, F., & Adornetti, I. (2014). Against linguistic Cartesianism: Toward a naturalistic model of human language origins and functioning. *Language & Communication*, 37, 29–39.
- Ferretti, F., Adornetti, I., Cosentino, E., & Marini, A. (2013). Keeping the route and speaking coherently: the hidden link between spatial navigation and discourse processing. *Journal of Neurolinguistics*, 26(2), 327–334.
- Fincham, J.M., Carter, C.S., van Veen, V., Stenger, V.A., & Anderson, J.R. (2002). Neural mechanisms of planning: A computational analysis using event-related fMRI. *PNAS*, 99, 3346–3351.
- Galetto, V., Andrectta, S., Zettin, M., & Marini, A. (2013). Patterns of impairment of narrative language in mild traumatic brain injury. *Journal of Neurolinguistics*, 26(6), 649–661.
- Gärdenfors, P. (2004). Cooperation and the Evolution of Symbolic Communication. In D.K. Oller, & U. Griebel (Eds.), *Evolution of Communications Systems: A Comparative Approach*. Cambridge: MIT Press, 237–256.
- Gärdenfors, P. (2014). The Evolution of Sentential Structure. *Humana.Mente Journal of Philosophical Studies*, 27.
- Gilbert, S., & Burgess, P. (2008). Executive Function. *Current Biology*, 18, R110–R114.
- Giora, R. (1985). Notes towards a Theory of Text Coherence. *Poetics Today* 6(4), 699–715.
- Giora, R. (2014). Cognitive and Social Aspects of Coherence. In T. Holtgraves (Ed.), *Handbook of Language and Social Psychology*. New York, NY: Oxford University Press, 141–153.

- Glosser, G., & Deser, T. (1990). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language, 40*, 67–88.
- Gottschall, J. (2012). *The Storytelling Animal. How Stories Make Us Human*. Boston: Houghton Mifflin Harcourt.
- Halliday, M. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. K., & Hasan, R. (1989). *Language, context, and text: aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.
- Johanson D., & Edgar B. (1996). *From Lucy to language*. New York: Simon & Schuster editions.
- Jurado, M.B., & Rosselli, M. (2007). The Elusive Nature of Executive Functions: A Review of our Current Understanding. *Neuropsychological Review 17*, 213–233.
- Kliegel, M., Eschen, A., & Thöne-Otto, A. I. (2004). Planning and realization of complex intentions in traumatic brain injury and normal aging. *Brain and Cognition, 56*(1), 43–54.
- Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron, 50*(6), 963–974.
- Linell, P. (1998). *Approaching dialogue: talk, interaction and contexts in dialogical perspectives*. Amsterdam- Philadelphia: Benjamins.
- Lczak, M.D. (1982). The problem of assessing executive functions. *International Journal of Psychology, 17*, 281–297.
- Mahaney, R. A. (2014). Lithic analysis as cognitive science: a framework, *Lithic Technology, 39*(3), 173–189.
- Marini, A., Galetto, V., Zampieri, E., Vorano, L., Zettin, M., & Carlomagno, S. (2011). Narrative language in Traumatic Brain Injury. *Neuropsychologia, 49*, 2904–2910.
- Marini, A., Spoletini, I., Rubino, I. A., Ciuffa, M., Bria, P., Martinotti, G., Banfi, G., Boccascino, R., Strom, P., Siracusano, A., Caltagirone, C., & Spalletta, G. (2008). The language of schizophrenia: An analysis of micro and macrolinguistic abilities and their neuropsychological correlates. *Schizophrenia Research, 105*(1), 144–155.

- Marini, A., Zettin, M., & Galetto, V. (2014). Cognitive correlates of narrative impairment in moderate traumatic brain injury. *Neuropsychologia*, *64*, 282–288.
- McDonald, S. (2008). Communication and language disturbances following traumatic brain injury. In B. Stemmer, & H. Whitaker (Eds.), *Handbook of neurolinguistics*. San Diego: Academic Press, 485–494.
- Perkins, M. (2007). *Pragmatic Impairment*. New York: Cambridge University Press.
- Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philosophical Transaction of the Royal Society of London B*, *360*, 781–795.
- Pickering, M., Clifton, C., & Crocker, M. (2001). Architectures and Mechanism in Sentence Comprehension. In M. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and Mechanisms for Language Processing*. Cambridge: Cambridge University Press, 1–28.
- Reinhart, T. (1980). Conditions for text coherence. *Poetics Today*, *1*, 161–180.
- Sanford, A. J. & Moxey, L. M. (1995). Aspects of coherence in written language: a psychological perspective. In M.A. Gernsbacher & T. Givón (Eds.), *Coherence in Spontaneous Text*. Amsterdam & Philadelphia: John Benjamins, 161–187.
- Stout, D. (2010). The Evolution of Cognitive Control. *Topics in Cognitive Science*, *2*(4), 614–630.
- Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society of London B*, *366*, 1050–1059.
- Stout, D., Toth, N., Schick, K., & Chaminade, T. (2008). Neural correlates of Early Stone Age toolmaking: technology, language and cognition in human evolution. *Philosophical Transactions of the Royal Society B* *363*(1499), 1939–1949.
- Strauss Hough, M., & Barrow, I. (2003). Descriptive discourse abilities of traumatic brain-injured adults. *Aphasiology*, *17*(2), 183–191.
- Stuss, D., & Alexander, M. (2007). Is there a dysexecutive syndrome? *Philosophical Transactions of the Royal Society of London B*, *362*, 901–915.
- Sugiyama, M. (2001). Food, foragers, and folklore: The role of narrative in human subsistence. *Evolution and Human Behavior*, *22*(4), 221–240.

- Tanskanen, S.K. (2006). *Collaborating towards Coherence. Lexical Cohesion in English Discourse*. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Tanji J. Shima, K., Mushiake, H. (2007). Concept-based behavioral planning and the lateral prefrontal cortex. *TRENDS in Cognitive Science*, 11(2), 528–534.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT press.
- Uomini, N.T., & Meyer, G.F. (2013). Shared Brain Lateralization Patterns in Language and Acheulean Stone Tool Production: A Functional Transcranial Doppler Ultrasound Study. *PLoS ONE* 8(8), e72693.
- Vallesi, A., & Crescentini, C. (2011). Right fronto-parietal involvement in monitoring spatial trajectories. *Neuroimage*, 57(2), 558–564.
- Welsh, M.C., & Pennington, B.F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, 4, 199–230.
- Wynn, T. (2002). Archaeology and cognitive evolution. *Behavioral and Brain Sciences*, 25(03), 389–402.
- Zalla, T., Plassiart, C., Pillon, B., Grafman, J., & Sirigu, A. (2001). Action Planning in a Virtual Context after Prefrontal Cortex Damage. *Neuropsychologia*, 39, 759–770.
- Zalla, T., Pradat-Diehl, P., & Sirigu, A. (2003). Perception of Action Boundaries in Patients with Frontal Lobe Damage. *Neuropsychologia*, 41(12), 1619–1627.
- Zlatev, J. (2008). From proto-mimesis to language: Evidence from primatology and social neuroscience. *Journal of Physiology – Paris*, 102, 137–152.
- Zlatev, J. (2014). Human uniqueness, bodily mimesis and the evolution of language. *Humana.Mente. Journal of Philosophical Studies*, 27.



# Travelling in Time and Space at the Origins of Language

*Francesco Ferretti*<sup>†</sup>  
francesco.ferretti@uniroma3.it

## ABSTRACT

In this paper we propose a narrative hypothesis on the nature of language and a proto-discursive hypothesis on the origin of our communicative abilities. Our proposal is based on two assumptions. The first assumption, concerning the properties of language, is tied to the idea that global discourse coherence governs the origin of our communicative abilities as well the functioning of these abilities. The second assumption, concerning processing devices, is connected to the idea that the systems of spatial and temporal navigation are implicated in discourse coherence processing. Analysis of the relationship between these two assumptions allows us to integrate the model of language based on clues proposed by Sperber and Wilson with Relevance Theory with the discursive foundation of human communication. In this respect, our proposal can be considered as a tentative extension of Relevance Theory (both at the level of properties and the level of cognitive systems).

Keywords: discourse coherence, mental time travel, mindreading, origin of language, protodiscourse, relevance theory, spatial navigation, storytelling.

## 1. Introduction

In this paper we propose a narrative hypothesis on the nature of language and a proto-discursive hypothesis on the origin of language. These hypotheses are based on two assumptions: the first concerns the level of the properties of language; the second pertains to the level of processing devices. Regarding the properties, to argue that human communication has a narrative foundation and a proto-discursive origin necessitates assigning a leading role, among the

<sup>†</sup> Department of Philosophy, Communication and Visual Arts, Roma Tre University, Italy.

characteristics of language, to global coherence (a pragmatic property). Regarding cognitive architecture, the reference to global coherence implies the presence of cognitive devices capable of processing a type of information very different from that processed by the devices that analyze the constituent structure of sentences. At the basis of our hypothesis is the concept that the narrative foundation of language and its proto-discursive origin are closely related to the functioning of cognitive systems that allow individuals to identify a goal to move toward as well as to construct and keep the correct route in order to reach it. In other words, both the actual functioning of language and its evolutionary roots rely on processing devices governing navigation in space and time.

## 2. The narrative nature of human language

Sciences focused on the mind as well as on evolution have produced results that, in turn, greatly influence the study of human narrative capacity, which is currently the center of a compelling debate (Boyd, 2009; Collins, 2008, 2013; Dautenhahn, 2002; Gottschall, 2012; Gottschall and Wilson, 2005; Herman, 2013; Hirstein, 2005; Sugiyama, 2005). One of the hottest conceptual issues in this debate is the question of whether the ability to tell stories (i.e. storytelling) is an evolutionary side effect with no impactful role in the adaptation of our species (Bloom, 2010), or if the human capacity for narrative does indeed play a specific evolutionary role (Boyd, 2009; Pinker, 2007). Gottschall (2012), for one, asserts that storytelling capacities play an important adaptive role in human evolution. In his view, narrative ability works in a very similar way to a flight simulator: it allows humans to gain experience of the most intricate affairs of life by sitting safely in a literal or figurative armchair.

In this article we do not intend to take a position on the nature of human narrative ability, rather we aim to propose a hypothesis on the nature of language. Nevertheless, our inquiry is closely tied to the analysis of storytelling ability: at the basis of our hypothesis is the idea that the conversational exchanges that characterize human communication have an eminently narrative dimension, and that the discursive character of language is the key to investigate its nature, its functioning and its origin. In effect the discursive dimension of language represents the point of convergence between the

properties (and the processing devices) of language that govern its functioning and the properties (and the processing devices) which led to its origin. Following this, an important methodological indication emerges—the possibility to investigate the steps that have given rise to language through the study of its actual functioning. Following such an indication, the first part of this paper will focus on the analysis of the functioning of human communication.

### 3. From microanalysis to macroanalysis

Although it may seem an interpretative hypothesis distant in time and long ago overtaken by events, the “principle of formality” (the idea that the computational devices only have access to the form of representations) at the base of cognitive orthodoxy (Haugeland, 1997; Fodor, 1975) continues to have an enormous influence on models of human communication that, focused on the primacy of syntax, consider the constituent structure of sentences as the essential and distinctive feature of language.

At the basis of such an interpretive hypothesis is a general perspective about the nature of the mind, and in particular about the (propositional) nature of thought characterizing the great part of the philosophy of language of the twentieth century. We will not discuss the issues related to this hypothesis that are already well known. For the purposes of our argument it is sufficient to focus on two aspects of the perspective of language related to the principle of formality: the implicit (explicit, sometimes) adherence to the code model of communication (Shannon & Weaver, 1949), and the adherence to the idea that the sentence (and its internal constituents) represents the essence of human language. These two aspects are closely related. In fact, the code model is founded on the close parallelism between thought and language. According to Fodor (1975, 1987), this parallelism is a consequence of the fact that the logical structure of sentences is nothing but the product of the propositional structure of thought. Before considering the reasons that led us to abandon the linguistic perspectives inspired by cognitive orthodoxy, it is necessary to say a few words about the code model, a model that Fodor (1975, p. 106) considers «not just natural but inevitable».

### 3.1 Communication based on clues

The code model of communication seizes upon the idea that «we have communicated when you have told me what you have in mind and I have understood what you have told me» (Fodor, 1975, p. 109). According to this model, the thought (i.e. the message) is encoded by the speaker in a succession of sounds that the listener decodes to be able to share the thought (the message) that the speaker has intended to communicate. Perhaps because of its intuitive plausibility, the code model has been considered the model of communication for centuries, and these days it is at the center of theoretical reflection on the nature of language (consider Chomsky's Universal Grammar, just to name one famous case).

Despite its popularity, the code model is completely ineffective from an explicative point of view. The most compelling criticism of this model has been made by Sperber and Wilson (1986; 2002) with Relevance Theory (RT). Revisiting Grice's intuition (Grice, 1968) from a cognitive perspective about the distinction between "speaker's meaning" and "sentence's meaning", the two authors have strongly questioned the idea that communication is governed by a parallelism between what is said and what is thought. The case of figurative language is the most obvious example of the difference between the literal meaning of the expressions uttered by the speaker and the informational content that she intends to communicate.

Data in favor of RT come from language pathologies. Autism is widely used in studies of cognitive pragmatics as evidence in favor of the role of the speaker's intentions in the processes of language production and comprehension (Happé, 1993; Happé & Frith, 1995; Sperber & Wilson, 2002; Wearing, 2010). One of the prevailing ideas in the literature is that people with autism remain stuck on the literal meaning because of a specific deficit in the system of Theory of Mind (ToM), the cognitive device underlying the capacity for mentalization that allows humans to attribute mental states to others as well as to themselves (Baron-Cohen et. al., 1985; Happé & Frith, 1985; Happé 1993). From this perspective, the difficulties typical of subjects with autism related to the pragmatics of language depend on what Baron-Cohen (1995) called "mindblindness", the difficulty of these individuals to see people around them as endowed with mental states. Even though the idea that the communicative deficits of autistic people are simply caused by an

impairment of ToM has been revised and expanded (Frith & Happé, 1994; Tager-Flusberg, 2007), the relationship among autism, mindreading systems and RT has been confirmed recently (e.g., Wearing, 2010).

Questioning the code model, emphasizing the gap between what is said and what is meant, allows us to consider the idea that the communicative expressions are only clues that the speaker offers the listener to enable her to interpret his communicative intention. Considering linguistic expressions as clues (rather than as encoded signs) is of great interest for our discussion. In addition to allowing us to understand the functioning of language, the clues model of communication is an important starting point for the analysis of the origin of language. We will return to this topic in the last part of the paper. Before opening the *pars costruens* of our argument, we must still consider a second difficulty of the code model: the priority assigned to the sentence in the production-comprehension processes.

### 3.2 From sentence to discourse

We do not discuss here the fact that the constituent structure of the sentences represents an important aspect of language processing. The question we ask is whether an analysis focused only on what happens within sentences is a sufficient condition to account for the production-comprehension of human language. The case of subjects who, in spite of their ability to produce well-formed sentences, are unable to efficiently communicate, invites us to respond negatively to the question. The cases of patients suffering of forms of derailment (as schizophrenics) (Marini et al. 2008), or subjects unable to maintain the route of speech as people with brain injuries (Marini et al., 2011; see Adornetti, this volume), show that the ability to construct and understand well-formed sentences is not a sufficient condition for the production and comprehension of discourse. Cases of this kind pave the way to two considerations: first, discourse processing requires different principles (and different processing devices) from those governing the construction of sentences; second, and more important for our purposes, the sentence cannot be considered to be the essence of language.

At the basis of this paper is the concept that the most important issues for the analysis of language functioning and origin are related to properties and processes that regulate the relationship between sentences (macro-analysis),

rather than those that govern the relationship between internal constituents of the sentence (microanalysis) (see Davis et al., 1997; Marini et al., 2008). More specifically, our idea is that the origin and functioning of language have to do with global coherence, a pragmatic property that allows individuals to “keep the route” in discourse production. Central to our proposal is the notion that global coherence is primarily a property concerning cognition (and only secondarily, language) and that the processing of this property needs specific computational systems—although not specific for language. Given the importance of coherence in our hypothesis, the first step of our argument is to demonstrate coherence’s autonomy and its independence from other linguistic properties.

The first move in this direction is the defense of the non-reducibility of coherence to cohesion, that is the linguistic connections between consecutive statements (i.e. by means of anaphora or pronouns). In sharp contrast with the tradition that maintains that discourse coherence is dependent on the cohesive ties between sentences (e.g. Daneš, 1974; Halliday & Hasan, 1976), empirical evidence and theoretical arguments show that cohesion is neither a necessary nor sufficient condition of coherence (Giora, 1985). In this volume, the issue of the irreducibility of global coherence to cohesion is analyzed in the article of Adornetti (to which I refer for details): we consider the reasons proposed in this article compelling, and we will assume them going on in our argument. That said, before proceeding, it is necessary to emphasize the important role that the debate over cohesion *vs* coherence plays in the discussion of cognitive architectures.

The advantage of reducing coherence to cohesion is represented by reference to a unique processing system: if the external relations between sentences depend the analysis of some internal components of sentences, then the device used for microanalysis also is able to account for macroanalysis. Having shown that discourse coherence cannot be interpreted in terms of cohesion therefore implies a reference to different cognitive systems than those involved in microanalysis. To explain the nature of these devices is a key point of our theoretical proposal. Before taking into account the issue of cognitive architectures, however, it is necessary to consider a second aspect of the conceptual debate on global coherence. This is the “friendly fire” of the proponents of RT: according to Sperber and Wilson, although not reducible to cohesion, coherence is reducible to relevance. Given the importance that the clues model of communication and RT play in the theoretical framework of this

paper, the attack that the two authors make on the autonomy of coherence deserves to be carefully analyzed.

#### 4. Extending Relevance Theory

##### 4.1 Why Relevance is not enough

At the basis of our proposal is the idea that the origin of language and its functioning are tied to a perspective in which the clue model fits with the discursive nature of human communication. From this perspective, the first step is to ask if coherence is reducible to the principle of relevance. Sperber and Wilson are explicit in this regard: since relevance is the (unique) principle through which any aspect of human language can be analyzed, even discourse coherence must be interpreted in reference to this principle (Sperber and Wilson 1986/1995, p. 289; Wilson, 1998). Giora (1997, 1998) criticized Sperber and Wilson's thesis. In her view, in fact, relevance is not able to account for any aspect of language. Specifically, when we move from the sentence level to the discourse level, relevance is not a criterion that can explain global coherence. In support of her hypothesis, Giora shows examples of expressions that respect the principle of relevance, but are incoherent in terms of their relation to each other, and examples of expressions that are coherent, even if they are not relevant in the context. A useful example for demonstrating Giora's position concerns the situations in which it is possible to distinguish between different *degrees of coherence*. Consider the cases of (1a) and (1b):

- (1a) This first time she was married her husband came from Montana. He was the kind that when he was not alone he would look thoughtful. He was the kind that knew that in Montana there are mountains and mountains have snow on them. He had not lived in Montana. He would leave Montana. He had to marry Ida and he was thoughtful (taken from Ida by Gertrude Stein).
- (1b) The first time she was married her husband came from Montana. He was the kind who loved to be alone and thoughtful. He was the kind who

loved mountains, and wanted to live on them. He loved Montana. But he had to marry Ida and leave Montana (Giora, 1997, p. 26).

Giora's interpretation is that the difference in coherence between the two discourses cannot be explained in terms of "contextual effects weighed against processing effort", that is in reference to the principle of relevance as conceived by Sperber and Wilson. In fact, according to Giora, as the two discourse are both relevant, «they nevertheless differ drastically in terms of coherence: (1b) is more coherent than (1a)» (Giora, 1997, p. 26). Because RT is unable to explain the coherence differences between (1a) and (1b), the general conclusion to be drawn from these considerations is that «coherence is not a derivative notion» (Giora, 1997, p. 22). Good news for the autonomy of coherence.

What makes the evaluation of coherence specific? Giora's thesis is that narrative comprehension relies on the identification of the causal links governing the relationship between segments of discourse: in order to judge a discourse coherent or incoherent (or not fully coherent), in other words, it is necessary to refer to the concept of well-formedness. We will not go into the details of the debate between Giora and Wilson. The issue that needs to be highlighted for our purposes is that the argument used by Wilson (1998) to argue that the judgments of coherence are completely reducible to the judgments of relevance rests precisely on well-formedness analysis. Her opinion is that the question of well-formedness is not a matter of relevance because RT is a theory of comprehension, and the issues relating well-formedness concern judgments and evaluations not involved in the psychological processes of comprehension. According to Wilson, in fact, the sense of incoherence that listeners feel with the fragments of discourse presented by Giora must be interpreted with reference to:

the manifest waste of effort spent in looking for relevance in the wrong direction, or failing to find it at all. The resulting interpretations will be inconsistent with the principle of relevance, *whether or not the discourse segments are related*. By the same token, what makes a discourse (...) acceptable in the circumstances described is the fact that it has an interpretation consistent with the principle of relevance, *whether or not the discourse segments are related* (Wilson, 1998, p. 67 italics added)

That being the case, all you need to process a discourse is a system capable of grasping the communicative intention of the speaker, independently of how the speaker organizes the segments sequence in the communication flow. The explanatory weight of the clues model is entrusted to the study of the processing systems involved in the effort to looking for relevance. In favor of RT there is a model of cognitive architecture perfectly suited to the theory: if all you need to account for discourse coherence is relevance, then all you need to process coherence is a system of mindreading.

From Wilson's argument, it is clear that in order to maintain that coherence is reducible to relevance it is necessary to maintain that the way in which discourse segments are related can be excluded from the comprehension processes. Nevertheless, contrary to this argument, Ditman and Kuperberg (2007), in an article on the ability of schizophrenics subjects to project over time, showed that the problems these patients had in maintaining the coherence links across sentences was due to the fact that «building a coherent representation of discourse (...) requires the establishment of logical and psychological consistency between the events and propositions described in individual sentences» (Ditman & Kuperberg, 2007, p. 992). It is difficult to account for the psychological and logical consistency between events and propositions without referring to the causal relationships between the events narrated in a discourse and the segments of the speech used in the narrative to express them. The emphasis that Giora puts on the issue of well-formedness fits with the idea that the way to organize the temporal sequence of expressions (to put in the correct way the segments of the speech in the narrative flow) plays a decisive role in the inability of schizophrenics to build a coherent representation of discourse. It goes well with this idea especially because of the fact that well-formedness, invoked to explain discourse coherence, is not (given the independence of coherence from cohesion) a property that linguistic grammar imposes on thought - coherence is a property of discourse because well-formedness is primarily a property of the flow of thoughts. That said, since it is difficult to argue that an essential character of the flow of thoughts is not implicated in comprehension processes, the criticism toward well-formedness is not enough to justify the reducibility of coherence to relevance, despite what Wilson is willing to recognize. So much for the question of the properties of language. It is time to take into account the issue of cognitive architectures.

#### 4.2 Why mindreading is not enough

One major difficulty in Giora's model (on this point Wilson is absolutely right) is that she does not have a proposal regarding the specific cognitive systems involved in the discourse processing. In our opinion, the litmus test of the dispute on the reducibility of coherence to relevance concerns the cognitive architectures: since RT focuses on a single processing system (mindreading), two questions have to be addressed for the purposes of our discussion. Regarding the actual functioning of language: Is a mindreading system sufficient to account for the human ability to process discourse? Regarding the origins topic: Is a mindreading system sufficient to account for the transition from animal communication to human language?

What we want to discuss here is not whether a mindreading system has a role in language processing: the point to be discussed is whether the mindreading system is really the only biological adaptation (Tomasello, 1999) at the base of the origin and functioning of language. Sperber and Wilson (1986, 2002; but above Sperber, 2000 and Origgi & Sperber, 2000) agree with Tomasello. From their point of view, the thesis that relevance is the only explanatory principle of language fits with the idea that mindreading is the only system underlying our communication skills.

From our point of view the processing systems required to explain language have to draw together the clues model of communication with the discursive foundation of language governed by coherence. In this perspective the problem of the models based on mindreading is that, by focusing exclusively on the interpretation of the speaker's intention, they leave out the discursive foundation of human communication. One way to try to understand how to put together the clues model with the narrative foundation of language is to ask for what kind of systems are required to process discourse coherence. The metaphor of the flow of discourse given by the ability to "keep the route" in conversation is the starting point of the *pars costruens* of our article. At the basis of our hypothesis, in fact, is the idea that coherence is guaranteed by the systems that allow humans to navigate through space and time.

### 5. Navigational communication

Our starting point here is a quote from Chafe (1987, p. 48), taken by Wilson in her dispute with Giora. Wilson is right to claim that:

[D]iscourse is best approached in terms of process than structure: “it is more rewarding, I think, to interpret a piece of discourse in terms of cognitive processes dynamically unfolding through time than to analyze it as a static string of words and sentences” (Wilson, 1998, p. 70).

We agree with this perspective, provided, however, one seriously considers the “unfolding through time” of the computational processes involved in the flow of discourse. In fact, in discourse processing the projections back and forward in time (monitoring what the speaker says with respect to what she has already said and anticipating what the speaker will say) are at the basis of the construction of the route in the flow of speech. Now, in spite of the emphasis that Wilson reserves in the Chafe quote, the theory of relevance is not equipped in terms of cognitive architectures to account for processes involving the temporal plane. For this purpose, the system of mindreading is not a sufficient condition: what we need is *Mental Time Travel* (MTT), a cognitive system that enables an individual «to mentally project themselves backwards in time to re-live, or forwards to pre-live, events» (Suddendorf & Corballis 2007, p. 299; Corballis, 2011).

There are two things to consider in regard to MTT.

The first follows what Gärdenfors (2003, 2004, see also his article in this volume) argues with respect to the way Tomasello (1999) addresses the question of the origin of human communication: in the same way in which the speaker’s communicative intention is a necessary but not sufficient condition to explain human language functioning, the mindreading device is a necessary but not sufficient condition to process language. In Gärdenfors’ (2003, 2004) and Osvath and Gärdenfors’ (2005; Gärdenfors & Osvath, 2010) opinions, to account for the origin of language, it is necessary to access the human capacity to anticipate the future (anticipatory cognition): from this perspective, language is the product of cognitive systems that allow individuals to break away from the here and now of the present situation in order to plan future goals. At the root of human communication is a form of future-oriented

cooperation: without a specific system of temporal processing, human language would never have originated, nor would it work the way it works (Cosentino, 2011; Cosentino & Ferretti, 2014; Ferretti & Cosentino, 2013).

The second consideration relates directly to the relationship between MTT and human narrative abilities. Neisser (2008) points out that remembering is much more akin to telling a story than to playing back a tape or looking at a picture. Corballis (2011, p. 111; see his paper in this volume) maintains that «the same constructive process that allows us to reconstruct the past and the construct possible futures also allows us to invent stories» (Corballis, 2011). Behind the considerations made by Gärdenfors and Corballis lies the first reason that Relevance Theory needs to be expanded: if language (and narrative abilities, specifically) call into question a navigation device in time, then mindreading cannot be considered a unique processing system at the foundation of human communication. But there's more.

In addition to underlie the relationship between MTT and human narrative abilities, Corballis also makes a more general statement of great relevance to our discussion. In his view, in fact, time navigation is closely related to space navigation:

Several of the critical properties of language, then, probably evolved from the relaying of events, whether past, present, future or fictional, most of them located at times other than the present, including imaginary time. Other times also means other places, since we are peripatetic creatures, restlessly moving about the planet – and occasionally off it (Corballis, 2011, p. 114).

Birds' migration is the most intuitive and clear case demonstrating the close connection between time and space (Berthold, 2000). The same thing applies to humans:

Despite every navigator's preoccupation with distances and angles, latitudes and longitudes, and headings and compasses, nearly all human navigation rests on a basis of understanding and measuring time. We need to know when to start, what direction to choose relative to the sun or stars, how long we've been moving if we wish to compute distance travelled, when to stop, or that most challenging task of all, how to determine relative time so we can deduce longitude (Gould and Gould, 2012, pp. 36-37).

The close link between space and time representation is well demonstrated by brain anatomy (Corballis, 2013). The discovery of place cells allowed O'Keefe and Nadel (1978) to argue that the hippocampus is the basis of spatial cognition in rodents and is the substratum for episodic memory of humans (Dudchenko, 2010). Having said that, there is also a relationship of priority between space and time. In effect, navigation in space has logical and temporal priority over navigation in time; that is to say, the ability to project oneself in time is based on the ability to project oneself in space. When one looks at the coherence of discourse as a phenomenon linked to the ability to build and maintain the route toward a goal, it is primarily the spatial navigation metaphor to which we look.

Speaking of spatial navigation, in this paper we analyze almost exclusively the distinction between processes that occur in the head (inner navigation) and processes (route navigation) that occur during actual navigation on the ground (Yoder, Clark, & Taube, 2011). It is a known and empirically verified fact that route-based representations are dependent on brain structures (i.e., lingual gyrus, calcarine cortex, fusiform gyrus, parahippocampal cortex) that are different from those implicated in the use of mental maps (Epstein, 2008). Very roughly, a distinction from inner and outer navigation enables us to distinguish between internal processes as the sense of direction and the construction of a mental map and external processes such as maintenance and realignment of the route during effective navigation. Such a kind of distinction is at the foundation of the tight relationship between spatial navigation and discourse production.

### 5.1 Inner navigation

Spatial navigation represents, even intuitively, a good metaphor for thinking about the processes at the foundation of discourse. Lewis' definition of a system of navigation (1994, p. 82), for which «the first requirements (...) is to enable the voyager to take his departure and continue towards his objective in the right direction», and Gallistel's (1990) idea for which navigation is «the process of determining and maintaining a course or trajectory from one place to another» both illustrate our perspective. The ability to maintain a trajectory in the right direction is a core component of the process involved in

approaching a destination. Indeed, in order to reach the expected destination, one needs to keep the intended route and overcome geographic obstacles. In a very similar way, the process of discourse construction also relies on the ability to identify a goal (the content the speaker intends to convey to the listener) and to construct the route and to maintain the right trajectory to express it. Like navigation in space, the flow of communication is strongly linked to difficulties in maintaining the course to reach a given destination. In fact, in the same manner as in space navigation, the achievement of the communicative goal depends on the continuous realignments implemented by speakers to rebuild the route in the face of continual digressions imposed by the different points of view typical of verbal communication (Ferretti et al., 2013). Building the route and maintaining the right trajectory to the goal is equivalent, in narrative terms, to building and maintaining the global coherence of discourse. What kind of evidence can we offer to justify the involvement of navigation systems in the processing of discourse coherence?

Recent neuropsychological and neuroimaging evidence supports the hypothesis of a connection between the systems required for spatial navigation and narrative processing in humans. For example, in Marini et al. (2008), during a story description task a group of schizophrenic participants produced derailments and errors of global coherence. This result is particularly interesting when we consider that schizophrenic patients were found to be impaired on a task in which they were required to learn their way through a virtual park rich with navigationally relevant landmarks. Similar disturbances in spatial navigation and narrative discourse processing also have been reported in persons with different aetiologies. According to Marini et al. (2010), Williams syndrome affects the narrative aspects of language. This result is particularly important given that until not long ago, the language of Williams syndrome sufferers (including discourse) was supposedly to be completely spared. Even more interesting is that Williams subjects have been reported to suffer from a severe deficit of reorientation (Lakusta et al., 2010). Even though in their 2010 paper Marini et al. did not consider the potential relationship between Williams subjects' navigational difficulties and their discourse problems, in a more recent analysis of the topic, the viability of the interpretation of Williams narrative difficulties in terms of navigational problems was openly acknowledged (Ferretti et al., 2013).

The reference to the construction of mental maps and all the processes of planning and control of one's path to a goal (O'Keefe & Nadel 1978; Spiers &

Maguire, 2006) is certainly the most intuitive way to relate spatial navigation and discourse analysis to the principle of coherence (Ferretti & Adornetti, 2011; Ferretti et al., 2013). That said, both in spatial navigation and discourse processing the ability to build the route and maintaining the right trajectory to the goal does not uniquely depend on information processing inside the mind. To figure out which kind of processing is further needed, we have to shift the focus from inner navigation to route-navigation.

## 5.2 Route-navigation

When we move in space, we are always looking for identified points on the ground (e.g., a church with a cross, a rock with a particular shape, a tree struck by lightning in a forest) from which we draw crucial information about our location in space and whether we are following the correct route. In fact, one of the main causes of disorientation in space is the absence of landmarks, while conversely «people typically don't get lost when in the presence of familiar landmarks» (Dudchenko, 2010, p. 66). From a neuroscientific perspective, the inability to recognize the reference points in the external world is evidenced by the “landmark agnosia” (Barrash, 1998), a specific case of “topographical disorientation” (Aguirre & D'Esposito, 1999). The role of landmarks is so important for navigation in space that even insects use them: the case of the digger wasp is one of the most studied in this respect (Collett & Collett, 2002; Duriev et al., 2003).

What makes landmarks so valuable in navigation is what Jonsson (2002) defined as their “magnetic power”. In a very similar way, Nemmi et al. (2013) considered the reference points in the external world as “beacons” that attract the attention of the traveller. The idea of landmarks as beacons exemplifies the function of anchoring to the context provided by these important (external) points of reference. The landmarks (intended as the clues on the ground through which the traveller “finds confirmation” of the planned route in mental maps) are the tools that allow travellers to assess the consonance between the chosen route and the actual walking.

In *The Art of Memory*, Yates (1966) described the loci method used by ancient Greek and Roman orators to maintain their route of discourse in public debates. This method makes extensive use of the metaphor of navigation and, in particular, of the construction of (the mental representation of) specific

spaces along the route of which attractiveness plays a central role in the construction of the flow of discourse. We posit that the ancient rhetoricians offer us extraordinarily effective insights to investigate the role of landmarks in the mental processes of comprehension and production of discourse. It seems plausible that speakers’ sense of route consonance and the continuous realignments and revisions in expressing their communicative intentions in the flow of speech rely on points of support characterized by a strong magnetic power analogous to what happens in route-based navigation. The construction of scenes (Hassabis & Maguire, 2007) in the critical points of the flow of speech fulfils the same function of landmarks in actual navigation (Ferretti et al., 2013).

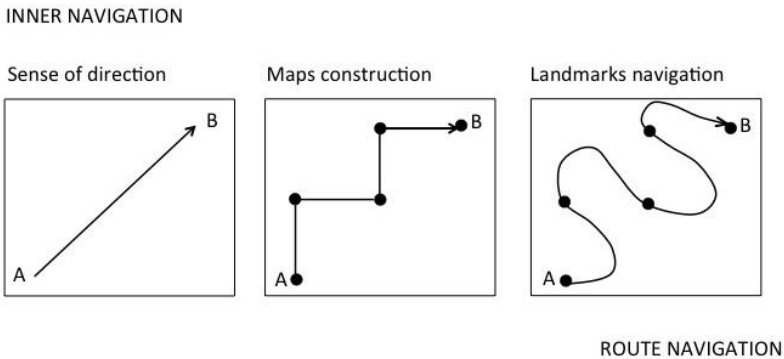


Figure 1  
Landmarks as point of convergence between inner navigation and route navigation.

As the relationship between the inner navigation and the human narrative abilities represents the general conceptual background of our discussion, the landmark navigation adds a further crucial step to our proposal. Looking at the landmarks in terms of the point of convergence between inner navigation and route-navigation allows us to look at the human narrative capacities in terms of the convergence between the clues model of communication and the discursive foundation of language (see Figure 1). If the identification of the goal and the recognition of the right direction to follow are useful to give an account of some aspects of discourse coherence, the landmark navigation can be profitably used to support the clues model of communication. In effect, the

specificity of such a model is closely tied to the episodic character of the expressive clues: any communicative expression of this kind is just a “prop” that occasionally occurs at a crucial point of a conversational exchange. From this point of view, expressive clues are a sort of “external scaffolding”, to use Clark’s (1997) terminology, through which communicators take advantage of external supports to proceed in conversation without an overly burdensome commitment in terms of elaboration processes. And it is exactly the episodic character of expressive clues that allow us to stress the close similarity between communication and route-navigation, as landmark navigation «is an episodic process» and «for landmark navigation to be accurate, one only needs to refer to the landmarks occasionally» (Yoder, Clark, & Taube, 2011, p. 561). Similar to actual navigation, expressive landmarks are useful but rather in specific and strategic points of discourse—similar to how we prop a conference with the images of power point. What evidence do we have in favour of the idea that the expressive clues used in conversational processes can be considered analogous to the use of landmarks in spatial navigation to build the right route toward a goal?

The first answer to this question comes from neuroscience. Ciaramelli (2008) reported the case of LG, a person with brain lesions at the ventromedial prefrontal and rostral anterior cingulate cortices, who invariably lost his way whenever asked to go somewhere on his own. Interpreting LG’s spatial difficulties, Ciaramelli maintained that «when travelling along routes that included a location he had attended frequently in the past, LG was ‘attracted’ to the familiar location, and failed to reach the goal location» (Ciaramelli, 2008, p. 2103). The hypothesis of landmarks as a magnet and beacon that we mentioned before is empirically confirmed by this study. More interesting for us is that Ciaramelli established a direct connection between becoming lost in space and becoming lost in thought and language. Specifically, she suggested that LG’s spatial disorientation (because of the attractive power of landmarks) involved a form of linguistic disorientation interpretable in terms of confabulation.

The second answer appeals to a different case. An interesting example to look at the role of the landmarks in the discourse construction concerns aspects of Inuktitut, one of the languages of the Barren Inuit—individuals that, because of the absence of natural landmarks in the environment around them, are continuously looking for reference points to correctly orient themselves in space (Kleinfeld, 1971, p. 132). The orientation difficulties imposed by the

environment to these individuals strongly constrains their language: for our purposes, the most interesting fact is that in communicating with each other Inuit *cannot help but* describe the location and orientation of the objects of which they speak. In this respect, an important aspect to underlying is the “contract” form (a form that fits well with the episodic character of the clues model) imposed by the “obligatory localizers” to the communicative expressions. In fact, as Kleinfeld (1971) maintains:

Adapted to the requirements of arctic ecology, the Eskimo language codes the domain of form and location with much greater economy than the English language (Gagne, 1968). For example, Gagne (1968) points out that the three-word Eskimo sentence “*ililavruk manna ilunga*” would be translated into the twenty word English sentence “Please put this slender thing over there cross-wise on that end of that slender thing to which I’m pointing” to convey the same amount of information about form and location (ivi, p. 134).

Ellard (2009), in agreement with what we have argued in this paper, asserts that the use made by the Inuit trekkers of «naming landmarks and embedding them into stories», is analogous to the way in which the digger wasps use landmarks to orient themselves in space. With an important difference, however. Unlike in the case of insects, an Inuit explorer can «shift from the use of space and geometry to navigate long distance to one based on a mental landscape of words, stories, and ideas» (Ellard, 2009, p. 40). The stories told by the Inuit make a vital contribution to the orientation of individuals in an environment poor of landmarks on the ground. That said, the ability to organize space navigation through the storytelling is not the only story to tell: now it should be clear at this point of our discussion that no storytelling would be possible if spatial representation (spatial navigation, specifically) does not enter strongly in the structural organization of the stories. In our opinion, it is possible to interpret the massive use of obligatory localizers coded in Inuktitut as a living fossil of the earliest forms of expression of all human languages: the obligatory localizers that characterize the language spoken by the Inuit, in other words, are what all languages likely had at the beginning of human communication and that the Inuktitut have maintained due to the special ecological conditions in which the Inuit continue to live.

## 6. At the beginning, to conclude

The merit of the clues model of communication is not only to engender the possibility of producing a convincing perspective of human language functioning, but it also to offer an interpretive outlook on the origin of our communication skills. First of all, because of the fact the clues model is able to offer a plausible explanation of how to cope with the (well-known) difficulties that affect the code model of communication with respect to the issue of origins. It is no coincidence that, from Saussure to Chomsky, the question of the origin of language has always been considered a type of reflection not worth wasting time upon: the argument underlying such a kind of reflection is that linguistic communication presupposes a shared expressive code and that such kind of code is an entity type that is given all at once or that it is not given at all. The clues model of communication allows obviating this kind of problem for the simple fact that it does not require a shared code to function.

Entrusting the origin of language to the clues model of communication means to entrust the explicative weight of the beginning of human communication to processing systems that are able to interpret expressive clues in terms of evidences of communicative intentions of the speaker. Sperber (2000; Origgì & Sperber, 2000) argues that the transition from animal communication to human language coincides with the transition from the code model to the clues model of communication governed by a mindreading system. The idea that mindreading plays a role in the origin of language is widely shared (e.g., Tomasello, 2008; Gärdenfors 2003; Seyfarth, Cheney, & Bergman 2005), and it is not questioned here. The point to be discussed here is whether the mindreading device is a sufficient condition to ensure the transition from animal communication to human language.

By the arguments brought forward to this point it should be clear that, as the narrative structure of the language cannot be interpreted in exclusive reference to the intentions of the speaker, mindreading cannot be considered the only device at the foundation of the origin of human communication. Language is strongly characterized by properties (the discursive coherence) and processing systems (the projection in space and time devices) that are very difficult to consider implicated in the functioning of communication without considering that they also are implicated in the origin of our communicative skills (see Ferretti & Adornetti, 2014). This type of analysis leads us to

consider the issue of the origin of language in reference to the protodiscursive foundation of human communication (see also Ferretti, 2013).

It is in reference to a perspective of this kind that the intent of linking the clues model of communication with the narrative perspective of language shows its explanatory power not only in reference to the issue of language functioning, but also in reference to the issue of origins. It is only through the projections in space and time, in fact, that the expressive clues produced by our ancestral relatives earn a significant distinction from animal communication. In addition to being an evidence of the communicative intention of the speaker, in fact, the expressions used in the early stages of protodiscursive communication were characterized by the projection of the clues in the narrative flow governed by coherence. If the ability to maintain the route in navigation can be seen as the condition for the construction of the flow of discourse in human communication, we have good reasons to think that the clues model (and the mindreading system strictly tied to it) must seek an ally in the navigation systems in space and time. From this order of considerations follows that the transition from the code model to the clues model is not a sufficient condition to ensure the transition from animal communication to human language. Then, we can conclude that the reasons we used to maintain that RT has to be extended in order to account for the functioning of language are the same reasons that lead us to sustain that RT needs to be extended and integrated also in order to explain the origin of language.

### Conclusions

Without spatial and temporal navigation systems we would not be able to maintain the route of discourse coherence: without discourse coherence human language would be very different from how we know it today. If coherence is among the essential features of language, then it is plausible to speculate that the origin of human communication skills must have involved processing systems which included, in addition to mindreading, projection systems in space and time.

Since its inception, human communication has been strongly characterized by a narrative structure. The opportunity to transform animal communication into language was made possible by the extraordinary processing power of cognitive systems available to humans. The “sewing machine” used to construct

narrative paths through the poor expressive clues used in the earliest forms of communication is a macro cognitive device based on the conjoint functioning of three different projection systems (Mental Time Travel, Mental Space Travel and Mental Mind Travel) exaptated in order to make communication more effective than that provided by the code model. Through the projective (and visionary) power of the macro cognitive device, humans began telling stories, an ability the origin of which, as should be clear at this point, coincides with the origin of language itself. The ability to tell stories as poems or dramas (as works of verbal art) is likely beyond the conception of narrative as we used it in this paper (see Collins, 2008, 2013). But verbal art is not something qualitatively different from typical human communication. From our point of view, poetry is just the product of evolution, certainly more sophisticated and culturally articulated, of the intrinsic discursive nature of human language.

## REFERENCES

- Adornetti, I. (2014). Making Tools and Planning Discourse: the Role of Executive Functions in the Origin of Language. *Humana.Mente – Journal of Philosophical Studies*, 27.
- Aguirre, G. K., & D’Esposito, M. (1999). Topographical disorientation: a synthesis and taxonomy. *Brain*, 122(9), 1613-1628.
- Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. Cambridge: MIT Press.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37-46.
- Barrash, J. (1998). A historical review of topographical disorientation and its neuroanatomical correlates. *Journal of Clinical and Experimental Neuropsychology*, 20(6), 807-827.
- Berthold, P. (2000). *Vogelzug. Eine aktuelle Gesamtübersicht*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Bloom, P. (2010). *How Pleasure Works*. New York: Norton.
- Boyd, B. (2009). *On the Origin of Stories. Evolution, cognition and fiction*. Cambridge: The Belknap Press of Harvard University Press.

- Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse*, Amsterdam: Benjamins, 21-51.
- Ciaramelli, E. (2008). The role of ventromedial prefrontal cortex in navigation: A case of impaired wayfinding and rehabilitation. *Neuropsychologia*, 46(7), 2099-2105.
- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge: MIT Press.
- Collett, T. S., & Collett, M. (2002). Memory use in insects visual navigation. *Nature Reviews Neuroscience*, 3, 542-552.
- Collins, C. (2008). Paleopoetics: Prefatory Notes Toward a Cognitive History of Poetry. *Cognitive Semiotics*, 2, 41-64.
- Collins, C. (2013). *Paleopoetics: The Evolution of the Preliterary Imagination*. New York: Columbia University Press.
- Corballis, M. C. (2011). *The Recursive Mind. The origins of human language, thought and civilization*. Princeton: Princeton University Press.
- Corballis, M. C. (2013). Mental time travel: a case for evolutionary continuity. *Trends in cognitive sciences*, 17(1), 5-6.
- Corballis, M. C. (2014). The gradual emergence of language. *Humana.Mente. Journal of Philosophical Studies*, 27.
- Cosentino, E. (2011). Self in time and language. *Consciousness and cognition*, 20(3), 777-783.
- Cosentino, E., & Ferretti, F. (2014). Communication as navigation: A new role for consciousness in language. *Topoi*, 33(1), 263-274.
- Dautenhahn, K. (2002). The origins of narrative: In search of the transactional format of narratives in humans and other social animals. *International Journal of Cognition and Technology*, 1(1), 97-123.
- Daneš, F. (1974). *Functional sentence perspective and the organization of the text*. In F.Daneš (ed.), pp. 106-128.
- Davis, G. A., O'Neil-Pirozzi, T. M., & Coon, M. (1997). Referential cohesion and logical coherence of narration after right hemisphere stroke. *Brain and Language*, 56(2), 183-210.
- Ditman, T., & Kuperberg, G. R. (2007). The time course of building discourse coherence in schizophrenia: An ERP investigation. *Psychophysiology*, 44(6), 991-1001.

- Dudchenko, P.A (2010). *Why People Get Lost: The Psychology and Neuroscience of Spatial Cognition*. Oxford: Oxford University Press.
- Ellard, C. (2009). *You are here: Why we can find our way to the moon, but get lost in the mall*. Random House LLC.
- Epstein, R.A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, 12, 388–396.
- Frith, U., & Happé, F. (1994). Autism: Beyond “theory of mind”. *Cognition*, 50(1), 115–132.
- Ferretti, F. (2013). Navigation, Discourse and the Origin of Language. In Brambilla P., & Marini A. (Eds.), *Brain evolution, language, and psychopathology in schizophrenia*, New York: Routledge, 22-32.
- Ferretti, F. & Adornetti, I. (2011). Discourse processing and spatial navigation. In B. Kokinov, A. Karmiloff-Smith, & N. J. Nersessian (Eds.). *European Perspectives on Cognitive Science*, Sofia: New Bulgarian University Press. Available online at <http://nbu.bg/cogs/eurocogsci2011/proceedings/pdfs/EuroCogScipaper251.pdf> (accessed 15 September 2014).
- Ferretti, F., & Adornetti, I. (2014). Against linguistic Cartesianism: Toward a naturalistic model of human language origins and functioning. *Language & Communication*, 37, 29-39.
- Ferretti, F., Adornetti, I., Cosentino, E., & Marini, A. (2013). Keeping the route and speaking coherently: the hidden link between spatial navigation and discourse processing. *Journal of Neurolinguistics*, 26(2), 327-334.
- Ferretti, F., & Cosentino, E. (2013). Time, language and flexibility of the mind: The role of mental time travel in linguistic comprehension and production. *Philosophical Psychology*, 26(1), 24-46.
- Fodor, J. A. (1975). *The language of thought*. Cambridge: Harvard University Press.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge: MIT Press.
- Gagne, R. C. (1968). Spatial concepts in Eskimo Language. In V. Valentine, & F. Vallee (Eds.), *Eskimo of the Canadian Arctic*. Toronto: McClelland and Stewart limited, 30-38.
- Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: Bradford/MIT Press.

- Gärdenfors, P. (2003). *How Homo Became Sapiens: On the Evolution of Thinking*. Oxford: Oxford University Press.
- Gärdenfors, P. (2004). Cooperation and the Evolution of Symbolic Communication. In D.K. Oller, & U. Griebel (Eds.), *Evolution of Communications Systems: A Comparative Approach*. Cambridge: MIT Press, 237-256.
- Gärdenfors, P. & Osvath, M. (2010). Prospection as a cognitive precursor to symbolic communication. In Larson, R. (Ed.), *Evolution of Language: Biolinguistic Approaches*, Cambridge: Cambridge University Press, 103-114.
- Gärdenfors, P. (2014). The Evolution of Sentential Structure. *Humana.Mente Journal of Philosophical Studies*, 27.
- Giora, R. (1985). Notes towards a Theory of Text Coherence. *Poetics Today* 6(4), 699-715.
- Giora, R. (1997). Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory. *Journal of Pragmatics*, 27(1), 17-34.
- Giora, R. (1998). Discourse coherence is an independent notion: A reply to Deirdre Wilson. *Journal of Pragmatics*, 29(1), 75-86.
- Grice, H. P. (1968). Utterer's meaning, sentence-meaning, and word-meaning. *Foundation of Language* 4 (3): 225-242.
- Gottschall, J. (2012). *The Storytelling Animal. How Stories Make Us Human*. Boston: Houghton Mifflin Harcourt.
- Gottschall, J., & Wilson, D. S. (Eds.). (2005). *The literary animal: Evolution and the nature of narrative*. Northwestern University Press.
- Gould, J. M., & Gould, C. G. (2012). *Nature's Compass. The Mystery of Animal Navigation*. Princeton and Oxford: Princeton University Press.
- Halliday, M. K, & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48(2), 101-119.
- Happé, F., & Frith, U. (1985). Theory of Mind and Autism. *Learning and Cognition in Autism*, 177-197.
- Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in cognitive sciences*, 11(7), 299-306.
- Haugeland, J. (Ed.). (1997). *Mind design II: philosophy, psychology, artificial intelligence*. Cambridge: MIT Press.

- Herman, D. (2013). *Storytelling and the Sciences of Mind*. Cambridge: MIT Press.
- Hirstein, W. (2005). *Brain fiction: Self-deception and the riddle of confabulation*. Cambridge: MIT Press.
- Jonsson, E. (2002). *Inner Navigation. Why We Get Lost and How We Find Our Way*. New York: Scribner.
- Kleinfeld, J. (1971). Visual memory in village Eskimo and urban Caucasian children. *Arctic*, 132-138.
- Lakusta, L., Dessalegn, B., & Landau, B. (2010). Impaired geometric reorientation caused by genetic defect. *Proceedings of the National Academy of Sciences*, 107(7), 2813-2817.
- Lewis, D. (1994). *We, the Navigators. The Ancient Art of Landfinding in the Pacific*. Second Edition. Honolulu: University of Hawaii Press.
- Marini, A., Galetto, V., Zampieri, E., Vorano, L., Zettin, M., & Carlomagno, S. (2011). Narrative language in Traumatic Brain Injury. *Neuropsychologia*, 49, 2904-2910.
- Marini, A., Martelli, S., Gagliardi, C., Fabbro, F., & Borgatti, R. (2010). Narrative language in Williams Syndrome and its neuropsychological correlates. *Journal of Neurolinguistics*, 23(2), 97-111.
- Marini, A., Spoletini, I., Rubino, I. A., Ciuffà, M., Bria, P., Martinotti, G., Banfi, G., Boccascino, R., Strom, P., Siracusano, A., Caltagirone, C., & Spalletta, G. (2008). The language of schizophrenia: An analysis of micro and macrolinguistic abilities and their neuropsychological correlates. *Schizophrenia Research*, 105(1), 144-155.
- Neisser, U. (2008). Memory with a grain of salt. In H.H. Wood & A. S. Byatt (Eds.), *Memory: an anthology*, London: Chatto and Windus, 80-88.
- Nemmi, F., Piras, F., Péran, P., Incoccia, C., Sabatini, U., & Guariglia, C. (2013). Landmark sequencing and route knowledge: an fMRI study. *Cortex*, 49(2), 507-519.
- O'keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Origg, G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the Human Mind: Language, Modularity and Social Cognition*. Cambridge: Cambridge University Press, 140-169.
- Oswath, M. & Gärdenfors, P. (2005). Oldowan culture and the evolution of anticipatory cognition. Lund: *Lund University Cognitive Studies* 122.

- Pinker, S. (2007). Toward a consilient study of literature. *Philosophy and Literature*, 31(1), 162-178.
- Seyfarth, R. M., Cheney, D. L., & Bergman, T. J. (2005). Primate social cognition and the origins of language. *Trends in cognitive sciences*, 9(6), 264-266.
- Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: The University of Illinois Press.
- Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed), *Metarepresentations: A Multidisciplinary Perspective*. Oxford: Oxford University Press, 117-137.
- Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and cognition*. Cambridge: Harvard University Press.
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, 17(1-2), 3-23.
- Spiers, H. J., & Maguire, E. A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. *Neuroimage*, 31(4), 1826-1840.
- Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans?. *Behavioral and Brain Sciences*, 30(03), 299-313.
- Sugiyama, M. S. (2005). Reverse-engineering narrative: Evidence of special design. In Gottschall, J., & Wilson, D. S. (Eds.), *The Literary Animal: Evolution and the nature of narrative*, Northwestern University Press, 177-196.
- Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Current Directions in Psychological Science*, 16(6), 311-315.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT Press.
- Wearing, C. (2010). Autism, metaphor and relevance theory. *Mind & language*, 25(2), 196-216.
- Wilson, D. (1998). Discourse, coherence and relevance: A reply to Rachel Giora. *Journal of Pragmatics*, 29(1), 57-74.
- Yates, F. A. (1966). *The Art of Memory*. London/New York: Routledge and Kegan Paul.
- Yoder, R. M., Clark, B. J., Taube, J. S. (2011). Origins of landmarks encoding in the brain. *Trends in Neurosciences*, 34(11), 561-571.

Commentary  
**Origins of Human Communication**

Michael Tomasello  
The MIT Press, Cambridge, 2008

*Alessandra Chiera*<sup>†</sup>  
alessandra.chiera@uniroma3.it

Over the last years an ever-increasing number of works has turned attention to the topic of the human language evolution from several points of view (e.g., Burling, 2005; Corballis, 2002; Christiansen & Chater, 2008; Christiansen & Kirby, 2003; Deacon, 1997; Dunbar, 1996; Fitch, 2010; Knight et al., 2000; Pinker & Bloom, 1990; Tallerman, 2005; Wildgen, 2004). The common thread connecting these even different approaches is the research of the distinctive traits that enabled language appearance. It is however possible to distinguish between about two cornerstones of the thinking behind this enterprise: on the one hand, a group of scientists emphasizes the features that make language a unique ability of *Homo sapiens* and that cannot be interpreted in terms of skills shared with closely related animals; on the other hand, referring to a strictly Darwinian tradition, some scholars state that the essence of human language has to be investigated starting from the abilities which underlie both animal and human communication.

Noam Chomsky, the leading figure of the former discontinuist perspective, has highlighted the centrality of a specific component – Universal Grammar (UG) – at the core of the language faculty which represents a unique sudden endowment of our species completely autonomous from other cognitive systems. By virtue of this specialty inherent human beings, looking at the non-linguistic devices that are in common with other species appears totally worthless within an account of human language (Chomsky, 1988, 1996): the latter, in such a definition, is a human-only system.

On the contrary, the continuist perspective stresses the relationship between communication and other cognitive skills rejecting the idea that human language might have arisen from a single unexpected break in the

<sup>†</sup> Department of Cognitive, Educational and Cultural Studies, University of Messina, Italy; Department of Philosophy, Communication and Visual Arts, Roma Tre University, Italy.

evolutionary course. From this point of view, the animal kingdom instead represents a great test bench for pinpointing the prerequisites of language.

In his book *Origins of Human Communication* (2008), Michael Tomasello examines the evolutionary roots of human language and offers a theory consistent with such latter tradition, proposing that the fundamental elements of our communication system do not concern with a linguistic component isolated by other capacities but are rather traceable in general cognition. Already at the time of publication of his scathing review of Pinker's *The Language Instinct* (1994) - unsurprisingly entitled *Language Is Not an Instinct* (1995) - Tomasello had called the need for a specific language device into question. By taking into account the evidence brought by Pinker, the review highlighted their compatibility with a series of different models as much biologically founded as the UG hypothesis: in the specific Tomasello's view, the biological foundation of language was interpreted «just not in the form of specific linguistic structures preformed in the human genome» (Tomasello, 1995, p. 32) but in terms of general cognitive and social abilities «some of which are shared with other primates and some of which are uniquely human» (Tomasello, 2008, p. 208). The rejection of the Chomskian tradition is definitively accomplished in *Origins of Human Communication*:

It is not that the evolution of some kind of innate syntactic template such as universal grammar is impossible, it is just that currently there is no evidence for it empirically, no precise formulation of it theoretically, and no need for it at all – if the nature of language is properly understood (Tomasello, 2008, p. 315).

The main reason for renouncing UG and taking the side of pre-linguistic skills at the bottom of human language origin entails a simple but effective argument: considering the rise of human communication starting from a symbolic code means assuming a preexisting form of communication which is merely encoded (Tomasello, 2008, p. 58). But this approach falls into a clear fallacy: it takes for granted from the beginning something that must be explained. If our aim consists in accounting for human language, we cannot start from language but rather from how non-conventional encoded communication has been able to make inroads. That is, language is not linguistic in nature but relies on some kinds of more basic capacities. Thus, research in the pre-linguistic abilities exhibited by children and some

nonhuman animals may fruitfully inform a model of language evolution outlining the main skills that fostered it.

Put aside the thesis of an innate linguistic component which defines the speciality of human species, Tomasello suggests instead – on the base of studies in such areas – that some forms of *mental attunement* have to be considered foundational (Tomasello, 2008, p. 59). This burning move prepares the ground for an appealing model that emphasizes the pragmatics nature of language origins. Many authors have drawn attention to the appearance of communicative systems from a pragmatics point of view (e.g., Arbib, 2012; De Ruiter et al., 2010; Sperber & Origgi, 2010; Sperber & Wilson, 2002). In *Origins of Human Communication*, Tomasello develops a specific pragmatics account centered on the idea that communication is a kind of coordination problem born in social contexts. The proposal is worth considering as it is outlined.

Referring to the first insights of Paul Grice (1957, 1975), he hypothesizes a strong relationship between the cooperative structure typical of human communication and the cooperative structure of human social interactions. More specifically, communication can be considered to be a form of social interaction supported by cooperative interests and, lastly, as part of a biological adaptation for collaborative activities (Tomasello, 2008, p. 110). The main focus is interestingly on a capacity hold to be essential both in human cooperation and communication, namely the capacity of constructing common ground and joining the same attentional frame. Along with Tomasello, we consider the notion of common ground as a very key concept within a theory of human communicative systems. Already H.H. Clark (1996) has characterized communication as a joint activity which largely depends on the ability to keep common attention, to share the relevant background knowledge and joint experience in order to get the content across and make sense in the exchanges. The focal point about how we get to manage this common ground is that it takes a specific infrastructure at the bottom. The question of the underlying infrastructure represents a very settling topic.

Outside and within the paradigm which conceives communication as a form of interaction, there are many attempts to investigate this infrastructure. And of these, the so-called *Tacit Communication Game* (TCG) represents an interesting experimental study worth mentioning. The scholars (de Ruiter et al., 2010; Noordzij et al., 2010; Newman-Norlund et al., 2009) submit a communication task that involves the control of geometrical shapes on a grid

by the side of two subjects focusing on what they refer to as “interactional intelligence” (Levinson, 1995) namely the ability to convey and recognize each other’s intentions independently from a linguistic pre-established code. Looking at this capacity allows to identify the mental strategies used by people in order to construct a common ground and get across meaning when they do not have a common code. The results reveal some important data concerning the ability to communicate even in sub-optimal situations and to develop novel ways of interaction; more specifically, they highlight that communication is a strictly cooperative enterprise and this feature appears necessary to produce a code starting from poor means of expression. The hard work in nailing the TCG task indeed rests on the capacity of speaker and hearer to collaborate and progressively give each other’s feedbacks, moreover showed by neurophysiological analysis concerning the comparison of cerebral responses: the activated regions were the same during the planning phase of the sender and the comprehension phase of the receiver. According to the authors, the findings support the idea of a specialized communicative intelligence consistent with the cognitive architecture suggested by Sperber and Wilson (2002) that emphasize the specific role of a pragmatic module of Theory of Mind (ToM) in language functioning and origin.

Obviously Tomasello fully agrees with an approach that stresses the need of cooperative prerequisites for the origin of human communication, nevertheless, he suggests a different infrastructure at their base. The ability of attributing mental states to others represents, even in his model, a core competence for communicative systems but it is included in a domain-general device that entails *cooperative* motivations and that triggered shifts to shared activities driven by joint goals, that is, new inferential processes prerogatives solely for humans. Here is the critical point that needs to be taken into account more in detail before making some comments.

This device is a sole infrastructure of *Homo sapiens* because it appeared when human cooperative activities differentiated themselves from the other animal forms of cooperation. In other words, Tomasello suggests that humanity entailed the rise of a new collaborative species with truly cooperative and altruistic goals that have to be distinguished by the other animal ways of collaborating. Even great apes – our closest animal relatives – have social goals but not cooperative ones: the difference making difference consists in their motivations which, he points up, are deeply selfish, competitive and supported by *I-mode modality* contrary to the *we-mode modality* that drives human

motivations. According to the author, in order to comprehend human language what needs to be analyzed is such crucial step; actually, it accounts for a structural change in the nature of communication. When this new social-cognitive and emotional infrastructure arose, it spawned first forms of communication that initially had individualistic imperative uses – as requests involving mutual interests – but gradually fostered more cooperative interactions. The complex recursive mindreading sustained by the tendency to be helpful has, in this way, made the forms of collaborative activity more organized and consequently even «the mentalistic and altruistic structure» (Tomasello, 2008, p. 334) of human cooperative communication which rests on that scaffolding has been refined.

Tomasello is not denying that apes hold some essential mechanisms underlying the cooperative dimension of language; he acknowledges that they are able to understand that others have perceptions and aims, they feel forms of empathy, they hatch out plans in order to influence others' behavior for their benefit and exhibit a series of capacities typical of individual intentionality. What non-human primates miss, making them intellectually different from humans, is sharing intentions engaging with others in a truly cooperative way that involves the ability to be interchangeable and really disinterested. Without this further crucial step, language simply cannot come alive.

In what follows we will discuss those which are considered to be in our opinion two main criticisms of Tomasello's theses. The first one deals with the idea that an exclusive scenario of cooperation and altruism that deployed itself only within the rise of *Homo sapiens* may account for a continuist model of language as Tomasello holds to be true; moreover, tied to this doubt, some counter-arguments have been developed to strongly call the hypothesis of humans as the only cooperative beings into question.

The second criticism concerns more closely the cognitive architecture which is accorded the fundamental status in Tomasello's opinion: he claims that a general mechanism of mindreading represents the device through which addressing the language origins issue and more specifically accounting for a key switch that led to a new ability involved in mental attunement and sharing common ground, named the *bird's eye view*. We will claim that it is controversial that the ToM mechanism alone – irrespective of whether it is interpreted as a specialized device or a general one – could explain the pressures that fostered these essential features of human communication.

Concerning the question of *Homo sapiens* as the only cooperative species, Tomasello's position has been refined and strengthened several times shifting conclusions about unique features but it is quite clear: as already said, although apes show many complex cognitive and social abilities, there is something missing in their cooperative skills which is at the base of their inability to develop language. That this lacking competence marks a crucial difference between humans and other animals is often underlined by Tomasello so much that he entitled his article with Rakoczy (2003) *What Makes Human Cognition Unique* wherein they consider the shift from individual to shared and collective intentionality the core transition inherent humanity. And indeed, a point frequently stressed concerns the thesis that in order to account for human communication one doesn't ultimately look into animal interpersonal activities but rather into human cooperation which is «unique in the animal kingdom in many ways, both structurally and motivationally» (Tomasello, 2008, p. 6). To this extent, the only scenario we might be interested in is that of how the human lineage formed collaborative alliances and social groups. In other words, although a kind of psychological infrastructure needs to be in place before language could arise in humans, actually the cooperative character rising in human nature sets its communication systems apart from those of all other living species.

Why Tomasello arrives at conclusions so widely different from an approach previously stated to be truly continuist? The answer lies, according to him, in the experimental findings achieved with non-human primates who point out a fairly individualistic behavior opposed to the human eusociality, well manifest already in very little children. What makes humans *super cooperators* individuals (Nowak & Highfield, 2011) is the co-evolution of cognitive and cultural abilities (Tomasello, 2008, p. 354) wherein cumulative cultural learning plays a settling role. Many counter-arguments have been developed to contrast the idea that cooperation is an exclusive human skill. For example, de Waal has produced several pioneering works on the pro-social attitudes of non-kin primates, on their sense of fairness and emphatic feelings, showing moreover that apes are able to monitor interactions in which they participate keeping track of each individual's support to common aims (de Waal, 2009). Other authors have tested the collaborative capacities of primates as well. Hare and Kwetuenda (2010) have recently experimentally documented that bonobos exhibit altruistic behaviors preferring to share food with a conspecific individual rather than consuming it alone. Along this line, Boesch and

colleagues (2009, 2010) has founded that wild chimpanzees help neighboring conspecifics in flights against opponents and do care for the welfare of other unrelated group members, for instance adopting orphaned youngsters. Moreover, Greenspan and Shanker (2004) have shown that the bonobo (*Pan Paniscus*) Panbanisha possesses advanced representational abilities allowing him to use gestures in declarative and informative forms besides in imperative ones. On the whole, there is a large number of evidence in favour of the existence of mutualistic and altruistic behaviors in animals that highly weakens Tomasello's evolutionary story (Reboul, 2010).

Furthermore, in *Origins of Human Communication* the idea that the cooperative dimension of human cognition is dominant to competition while the latter represents the hallmark of all nonhuman interactions is taken for granted. To this extent, Tomasello ignores coercion and deception as critical evolutionary forces even in the cognitive evolution of humans. Against this position, many theories claim that complex skills evolved in a competitive scenario wherein the fight between cheaters and cheater detectors fostered new pressures – that is ascribable to the Machiavellian Hypothesis developed by Whiten and Byrne (1997). Moreover, Cavalli-Sforza (2010) has emphasized that group competition has characterized the evolution of *Homo sapiens* more than cooperation, making it the most overbearing ape.

Such considerations suggest that an exclusive competitive account of animal behavior as well as an exclusive cooperative description of human behavior must be blended. What comes to light is that the debate concerns the emphasis put to the similarities on the one hand and the differences with animals on the other hand. Tomasello, though addressing an evolutionary account and contrary to the early assumptions, stresses the aspects that in his opinion provide human uniqueness. Nevertheless, the considered findings suggest to set less clear boundaries between ape and human socio-cognitive skills (Ferretti & Adornetti, 2012). Moreover, the thesis of human speciality based on the role of cultural development evokes the dual inheritance approach carried on by Richerson and Boyd (2005) and the idea that when biology meets culture something totally changes – theses that are at least liable to suspicion from a Darwinian perspective; namely, Tomasello's model is only plausible to the extent that we support a discontinuist view of human nature.

Let us see the second criticism. Tomasello argues that a crucial phase in the evolution of the cognitive infrastructure supporting human language is represented by the emergence of an essential ability that is, creating common

ground. This critical dimension of human communication enabled the richness of meaning exchanges and the possibility to go beyond the ambiguity of reference, something that is absolutely unknown to any animal form of communication. It is not specified exactly how humans acquired this tool that allowed them to act together on a common ground but it is maintained that human culture played a leading role in fostering it. In this last part, we will state that an effective functional communication actually requires such capacity but it can be explained in fully biologically and continuist terms. More specifically, in our opinion this explanation is achievable provided that the early forms of mental attunement and cumulative common ground engaged on first communicative exchanges are interpreted as proto-conversational.

These critical features in fact arise only in the speech triangle context wherein speaker, listener and topic need to be aligned. It is here that common conceptual ground shared by communicative subjects has to be inferred, maintained and monitored in a situation in which it is not static but continuously dynamics and cumulative. Tomasello has never explicitly focused attention on the role of conversation in the origin of language, except in the brief response to commentaries written with Carpenter and colleagues (2005) where they highlighted that apes are not able to engage in proto-conversation. Because of this omission, his model falters in explaining the core element of human cooperative communication which rests on its dialogic nature (Pickering & Garrod, 2004). A theoretical account of the origins of human language that lies on conversational dimension may explain how mental attunement and keeping track of common ground have led to a key notion of Tomasello's overall theory, that is, the notion of *bird's eye view*. The ability to have a bird's eye view of the scene entails that an individual involved in a social activity as communication can understand the global scene from a neutral perspective. In this way, one can simultaneously engage in shared activities, in paying attention to the topic and understanding if the interlocutor is also paying attention to it. It is a kind of multiple dimension in which the subjects are able to act together and progressively increase their common ground. This does not imply that we converse performing cumbersome recursive computations such as "what would he think I would think if he..." (Clark & Marshall, 1981). Just in some situations, when something goes wrong and we have to ask for clarifications or realize repairing strategies, we explicitly experience the efforts required to achieve the greater level of sharing common ground, that is, the conceptual alignment between speaker and hearer (Garrod

& Pickering, 2009). Therefore, the bird's eye view makes this alignment possible; obviously, this kind of phenomenon is a late product of evolution but it represents the needle that from the beginning guides the communicative exchanges.

The next step involves the following question: is it enough an infrastructure centered on a ToM mechanism to account for the properties just considered? In Tomasello's view, the mindreading device sustaining *we-mode* cooperation represents all that is necessary. Nevertheless, the idea that a single device may explain the complex abilities that triggered human communication is controversial (e.g., Ferretti & Adornetti, 2014). It has been recently emphasized that although an approach focused on identifying the mechanisms underlying specific aspects of language elaboration has undeniable advantages, however, what characterizes human language is traceable in a more holistic perspective. A similar observation is even more likely considering the properties that we conceive with Tomasello as essential, especially from a conversational point of view. Consistent with this idea, many authors have highlighted the explicative value of reframing the question of language in terms of brain *network* (e.g., Fedorenko & Thompson-Schill, 2014; Novick et al., 2009; Ramachandran, 2012). To this extent, human language was triggered by cross linkages between different areas wherein functionally specialized and domain-general systems started to work together eliciting global connections. In this approach, the breakdown of devices involved in language elaboration would lose distinguishing features that appear only at the level of this widespread activation. For instance, the domain-general cognitive control system takes advantage of such network and seems to play a key role in language processing (Knudsen, 2007; Thompson-Schill et al., 2005). A speculative hypothesis might claim that such a kind of system has been critical in making the early poor means of expression well-articulated, fostering forms of explicit control and monitoring of the conversational setting.

Probably, this structured network that allows information to be broadcast even explicitly is a specific characterization of *Homo sapiens* that lacks in other non-human primates (de Winter & Oxnard 2001; Gazzaniga, 2008; Shea et al., 2014). Nevertheless, this consideration does not undermine the Darwinian framework: referring to the cognitive architecture underlying the switch from proto-conversational to effective interactive communicative exchanges as a matter of networking, rather than a rise of a unique endowment or a feature prompted by human culture, provides a truly evolutionary account. In fact,

what establishes the peculiarity of human communication does not rely upon a magic ingredient lacking in the animal kingdom but on the cerebral reorganization which functioning might have allowed new abilities as the bird's eye view.

Tomasello has contributed to add an important piece to the research but the puzzle regarding language origins and evolution requires still a very long way to be solved. The commitment of providing a comprehensive theory accounting for the many small steps that gradually led from the communicative skills of our closely related animals to the complexity of modern human language is a very living matter.

#### REFERENCES

- Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis*. Oxford: Oxford University Press.
- Boesch, C. (2009). *The real chimpanzee: sex strategies in the forest*. Cambridge: Cambridge University Press.
- Boesch, C., Bole, C., Eckhardt, N., & Boesch, H. (2010). Altruism in forest chimpanzees: the case of adoption. *PLoSOne*, 5(1), e8901.
- Burling, R. (2005). *The talking ape: How language evolved*. Oxford: Oxford University Press.
- Cavalli-Sforza, L. L. (2010). *La Specie prepotente*. Milano: Editrice San Raffaele.
- Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. Cambridge: MIT Press.
- Chomsky, N. (1996). *Language and mind: Current thoughts on ancient problems (part I)*. Paper presented at Universidad de Brasilia.
- Christiansen, M.H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, 31(05), 489-509.
- Christiansen, M.H., & Kirby, S. (Eds.). (2003). *Language evolution*. Oxford: Oxford University Press.
- Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H.H., & Marshall, C.R. (2002). Definite reference and mutual knowledge. *Psycholinguistics: critical concepts in psychology*, 414.

- Corballis, M.C. (2002). *From hand to mouth: The origins of language*. Princeton: Princeton University Press.
- De Ruiter, J.P., Noordzij, M.L., Newman-Norlund, S., Newman-Norlund, R., Hagoort, P., Levinson, S.C., & Toni, I. (2010). Exploring the cognitive infrastructure of communication. *Interaction Studies*, 11(1), 51-77.
- De Waal, F. (2009). *The age of empathy*. New York: Harmony.
- De Winter, W., & Oxnard, C.E. (2001). Evolutionary radiations and convergences in the structural organization of mammalian brains. *Nature*, 409(6821), 710-714.
- Deacon, T. (1997). *The symbolic species. The co-evolution of language and the brain*. New York: WW Norton & Company.
- Dunbar, R., (1996). *Grooming, gossip, and the evolution of language*. London: Faber and Faber Limited.
- Fedorenko, E., & Thompson-Schill, S.L. (2014). Reworking the language network. *Trends in cognitive sciences*, 18(3), 120-126.
- Ferretti, F., & Adornetti, I. (2012). *Dalla comunicazione al linguaggio. Scimmie, ominidi e umani in una prospettiva darwiniana*. Milano: Mondadori Università.
- Ferretti, F., & Adornetti, I. (2014). Against linguistic Cartesianism: Toward a naturalistic model of human language origins and functioning. *Language & Communication*, 37, 29-39.
- Fitch, W.T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.
- Garrod, S., & Pickering, M.J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292-304.
- Gazzaniga, M.S. (2008). *Human. The Science Behind What Makes Us Unique*. New York: Ecco.
- Greenspan, S.I., & Shanker, S.I. (2004). *The first idea. How Symbols, Language and Intelligence Evolved from our Primate Ancestors to Modern Humans*, New York: Basic Books.
- Grice, H.P. (1957). Meaning. *The philosophical review*, 66, 377-388.
- Grice, H.P. (1975). Logic and conversation. In P. Cole, & Morgan, J.L. (Eds.), *Speech Acts*. New York: Academic Press, 41-58.

- Hare, B., & Kwetuenda, S. (2010). Bonobos voluntarily share their own food with others. *Current Biology*, 20(5), R230-R231.
- Hindy, N.C., Solomon, S.H., Altmann, G.T., & Thompson-Schill, S.L. (2013). A cortical network for the encoding of object change. *Cerebral Cortex*, bht275.
- Knight, C., Hurford, J., & Studdert-Kennedy, M., editors (2000). *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge: Cambridge University Press.
- Knudsen, E.I. (2007). Fundamental components of attention. *Annual Review of Neuroscience*, 30, 57–78.
- Levinson, S.C. (1995). Interaction biases in human thinking. In Esther N.G. (Ed.) *Social intelligence and interaction*. Cambridge: Cambridge University Press, 221–260.
- Newman-Norlund, S.E., Noordzij, M.L., Newman-Norlund, R.D., Volman, I.A., Ruiter, J.P.D., Hagoort, P., & Toni, I. (2009). Recipient design in tacit communication. *Cognition*, 111(1), 46–54.
- Noordzij, M.L., Newman-Norlund, S.E., de Ruiter, J.P., Hagoort, P., Levinson, S.C., & Toni, I. (2010). Neural correlates of intentional communication. *Frontiers in neuroscience*, 4.
- Novick, J.M., Kan, I.P., Trueswell, J.C., & Thompson-Schill, S.L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, 26(6), 527–567.
- Nowak, M., & Highfield, R. (2011). *SuperCooperators: Altruism, evolution, and why we need each other to succeed*. New York, NY: Martin Free Press (The).
- Pickering, M.J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(02), 169–190.
- Pinker, S. (1994). *The Language Instinct*. New York: William Morrow and Company.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and brain sciences*, 13(04), 707–727.
- Ramachandran, V.S. (2012). *The tell-tale brain: A neuroscientist's quest for what makes us human*. New York: WW Norton & Company.
- Reboul, A. (2010). Cooperation and competition in apes and humans: A comparative and pragmatic approach to human uniqueness. *Pragmatics & Cognition*, 18(2), 422–440.

- Richerson, P.J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.
- Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C.D. (2014). Supra-personal cognitive control and metacognition. *Trends in cognitive sciences, 18*(4), 186-193.
- Sperber, D., & Origgi, G. (2010). A pragmatic account of the origin of language. In R. K. Larson, V. Déprez, & H. Yamakido, (Eds.), *The Evolution of Human Language: Bilingualistic perspectives*. Cambridge: Cambridge University Press, 124-132.
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language, 17*(1-2), 3-23.
- Tallerman, M. (Ed.). (2005). *Language origins: Perspectives on evolution*. Oxford: Oxford University Press.
- Thompson-Schill, S.L., Bedny, M., & Goldberg, R.F. (2005). The frontal lobes and the regulation of mental activity. *Current opinion in neurobiology, 15*(2), 219-224.
- Tomasello, M. (1995). Language is not an instinct. *Cognitive development, 10*(1), 131-156.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). In search of the uniquely human: Response to commentaries. *Behavioral and Brain Sciences, 28*, 721-727.
- Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind & Language, 18*(2), 121-147.
- Thompson-Schill, S.L., Bedny, M., & Goldberg, R.F. (2005). The frontal lobes and the regulation of mental activity. *Current opinion in neurobiology, 15*(2), 219-224.
- Whiten, A., & Byrne, R.W. (Eds.) (1997). *Machiavellian intelligence II: Extensions and evaluations*. Cambridge: Cambridge University Press.
- Wildgen, W. (2004). *The evolution of human language: scenarios, principles, and cultural dynamics*. Amsterdam: John Benjamins Publishing.



Book Review  
The Recursive Mind. The Origins of Human  
Language, Thought and Civilization

Michael C. Corballis  
Princeton University Press, 2011

*Silvia Felletti*<sup>†</sup>  
silvia.felletti@istc.cnr.it

Language is perhaps the most common way to identify the source of humans' uniqueness within the animal kingdom. Its variety, creativity and flexibility make it hard to believe that our language is something that we could share with other species. It is mainly for this reason that the study of animal communication has long been – and still is – considered by some a useless activity to undertake, when the aim is to reveal the mechanisms underlying the human faculty of language. Among them, Noam Chomsky has claimed that studying animal communication is just a waste of time, since human language is based on an «entirely different principle» (Chomsky, 1966, p. 78). In his view, language is not an outcome of evolution, but rather a consequence of a sudden and fortuitous genetic mutation that occurred in some individual that he nicknamed “Prometheus”.

Although Chomsky is owed enormous credit for his essential finding that language is an innate faculty in humans and that it cannot be explained in terms of learned sequences, his idea, however, of a sudden appearance of language in *Homo sapiens* has not had the same acceptance. Many scientists have in fact tried to reconcile Chomsky's idea of the innate nature of human language with Darwin's natural selection theory, that is crucial for any theory on the origins of language worthy of its name – any true naturalist will agree on this (see, for example, Pinker and Bloom's attempt, 1990).

Michael Corballis should be included among the true naturalists, as he clearly takes a stand against such “emergentist” approaches that, in his own words, «do smack of the miraculous» (2011, p. viii). In his book, *The*

<sup>†</sup> Sapienza, University of Rome, Italy; Institute of Cognitive Sciences and Technologies, CNR Rome, Italy.

*Recursive Mind* Corballis claims that language did not appear suddenly in humans 50-100,000 years ago as Chomsky proposed, but instead evolved during the last 6 million years, arising from pre-existing cognitive systems that we share with other animals. The distinction of Corballis' proposal is that these systems share a characteristic that is critical for human language and mind, that is *recursion*. Recursion is a general property of our minds that allows us to embed elements into other elements of the same kind to form potentially infinite thoughts, behaviours, or linguistic sequences.

*The Recursive Mind* is divided into four main parts, explaining in a simple and informative style what it means for humans to have a recursive mind, how it has influenced the origin and the nature of our language, as well as the overall evolution of hominids, and why recursion is the primary characteristic that distinguishes the human mind from that of other animals.

The first part of the book deals with language and recursion. Recursion in language is the capacity to embed structures within structures to generate expressions such as “John, who’s married to Mary, whom I’ve known for years, works at the airport”. As long as our mind can stand it, sentences like this can expand unlimitedly – even though, realistically, we can get up to only four or five levels of embedding before losing the thread. The reason why language is recursive is most likely because it is used to express recursive thoughts, and indeed, in the author’s view, it is in thought rather than in language that recursion originates. Against the view that language rests upon a “language of thought” (Fodor, 1975), Corballis claims that it depends on a set of other cognitive devices that are not specific for language and that are themselves recursive. Recursion seems to be a general principle that organizes very different spheres of human mental activity, such as mindreading and episodic memory. Thus, recursive speaking may well depend on the recursive nature of non-linguistic thought.

But is language exclusive to our species? Although Corballis affirms that «it remains highly likely that language itself is specific to humans» (2011, p. 35), the essential concern in order to establish a connection between humans and other species is to ascertain whether animals other than humans possess anything resembling human language. The author thus analyses animal communication and the several attempts which have been made to teach human language in all its forms to animals. Attempts have been made with birds, like the famous parrot Alex, who could answer questions he heard for the first time such as “how many green keys are there?” (Pepperberg, 2009). But the

experiments concentrated mainly on the so-called “human raised apes”, like the chimpanzees Washoe (Gardner & Gardner, 1969) and Nim Chimpsky (Terrace et al., 1979) - the latter being a pun on Chomsky’s name -, who learned American Signed language, and the bonobo Kanzi (Savage-Rumbaugh et al., 1998) who learned to communicate with humans using lexigrams on keyboards. The results of the research on Kanzi were particularly relevant: they showed that Kanzi is capable of carrying out tasks based on the understanding of quite long sentences, even when they are new and rather eccentric to him, such as “would you please put the soap in the refrigerator?”. Such studies show at least an apparently simple but surprising ability, that is to discern the single words composing sentences. It is not a banal operation, as we could easily verify by listening to someone speaking a language completely strange for us.

It is up to the reader to decide whether Washoe or Kanzi succeeded or not in learning language, as there is no general consensus on the issue yet. Many insist that the language acquired by apes is not true language, as it does not manifest a true recursive grammar: «they just don’t get it» (Pinker, 1994, p. 340). However, their language can actually be regarded as a form of what Derek Bickerton (1995) defined *protolanguage* a kind of symbolic communication that efficiently transmits information but lacks grammar. Protolanguage is probably what our ancestors hominins spoke, and it is often considered to be the precursor of modern language (see Tallerman, 2007).

Whatever opinion you may have on this, one thing remains certain: while all attempts to teach great apes how to speak have been fatally unsuccessful, significantly better results have been achieved when they have been taught to use written symbols or signed languages. The apes’ difficulty to reproduce the articulated sounds of human language is due to the different physical structure of their vocal tract compared to the human one. But their failure can be also attributed to the common misconception that human language can be compared to animal vocalizations. The sounds emitted by animals, despite being sometimes very similar to our voice - as for some talking birds - have little in common with human speech. Firstly, they are mainly under emotional control, as in the case of the calls emitted involuntarily by chimpanzees who spot a source of food or a predator. Secondly, those sounds are genetically fixed, and, with only few exceptions, tied to the external environment. Moreover, they lack a real intention to inform others, as testified by the fact that chimps will make alarm calls in case of predators even if nobody is around.

This is because vocalizations are largely involuntary: in 1986 Jane Goodall recorded that a chimpanzee cannot help but make a call when she discovers food, even if she would prefer to keep it secret in order to eat it all by herself. In general, it seems that the production of a certain sound in absence of the corresponding stimulus is impossible even for animals as close to us as chimpanzees.

On the other hand, as Corballis points out, great apes' gestures - both the ones directed to humans and those observed in the wild between the same species - are intentional and sensitive to the receiver's attentional state. Apes are in fact capable of following gaze direction and of using it to understand where the agent's attention is directed. Such characteristics are significant, says Corballis, because they are fundamental prerequisites for language. Besides, gestures are more similar to language in that they are typically performed by one individual towards another, while vocal calls are usually directed to an entire group.

This similarity between human language and apes gestures has led Michael Corballis to formulate his theory on the origin of language, already explained in his 2002 book, *From Hand to Mouth*. His idea is that language as a complex and structured behaviour cannot plausibly have evolved recently in our species - with "recently" meaning some 50-100,000 years ago. The emergence of complex structures can only be gradual (Pinker, 1994; Pinker & Bloom, 1990), thus we must look at a time far earlier, that is to say from two to six million years ago, and at the ingredients of language that we share with our primate relatives.

A recent discovery supports the theories on the gestural origins of : this is the finding of the mirror system by Giacomo Rizzolatti and his colleagues at the University of Parma (e.g. Gallese et al., 1996; Rizzolatti et al., 2001). They found that an area in the macaques' brain - the area F5 - that is activated when they make hand movements intentionally, and also during the observation of the same movements performed by another individual. An homologous of this area was also later discovered in humans, located in the Broca's area, critical for language and complex hand movements. The significant difference between the chimps' and humans' mirror system is that in humans it responds not only to transitive acts, such as reaching for an actual object, but also to intransitive acts, in which the movement is mimed and there are no object within the field of view. This capacity could have led to an understanding of symbolic acts and thus to symbolic behaviour and communication. Moreover, mirror neurons in

humans are also activated when we hear or read about an action, suggesting that our mirror system involves language too. The discovery of the mirror system has been used to build a theory of action understanding, and speech perception in particular, by which we do not perceive speech in terms of acoustic patterns, but in terms of how we ourselves would articulate it (see Rizzolatti & Arbib, 1998; Arbib, 2012). So, according to Corballis, intentional communication could actually have originated from action understanding, and mirror neurons could be a suitable candidate for bridging a gap between human language and animal communication.

Since a number of great apes have been able to successfully learn different forms of language-like gestures, it is reasonable to suppose that hominins could also, and that they once used to communicate this way. Such communication maybe lacked grammar that gradually evolved later, «driven by practical concerns and not by biological predisposition» (Corballis, 2011, p. 29). This process goes by the name of *grammaticalization*, and can be seen even today, for example in second-generation users of sign languages or in creole speakers. Time was also critical for *conventionalization*, the process by which gestures lost their iconicity and thus their resemblance to real objects of the world. More arbitrariness meant higher difficulty at the moment of acquiring symbols, but also more variation and flexibility at the stage of production. The next step towards spoken language also meant greater economy of use and freedom to use the hands for other purposes.

But how come we switched from gestures to speech? Corballis' account highlights two main steps. The first is the incorporation of facial gestures, being the face the connection between hands and mouth. Even today, sign languages of the world make use of facial gestures as well as those of the body. Facial gestures increasingly involved invisible movements of the tongue, and in the second step the activation of the vocal cords made these movements perceivable by the ear.

Humans might then have passed through a brain evolution for language, and also undertaken genetic mutation that led to the development of “language genes” such as the FOXP2 (the last is a controversial point, see Fisher & Scharff, 2009 for a review). We also took advantage of anatomical changes that favoured the rise of language: first, the upright stance freed up our hands allowing us to transport and craft objects, and also to gesticulate. Second, the lowering of the larynx created a right-angled vocal tract which was able to articulate the range of vocals that characterizes speech. Even though

bipedalism brought disadvantages such as back pain, and increased birth pain, on the other hand the evolution of speech to which it led brought important advantages such as the possibility to communicate at long distances or in the dark, and a reduced penetrability due to the absence of an iconic component. Today the switch is complete, but we can still witness the remains of the gestural origins of language in our everyday verbal exchanges.

Corballis argues that once the progression from manual gestures to speech was complete, grammar was added to language. But how did recursion and generativity become the most remarkable characteristics of modern language?

Corballis claims that these properties came from the recursive nature of two other cognitive devices, from which language takes its roots: Mental Time Travel and Theory of Mind. These topics are respectively the second and third part of the book.

Mental Time Travel is a phenomenon that joins two different but related capacities: memorizing past events and imagining possible future events (see Suddendorf & Corballis, 2007). Memory is certainly a recursive phenomenon: when we remember or think about the future we are in fact inserting previous or future experiences into present consciousness. Memories and plans can also be inserted into other memories, like when we remember ourselves in the past planning for the future. Humans undoubtedly master such activities: we can remember countless episodes and memorize lots of concepts, but also project ourselves in the future to make plans for our lives or to predict situations to always be ahead of the game. Finally, we are most likely the only species that can use its memories and resolutions for the future to build a structured personal identity.

Corballis suggests that language is tightly linked to mental time travel, as its main purpose is to share our past and future events, and indeed it could actually have evolved for this reason. Humans, in fact, seem to be particularly interested in the lives of others – hence the existence of gossip, which someone has pointed to as the reason why we talk so much (see Dunbar, 1998).

The second element, the Theory of Mind, is the ability to gauge is in the mind of other individuals, whether they are thoughts, emotions, intentions or desires. It is a key ability for activities such as cooperation, deception and for what has been called Machiavellian intelligence (De Waal, 1982). Mind Reading is also a recursive activity: there are in fact several orders of intentionality that we can build, e.g. “I think that she suspects that he knows what I’m about to do” – a mere fourth grade intentionality. As Sperber and

Wilson (1986), and Grice before them (1975) have taught us, the role of intentionality in language is essential. True language is not a mere transmission of words and content, but rather a cooperative activity in which participants have to show their communicative intentions and recognize the same in others. The importance of mindreading for language catches the eye when we analyse language in individuals with deficits of Theory of Mind, such as autistic subjects, who cannot go beyond the literal meaning of sentences.

It is debated whether other animals are capable of some grade of episodic memory, future thinking and mindreading. The most explored field since Premack and Woodruff's 1978 article is that of Theory of Mind: there is some evidence that chimpanzee and bonobos, our nearest related species, show at least some grade of mindreading (see Call & Tomasello, 2008). Even if they may not come to understand false beliefs, they could actually be aware of the attentional states of others, and to effectively use this ability to implement behaviours such as tactical deception (Whiten & Byrne, 1988; for a sceptical view, see Povinelli & Wonk, 2003). As for Mental Time Travel, there is some evidence of future prediction in animals. To name, a few scrub jays who are used to store away the kind of food that they know will not be available later in time (Correia et al., 2007); or Santino, the chimpanzee who has the curious habit to stockpile stones in order to throw them later at unlucky visitors of the Furuvik Zoo in Sweden (Osvath, 2009). But in many cases it is more difficult to exclude some instinctive determinants for such behaviour. Even conceding a glimmer of future thinking capacity in animals, Corballis claims that humans are for sure the only ones who can deliberately measure time to guide intentional behaviour.<sup>1</sup> The author is actually sceptic about the animals' ability to reach a level of recursion higher than the first, in both mindreading and Mental Time Travel. Anyway, nothing prevents us from hypothesizing that our abilities are just refinements of basic capacities already present in other species.

The last part of the book is dedicated to human evolution: here Corballis takes a clear position against the Cartesian discontinuity, and in favour of a Darwinian gradualist continuity between human beings and the other species. He analyses the steps by which we became humans, the facultative and then obligate bipedalism, the evolution of tools manufacturing, the migration from Africa, and the building of a "cognitive niche", and how we evolved our unique

<sup>1</sup> Corballis has recently changed his opinion on this subject, arguing that some animals other than human beings can possess a form of MTT: see Corballis, 2013.

way of thinking from mental structures that were already present in other species. The journey towards modern human beings began yet before the appearance of hominins, even if Corballis points to the Pleistocene (starting about 2 million years ago) as the period in which the most radical changes happened. It is at this moment that language started its evolution, eventually shifting from manual gestures to articulated speech, maybe as a consequence to a gene mutation or simply as a result of a cultural invention. Recursion came later mainly under the pressure of complex social interaction and the inclination to storytelling (but, according to Corballis, only secondarily to the crafting of tools).

In conclusion, the main claim of *The Recursive Mind* is that recursion is that recursion is the key property of the human mind that underlies our unique abilities, such as language, Mental Time Travel and Theory of Mind. But we could still abide with Darwin's postulate that such difference, «great as it is, certainly is one of degree and not of kind» (1871). Recursion is not a module, it did not evolve in some single, miraculous step, and different levels of recursion can be found in the behaviour of many non-human animals, even if they might not go beyond a first level of embedding. For these reasons, language and recursion could turn out to be the link between humans and other species, rather than the source of our uniqueness.

#### REFERENCES

- Arbib, M. (2012). *How the Brain Got Language. The Mirror System Hypothesis*. Cambridge: Oxford University Press
- Bickerton, D. (1995). *Language and human behavior*. London: University College London Press.
- Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in cognitive sciences*, 12(5), 187-192.
- Chomsky, N. (1966). *Cartesian linguistics: A chapter in the history of rationalist thought*. Harper and Row.
- Corballis, M. C. (2002). *From hand to mouth. The origins of language*. Princeton: Princeton University Press.
- Corballis, M. C. (2011). *The recursive mind: The origins of human language, thought, and civilization*. Princeton: Princeton University Press.

- Corballis, M. C. (2013). Mental time travel: a case for evolutionary continuity. *Trends in cognitive sciences*, 17(1), 5-6.
- Correia, S. P., Dickinson, A., & Clayton, N. S. (2007). Western scrub-jays anticipate future needs independently of their current motivational state. *Current Biology*, 17(10), 856-861.
- Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: Murray.
- De Waal, F. (1982). *Chimpanzee politics: Sex and power among apes*. London, UK: Jonathan Cape.
- Dunbar, R. (1998). *Grooming, gossip, and the evolution of language*. Cambridge: Harvard University Press.
- Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã. *Current anthropology*, 46(4), 621-646.
- Fisher, S. E., & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends in Genetics*, 25(4), 166-177.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.
- Fodor, J. A. (1975). *The Language of Thought*. Cambridge: Harvard University Press.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119(2), 593-609.
- Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science*, 165(3894), 664-672.
- Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge: Harvard University Press.
- Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics Volume 3: Speech Acts*. New York: Academic Press, 43-58.
- Osvath, M. (2009). Spontaneous planning for future stone throwing by a male chimpanzee. *Current biology*, 19(5), R190-R191.
- Pepperberg, I. M. (2009). *The Alex studies: cognitive and communicative abilities of grey parrots*. Cambridge: Harvard University Press.
- Pinker, S. (1994). *The language instinct: The new science of language and mind*. London: Penguin.

- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and brain sciences*, 13(04), 707-727.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(04), 515-526.
- Rizzolatti, G., & Arbib. M. (1998). Language within our grasp. *Trends in Neuroscience*, 21, 188-194.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2(9), 661-670.
- Savage-Rumbaugh, S., Shanker, S. G., & Taylor, T. J. (1998). *Apes, language, and the human mind*. Oxford University Press.
- Sperber, D., Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.
- Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans?. *Behavioral and Brain Sciences*, 30(03), 299-313.
- Tallerman, M. (2007) Did our ancestors speak a holistic protolanguage? *Lingua*, 117(3), 579-604.
- Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence?. *Science*, 206(4421), 891-902.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT press.
- Whiten, A., & Byrne, R. W. (1988). Tactical deception in primates. *Behavioral and Brain Sciences*, 11(02), 233-244.

Book Review

More than Nature Needs.  
Language, Mind, and Evolution

Derek Bickerton  
Harvard University Press, 2014

*Serena Nicchiarelli*<sup>†</sup>  
serenanicchiarelli@gmail.com

*There are no grotesques in nature;  
not anything framed to fill up empty cantons,  
and unnecessary spaces.*

Sir Thomas Browne, *Religio Medici*, Part XV (1642)

1870. United Kingdom. *Scenes from a divorce*. Alfred Russel Wallace, the co-inventor with Darwin of the theory of evolution through natural selection, completely captivated by the melodious song of the sirens hailing spiritualism landed definitively on the controversial shores of the treason. In effect, with his essay *The Limits of Natural Selection as Applied to Man*, Wallace turned his back on the explicative power of natural selection regarding the properties that best depict humans, denying «that all nature can be explained on the principles of which I am so ardent an advocate» (Wallace, 1870, p. 133) and wondering how sophisticated skills always in use among members of learned societies may have been inlaid in the brain of savage, de facto foreign to these capacities.

We see, then, that whether we compare the savage with the higher developments of man, or with the brutes around him, we are alike driven to the conclusion that in his large and well-developed brain he possesses an organ quite disproportionate to his actual requirements - an organ that seems prepared in advance, only to be fully utilized as he progresses in civilization (Wallace, 1870, p. 342).

<sup>†</sup> Department of Philosophy, Communication and Visual Arts, Roma Tre University, Italy.

The clashing note in the symphony orchestra of natural selection is briefly the fact that the brain of the savage is shown to be larger than he needs it to be.

So, if on one hand the course undertaken by Wallace to account for human evolution has been the use of supernatural explanations – «the existence of some power, distinct from that which has guided the development of the lower animals through their ever-varying forms of being» (ibid.), on the other hand the question appears to still be in search of a scientific solution.

It's precisely into the empty place of this missing tile – not by chance so called “Wallace's problem” – that Derek Bickerton lays the foundations of his last book *More than Nature Needs. Language, Mind, and Evolution*.

As indeed outlined by the same author on the first pages of his work, the specific topic of his argumentation is to explain how the human species acquired cognitive capacities that seem far more powerful than anything humans could have needed to survive, and specifically, with all due respect to Wallace, to illustrate how this could have come about without the intervention of any mysterious extra-evolutionary forces.

On the other side of the fence, Darwinian gradualism and its proud army composed of *steady slow and continuous changes* in evolution do not seem to convince Bickerton: if the march of evolution proceeds as a gradual process, and the act of natural selection was just a response to the requests placed on animals by their environment, then humans should have had a brain *little superior to that of an ape*. Thus, if Darwin surely was aware of the extent of the problem, having «no doubt that the difference between the mind of the lowest man and that of the highest animal is immense» (Darwin, 1871, p. 100), the argument he used as a counterattack, namely the continuous gradation of intellect between the “lower fishes and the higher apes”, turns out to be a statement not against the gap but for it. In fact, if the vast panorama of nature has countless animals with capacities and skills partway between those of a lamprey and a chimp, how is it that nothing resembling humans exists anywhere else in nature?

How is it that there are no animals with small or moderate amounts of self-consciousness, gradually increasing degrees of innovation and creativity, varying levels of artistic achievement (perhaps in only one or two of the arts), or at least a rudimentary language? The flat assertion of “no fundamental difference” is not (and could not have been, even in Darwin's time) a scientific statement. It was and is a pure declaration of faith (Bickerton, 2014b, p. 3).

So, the immense gap between human mental abilities and those of any other species is presented as an evolutionist's *Achille's heel* and the human mind seems to be a deeply unlikely evolutionary development. Nothing of unexpected. This particular conception is an outlet perfectly coherent with the author's whole body of work. In his *Language and species* (1990), in fact, his starting point is represented by the idea that the differences between language and the most sophisticated systems of animal communication of which we are so far aware of are *qualitative* rather than *quantitative*. Thus, he sketches the contours of his guiding light, "the Continuity Paradox", for which language must have evolved out of some prior system, and yet there does not seem to be any such system out of which it could have evolved. This particular point is restated in his consecutive work, *Adam's Tongue* (2009), in which the linguist acknowledges that «discontinuity exists, and that discontinuity is not limited to language – it extends to all aspects of the human mind. We have, first, to admit that it exists. Then we have to figure out how evolution could have produced it» (Bickerton, 2009, p. 9).

Thinking about language evolution represents an unavoidable exposure to the long-standing debates surrounding empiricist *versus* nativist theories and externalist *versus* internalist explanations. Joining state of the art research with forty years of studying language evolution, Bickerton overtakes this hackneyed "nature or nurture" refrain, underlying how the only constructive way to confront the issue is through a path of synthesis. More specifically, it should be a path that, unresponsive to philosophical or linguistic prejudices, takes into account a genetic component that furnished the basic mechanisms necessary for language and allowed subsequent variation to environmental factors. The first step oriented in the direction of such a synthetic view is to recognize that the attempt to uncover how language evolved is not just "the hardest problem in science", so as defined by Christiansen and Kirby (2003), because it is not simply a problem: it is a disguise that masks three separate problems. Furthermore, this particular kind of *three-question-marked-head* Cerberus, that guards the secrets of language evolution, represents the three problematic situations into which the general process has to be broken down and each of which, forming the backbone of the treatment, requires separate questions and separate answers in the central part of the book.

The first process under analysis is the human jailbreak from the confinement of animal communication. It emerged as a direct response of a specific ecological demand that an ancestral human relative had to face around

two million years ago, a process driven mainly by external evolutionary factors. Before defining where to put the starting grid of a like-language communication system, it would be useful to reflect about how rich or poor the cognitive state of our last alingual ancestors was. This is because the course of language development has been surely conditioned by the degree of complexity of prelinguistic cognition: intuitively, in the presence of an initially rich cognition it's possible to imagine a short trajectory for that course, whereas a longer road would have been traveled if prehumans were relatively poor in cognition. Embracing a specific tendency in the field (e.g., de Waal, 2006; Pollick & de Waal, 2007), the assumption is that communication in the last alingual relative of modern human beings has been profoundly similar to the communication existing among living apes. We then arrive at another stagnant situation according to which nonhumans must have advanced cognition and human-like concepts because there are so many things they can do, but at the same time (coup de théâtre!) they cannot have advanced cognition and human-like concepts because there are so many things they *can't* do (Bickerton, 2014b, p. 79). Thus, a way for avoiding the quagmire deriving from this so defined "Paradox of Cognition" is to delineate the difference between *online thinking*—when there is perception (it involves something that is present in the surrounding situation) - and *offline thinking*—for mental activity not triggered by an immediate external stimulus or by the thinker's current behavior (see also the distinction between *cued representations* and *detached representations* in Gärdenfors, 2003; Gärdenfors & Osvath, 2010). In this regard the pillar idea in the book is that there is no good reason for assuming nonhuman animals to be capable of an offline mental activity. The core of this belief is perched on the idea that the huge creativity and variability in human behavior rests exactly on the capacity to think offline: in reverse, the absence of this ability in all other animals would suffice to justify their slight creativity and minimal variability in behavior. The main role in carrying out offline mental activity is played by voluntary retrieval and by the existence of a hard neural linkage: any concept, in fact, has to be continuously accessible, immediately retrievable, and potentially connectable with every other one. As Bickerton affirms «it can hardly be an accident that these prerequisites, as well as being basic essentials for any complex thinking, are identical with those required for conducting fluent linguistic communication» (Bickerton, 2000, p. 270). Thus, the most meaningful breakpoint with other communication forms is represented by the capacity to transfer information about entities and events

that lay outside the immediate sensory range of the animals concerned: namely *referential displacement*.

The next step in the reasoning is to pursue a likely early hominid need to which a crucial property of language, precisely displaced reference, might have been the answer. That is because any enhancement in intelligence that is not caused by the specific requests of a particular species' niche is extremely improbable - if not impossible - (Odlin-Smee et al., 2003), and every evolutionary change answers a specific need. The need found by the author is the recruitment of fellow hominids to cooperate in butchery and in fending off rival scavengers of megafauna carcasses, beginning in east Africa some 2 million years ago (Bickerton & Szathmary, 2011). Confrontational scavenging requires recruitment. And recruitment requires cooperation. The choice between cooperation and defection is crucially based on the nature of information. The only manner in which the transfer of information, rich and precise enough to guarantee operative collaboration, is possible is by increasing existing models of communication through the addition of displacement. What resulted from this change, at this level, was no more than an enhanced form of animal communication. Over time, in social animals with large brains, the processes evolved into the enrichment of a signal inventory with the capacity of displacement sufficient enough to transform this set into a crude and structureless protolanguage. More specifically, this protolanguage was a system containing only the semantic components of language: "all that nature needed" (e.g. Bickerton, 1990, 2000, 2009, 2014a; Calvin & Bickerton, 2000).

Recruitment for confrontational scavenging forced the prehuman mind to accept the notion that the world might consist of nameable objects.

The ability to name the species would prove not only central to language when language finally emerged but would also establish the linkage between voluntary signals and their related concepts crucial for the development of both language and advanced cognition (Bickerton, 2014b, p. 88).

Through continued use for the more purposes, these displaced reference units (proto-words) would more closely resemble fully symbolic units, and the neural representation of each unit would be linked with a (presumably pre-existing) concept. The presence in the brain of representations of symbolic

units set the second process in motion: the acquisition of very basic structures for the output of the first process.

This second stage was driven by an internal development. When a new source of information becomes available, brains punctually regroup their resources, self-reorganizing in response to their own requirements, such as reducing neural connections to economize both time and energy. In the case of our brave primordial relatives, this new source was represented by a growing store of words and their associated meanings, and the wealth of information that these phenomena created. More specifically the brain, just as a capable interior designer, had to re-distribute spaces for permitting the storage of proto-words, redrawing at the same time its wiring pattern so as to link words with their opportune concepts and with one another, and also with the motor controls for speech.

Whereas in generative theory, both vocal language and Universal Grammar made their appearance simultaneously and independently from any external event (or UG preceded spoken language in the features of a language of thought) substantially because language emerged ready-made, “pretty much as we know it today”, in Bickerton’s proposal the scenery is overturned. In particular, he takes a position against the generative idea that language didn’t evolve to solve any special problem but emerged as a result of organism-internal developments, and that there need not be anything you could call proto-language (e.g. Chomsky, 2010; Piattelli-Palmarini, 2010). Thus, the assumption of Bickerton’s book is conversely that UG and the enhanced communication that would grow into protolanguage emerged separately, in the reverse order. As the linguist states, «from an evolutionary perspective, it seems obvious that words came first but had only a small subset of the properties of modern words, that their arrival precipitated syntax, and that their subsequent interactions with syntax built the set of modern properties» (Bickerton, 2014b, p. 105). The most crucial contribution of brain developments to language does not lay in the sphere of lexical parcellation but in the improvement and automation of the construction of meaningful propositions. This automation of the process of utterance consists of fixing on a stereotyped routine and then increasing the rapidity with which that routine can be executed. More precisely, *Attach, Close, and the phrase and clause algorithms* constitute, in the Bickertonian account, the totality of UG in the sense of specific computational mechanisms for generating syntax. And

syntactic infrastructure so resulted from self-reorganizing activity within the brain itself.

But originally this syntactic engine was not sufficient enough to permit those mental units to be externalized outside the brain, and there was no further round of grammatical processing. How this became possible is the story of the third process, managed by culture: the creation of the kind of language we know today. As we have seen, the human brain, once fertilized with words, developed a means that enabled our ancestral species to achieve and use language. Thus, members of our species began to use linguistic materials, constrained by fundamental elements of UG. On the basis of this bond, every further development corresponds to cultural innovations that would have to be acquired by inductive learning. For this reason, in Bickerton's argumentation, the issue about the third process is presented as strictly connected to the problem of linguistic variation. More specifically, after the brain imposed structure on the output of the first process, how was the final result not a single language but several thousand languages?

The idea of the linguist is that «once humans had the materials for a starter language, change was inevitably going to take place at a rate too fast to form a target for natural selection» (Bickerton, 2014b, p. 152). Besides the inherent instability of the phonetic elements (perhaps the first source of variables that can then be pulled in different directions by a variety of extralinguistic factors: social, cultural, or merely statistical), the other principal factor that contributed to variation and change is that UG was radically underspecified on a second level. On the one hand, there were things unspecified in UG that had to be specified in speech (this is because the brain, far from designing an optimal language, is merely satisfying its own needs for wiring economy and automated routine), on the other hand there were those cases where additional specification, though not strictly necessary for communicative purposes, was seen as enhancing the efficiency of communication.

As we have seen, the model proposed by Bickerton is structured into three steps. So, if protolanguage evolved to facilitate recruitment for confrontational scavenging, on the contrary, language (or rather that part of language instantiated in UG) did not evolve as a consequence of any particular human need, but rather it evolved to refine the brain's speed and accuracy in processing words and concepts. It was successful not to the extent that it improved human fitness but rather to the extent that it satisfied the brain's need for economy and automaticity. In the last step, culture provides the

developments and variation that characterize modern language. As noted by the linguist, «Neither brains nor individuals could have foreseen [...] that the syntax that grew out of protolanguage would create the most powerful cognitive mechanism that had ever existed» (Bickerton, 2014b, p. 162).

This precise model has its direct benchmark in the process of language acquisition by children under normal and abnormal conditions (creolization); both cases being regarded as living forms of protolanguage. The idea is that the child doesn't *learn* or *acquire* language, but rather produces it, «as an automatic reaction to the sound of a running stream of speech, with which they are almost constantly bombarded from birth onward (and even before)» (Bickerton, 2014b, p. 194). The core of language is a small set of algorithms that automatically create basic structures and that are invariant across language. Such UG is fully present throughout development, thus there's no need to postulate any form of grammatical maturation or any cognitive operation actively performed by child on it or with it: it simply sets itself in motion when stimulated by the words around it. After the first “one-word stage”, in which words are few in number, produced in isolation, and very gradually acquired over a period of several months, the next “two-word stage” serves to establish some of the major word-order relationships in the target language and to develop a critical mass of words. The increasing rapidity of vocabulary growth that this mass makes possible determines the transition to the next stage: “telegraphic speech”, consisting of strings of words in phrases or sentences and thus concerning grammatical inflections and simple prepositions. So, this *Language Bioprogram* (Bickerton, 1984) is a single monolithic grammar also used by children in creolizing situations in exactly the same way as it is by children in “normal”, established-language settings. All that differs is the quantity or quality of primary linguistic data in the two cases.

The approach, resulting from the reasoning, satisfies the need for a synthetic view, capable of integrating the evolution of language into an overall account of human evolution. The basic Bickertonian idea is that only by taking such an approach - that regards *syntacticized* language as neither fully innate nor fully learned, but rather compounded of a learned component and an innate component - we can show how nature could have provided our species with powers far in excess of their needs. In response to the opening-stated-purpose of the book, adopting a trio of different solutions takes the author beyond the sterile and seemingly unending arguments of empiricists and nativists alike. Natural selection, internal development, and culture have all

played roles in the evolution of language. It's just that they haven't played them at the same time or in the same process.

## REFERENCES

- Bickerton, D. (1984). The Language Bioprogram Hypothesis. *Behavioral and brain sciences* 7, 173-188.
- Bickerton, D. (1990). *Language and species*. Chicago: University of Chicago Press.
- Bickerton, D. (2000). *How protolanguage became language*. In C. Knight, M. Studdert-Kennedy, & R. Hurford (Eds.), *The evolutionary emergence of language*, Cambridge: Cambridge University Press. 264-284.
- Bickerton, D. (2014a). Some Problems for Bilingualistics. *Biolinguistics*, 8, 73-96.
- Bickerton, D. (2014b). *More than Nature Needs. Language, Mind, and Evolution*, Massachusetts: Harvard University Press.
- Bickerton, D. & Szathmary, E., (2011). Confrontational scavenging as a possible source for language and cooperation, *BMC Evolutionary Biology* 11:261.
- Calvin, W., & Bickerton, D. (2000). *Lingua ex machine*. Cambridge Mass: MIT Press.
- Chomsky, N. (2010). *Some simple evo-devo theses: How true might they be for language?*In R.K., Larson, V. Déprez & H. Yamakido (Eds.), *The Evolution of Human Language: Bilingualistic perspectives*. Cambridge: Cambridge University Press, 45-62.
- Christiansen, M.H. & Kirby, S. (Eds.) (2003). *Language evolution*, Oxford: Oxford University Press.
- Darwin, C. (1871). *The Descent of Man, and selection in relation to sex*, London: John Murray.
- de Waal, F. (2006). *Primates and philosophers. How morality evolved*. Princeton: Princeton University Press.
- Gärdenfors, P., (2003). *How Homo became Sapiens. On the Evolution of Thinking*. Oxford: Oxford University Press.
- Gärdenfors, P., & Osvath, M. (2010). *Prospection as a cognitive precursor to symbolic communication*. In R.K., Larson, V. Déprez & H. Yamakido (Eds.), *The*

- Evolution of Human Language: Bilingualistic perspectives.* Cambridge: Cambridge University Press, 103-114.
- Knight, C., Studdert-Kennedy, M., & Hurford R. (Eds.) (2000). *The evolutionary emergence of language.* Cambridge: Cambridge University Press.
- Larson, R.K., Déprez V., & Yamakido, H., (Eds.) (2010). *The Evolution of Human Language: Bilingualistic perspectives.* Cambridge: Cambridge University Press.
- Odlin-Smee, F.J., Laland, K.N., Feldman, N.W., (2003), *Niche construction: the neglected process in evolution.* Princeton: Princeton University Press.
- Piattelli-Palmarini, M. (2010). *What is language, that it may have evolved, and what is evolution, that it may apply to language?* In R.K., Larson, V. Déprez & H. Yamakido (Eds.), *The Evolution of Human Language: Bilingualistic perspectives.* Cambridge: Cambridge University Press, 148-162.
- Pollick, A.S., & de Waal, F. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Science*, 104, 8184-8189.
- Wallace A. R., (1870). *The Limits of Natural Selection as Applied to Man*, in *Contributions to the Theory of Natural Selection. A Series of Essays.* London-New York: Macmillian

[www.humanamente.eu](http://www.humanamente.eu)

## *Humana.Mente – Journal of Philosophical Studies*

was founded in Florence in 2007. It is a peer-reviewed international journal that publishes 2 issues a year. Each issue focuses on a specific theme, selected from among critical topics in the contemporary philosophical debate, and is edited by a specialist on the subject, usually an emerging researcher with both philosophical and scientific competence.

## *Origin and Evolution of Language*

EDITED BY FRANCESCO FERRETTI AND INES ADORNETTI

HUMANA.MENTE  
Journal of Philosophical Studies

ISSN: 1972 - 1293