Bodily mimesis and the transition to speech

Jordan Zlatev
Centre for Languages and Literature
Department of Cognitive Semiotics
Lund University
jordan.zlatev@ling.lu.se

Abstract
The chapter reviews evidence for the bodily mimesis hypothesis, which states that the evolution of language was preceded by an adaptation for improved volitional control of the body, giving our ancestors advantages in the domains of imitation, empathy and gestural communication. Much of this evidence is also shared by other gesture-first theories of language origins, but they face the problem of explaining the "switch" from a gestural (proto-) language to a spoken one. The bodily mimesis hypothesis fares better with this objection, since it (a) emphasizes the non-conventionality and non-systematicity of bodily mimetic signaling, (b) posits a long bio-cultural spiral of conventionalization and adaptation for speech and (c) insists that the transition to speech should be seen as only partial. Following Brown (2012), a cognitive semiotic explanation can further be given as to why speech has eventually taken on increasingly higher communicative load: vocalization is intrinsically less capable of iconic representation, and given a multimodal gestural-vocal communicative signal, the vocal element is bound to eventually take on the role of symbolic representation, involving higher levels of conventionality and systematicity.

1. Introduction

It is now generally accepted that the human capacity to imitate bodily actions far outstrips that of other animals, including apes (Custance, Whiten and Bard, 1995; Call, 2001). Another capacity, closely related to imitation, in which human beings excel, is intersubjectivity or empathy (Hurely and Chatter, 2005; Zlatev et al., 2008). Jointly, imitation and empathy function as springboards for the development of uniquely human capacities for intentional communication in childhood (Piaget, 1962; Tomasello, 1999; Zlatev, 2013). Considerations such as these have given rise to the *bodily mimesis hypothesis*, stating that an adaptation for improved volitional control of the body gave our ancestors advantages in the domains of imitation, empathy and (gestural) intentional communication. It is

assumed that this paved the way for the evolution of language, with no other biological adaptations being required apart from improved vocal control (Donald, 1991, 2001; Zlatev, 2008a, 2008b).

The first aim of this chapter is to spell out this hypothesis in some more detail and to sum up the empirical evidence in its favor. To some degree, both the hypothesis and the evidence for it overlap with so-called "gesture-first" theories of language origins (Hewes, 1973; Corbalis, 2002, 2003; Arbib, 2003, 2005), but there are some important differences, making bodily mimesis less vulnerable to the most common counterargument to gesture-first theories: why are all current languages of hearing people predominantly spoken rather than gestural, like the signed languages of deaf communities?

The second aim of the chapter is therefore to elaborate on the possible transition from a predominantly mimetic form of communication to a predominantly symbolic one, using the vocal channel. The hurdle has appeared as so great for conceptual as well as empirical reasons, i.e. treating human language as a purely symbolic ("arbitrary") code. It will be argued that the explanatory task appears differently, and as more manageable, if we rather acknowledge the inherently multimodal nature of linguistic communication, with differential roles for speech and gesture, and furthermore see speech itself not as completely arbitrary, but with a considerable degree of sound symbolism (Ahlner and Zlatev, 2010).

2. Bodily mimesis

Donald (1991) initially proposed that bodily mimesis served a crucial role in evolution in his general theory of human cognitive-semiotic origins, defining mimesis as "the ability to produce conscious, self-initiated, representational acts that are intentional but not linguistic" (ibid: 168). In another characterization, he explicates that "it manifests in 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: 240). Crucially, it allowed a qualitatively new form of culture to emerge: "Mimesis served as a mode of cultural expression and solidified a group mentality, creating a cultural style that can still be recognized as typically human." (ibid: 261). Thus, mimesis is manifested in the evolution of the following cognitive-semiotic capacities or functions, in ways that are uniquely human.

(1) Functions of bodily mimesis
- *Learning*: through imitation and teaching
- *Skill*: through conscious rehearsal
- *Imagination and planning*: through re-enactment
- *Communication*: through pantomime and other kinds of gesture
- *Culture*: through shared practices, concepts and beliefs

What has made the bodily mimesis hypothesis attractive is that evidence from a number of different sources can be said to converge toward it. Donald (1991) appealed to the paleoanthropology, neuroscience, and gesture studies of his day. In addition, evidence from human ontogeny (Zlatev, 2007), comparative psychology and "mirror neuron" neuroscience (Zlatev, 2008b), and experimental semiotics (Brown, 2012) have been argued to support the hypothesis as well. What follows is an updated summary of this supportive evidence.

2.1 Paleoanthropology

The hominin species with which bodily mimesis is most strongly associated is *Homo ergaster*, appearing about 1.8 mya in Africa, and the Asian version of this species, *Homo erectus,* attested between 1.5 and 0.1 mya: "the first universally accepted member of our own genus" (Fitch, 2010: 265). The body size of *H. erectus* had increased at least twice compared to the earlier *austalopithecines,* and the brain size even more, to almost modern proportions. The shape of the body had changed as well, giving rise to complete bipedalism, with the capacity for efficient long distance running – highly adaptive for hunting and/or scouting (Cela-Conde and Ayala, 2007). In terms of technology, there was a qualitative shift in style and complexity from older Oldowan to the larger symmetrical handaxes of Achulean technology, requiring considerable skill, practice and pedagogy. These biological and cultural adaptations, including the domestication of fire, from at least 400,000 ya (Weiner et al., 1998), made migration to most parts of Euroasia possible.

Yet, it is not clear if all these achievements coincided with the evolution of the vocal control necessary for speech. One possible marker of such control in the fossil record is an extended thoracic canal, needed for controlling breathing during speech (or singing). Based on earlier evidence, it was concluded that *Homo erectus* still had a thoracic canal in the range of *austalopithecines* (MacLarnon and Hewitt, 1999). This has been contested on the basis of more recent and extensive evidence, suggesting that the species may have had a thoracic canal in the range of modern humans (Gómez-Olivencia et al., 2007). The debate continues, but it remains that while it is clear that *H. erectus* must have had improved volitional control of the body and unprecedented level of culture, there is no firm evidence for the simultaneous evolution of speech. Bodily mimesis thus stands as the likely basis for achievements that are both remarkable, compared to those of earlier hominins, and yet limited compared to those of *Homo sapiens*.

2.2 Mirror-neuron systems

Gestural/bodily theories of language origins received a major boost with the discovery of so-called mirror neurons, responding both to one's own and to others' hand movements, in the 1990s. One argument for their relevance for

language was that they were initially found in area F5 in the premotor cortex of the macaque brain, which appears to be homologous to the left inferior frontal gyrus of the human brain, corresponding to the well-known "Broca's area" (Arbib, 2003, 2005). Extensive studies, using various imaging methods, confirmed that BA 44 and 45 (≈ Broca's area) and BA 22, 39, 40 (≈ Wernicke's area) overlap extensively with the (extended) human "mirror neuron system" (MNS), and are activated in tasks involving action recognition, imitation, pantomime and iconic gestures (Iacoboni, 2008).

Early enthusiasm that this would be sufficient to explain both the neural mechanisms of language and its evolution (Rizzolatti and Arbib, 1998) were, however, rather premature. Admittedly, there is a major gap between the "parity" of action recognition, and that of shared symbolic meanings (Hurford, 2004). In response to such criticism, Arbib (2003, 2005) proposed a more elaborated scenario for how the MNS was gradually extended over evolution from serving the function of action recognition (in monkeys), to "simple imitation" (in apes), to "complex imitation" and pantomime in early *Homo*, to "proto-sign" and eventually to speech. Apart from the stage of "proto-sign", consisting of "elements for the formation of compounds which can be paired with meanings in a more or less arbitrary fashion" (Arbib, 2003: 195), the model is consistent with the bodily mimesis hypothesis (Zlatev, 2008b). For example, BA 4 and BA 6 are not credited with being part of the human MNS, but they have been shown to activate during the perception and production of meaningless syllables (Wilson et al., 2004), and BA 44 and 45 likewise are differentially associated with speech. All this is consistent with the hypothesis that speech was only gradually recruited for intentional communication, "atop" older systems serving action, imitation and gesture.

2.3 Comparative psychology

One of the primary types of evidence used by Hewes (1973) in arguing for a gestural origin of language were the recent for the time findings of relative success in "ape language" studies using a simple form of American Sign Language (ASL). The large controversies that surrounded these studies have made it clear that apes indeed have highly limited abilities to use manual signs compositionally and "declaratively" (i.e. to provide information rather than to request an action), but also that they are capable of learning manual and other forms of non-vocal signs, and to use these flexibly, with close attention to the addressee's state of attention (cf. Zlatev, 2008a). These conclusions have also been confirmed by a number of naturalistic studies of spontaneous bodily communication in great apes, living both in the wild and in captivity (cf. Call and Tomasello, 2007). Tomasello (2008: 54) summarizes the contrast between the vocal and gestural modalities in fairly categorical terms: "… primate gestures are individually learned and flexibly produced communicative acts. […] vocal displays are mostly unlearned, genetically fixed, emotionally urgent, involuntary, and inflexible. […] They are

broadcast mostly indiscriminately." Since extant great apes are our best approximate model for the last common ancestor (LCA) of hominins and apes, it is reasonable that the LCA had similar skills, and that gesture/bodily mimesis was therefore within its "zone of proximal evolution" (Donald, 2001), unlike speech. While several researchers have argued that such an appraisal underestimates chimpanzee vocal capacities and their communicative functions (Slocombe and Zuberbuehler, 2005), it seems clear that there is at least a quantitative if not qualitative difference between the flexibility, volitional control, and referentiality of ape gestures as opposed to vocalizations (Pika, 2008). Thus again: producing signs with the body was more "at hand" than with the voice.

Looked from the other direction, what are the main differences between ape and human cognition, leaving language aside? It has been popular for some time to downplay such differences (e.g. Tallis, 2011), but in a recent extensive review article, Vaesen (2012) examines the evidence from nine cognitive domains (including language) related to tool production and use, and concludes that "striking differences between humans and great apes stand firm in eight out of nine of these domains" (ibid: 203). The seven non-linguistic domains in which human capacities clearly exceed those of apes according to this review are: (a) hand-eye coordination, (b) causal reasoning, (c) functional representations (e.g. for tools), (d) executive control (e.g. inhibition and planning), (e) social learning (e.g. imitation), (f) teaching and (g) social intelligence (e.g. passing false belief tasks). Rather than considering one of these as the crucial difference, Vaesen concludes that "no individual cognitive trait can be singled out as the key trait differentiating humans from other animals." (ibid: 203). This claim is quite in line with the bodily mimesis hypothesis, since mimesis is poly-functional. Indeed, there is a close correspondence between the functions associated with bodily mimesis under (1) and the features in Vaesen's list given above, especially when the latter are grouped as (a) motoric, (b-d) cognitive, and (e-g) social-cognitive.

In such a manner, the bodily mimesis hypothesis of the origins of human uniqueness can help generalize over a number of findings from comparative psychology.

2.4 Gestures and ontogeny

Several decades of extensive research on the spontaneous gestures of adults and their development in children have shown that gestures are ubiquitous in all human cultures and that they align temporally and semantically with speech, at least in adult language use (Kendon, 2004; McNeill, 2005). The explanations of these findings, however, differ. While McNeill (1985, 2005, 2012) considers speech and gesture (production) to be two parts of a single system, others point out that there are good reasons to regard them as two closely interacting, but distinct systems. The resolution of this controversy has direct implications for evolutionary hypotheses.

It is now generally accepted that gestures share semantic properties with what is being said, and that speakers of different languages gesture somewhat differently, in ways that can be related to the semantics of the respective languages (Kita and Özyürek, 2003). However, speakers also use gestures to represent objects and events iconically in ways that go beyond what is said, and in ways that are similar across languages (Zlatev and Andrén 2009). This is consistent with a model of "the two qualitatively different representations [which] are adjusted with respect to each other and co-evolve" (Kita and Özyürek, 2003: 30). Careful analyses have also shown that co-speech gestures synchronize with features of the interaction as a whole, including the responses of the addressee (Sikveland and Ogden, 2012), and are thus not automatically tied to speech production itself.

The developmental evidence also appears to support an analysis in terms of two interacting systems rather than a completely unseperable speech-gesture bond of the kind that McNeill envisages. On the one hand, there is general agreement that there is close interaction between gesture and speech in language development (Volterra et al., 2005; Goldin-Meadow, 1998; Andrén, 2010). Still it appears that both pointing and iconic gestures emerge prior to speech, at around 9-12 months, and play an essential role for the development of language (Bates et al., 1979; Liszkowski et al., 2012; Lock and Zukow-Goldring, 2012). Speech and gesture become gradually integrated in ontogeny, with at least some analyses showing "a gradual specialization from unimodal forms of communication, less demanding in cognitive, social and semiotic terms, to multimodal patterns involving the coordination of specific gestures and vocalizations" (Murillo and Belinchón, 2012: 31).

Of course, such apparent gestural primacy in ontogeny is not a strong argument for a corresponding primacy in evolution, since the old principle of "recapitulation" cannot be accepted without prior justification. Still, if gesture plays a scaffolding role for language in development, it is reasonable to suppose that it played an analogous role in evolution as well, since in both ontogeny and phylogeny (a) bodily movement comes under volitional control earlier than vocalization, as argued in 2.3, and (b) gesture affords a greater degree of *iconicity* than speech.

The last point, i.e. the iconic (resemblance-based) relation between at least some gestures and their meanings, has been a rather controversial topic. Intuitively, communicating with the whole body should be easier than only with the voice when lacking a common language, since this is indeed what people do when they need to communicate in such cases. On the other hand, many gestures are conventionalized, and some researchers have even argued that iconicity plays hardly any role at all in gestural communication (Streek, 2009). This controversy can be in part resolved by turning to semiotics, where the topic of iconicity has been thoroughly investigated.

2.5 Semiotic analysis and experiments

Semiotics is the interdisciplinary field investigating commonalities and differences between different communicative systems, such as visual representations, speech and gestures (in both spontaneous and artistic forms), and their dependence on and interaction with cognitive capacities including perception, movement and consciousness (cf. Sonesson, 1989). While traditional semiotics was based almost entirely on a form of conceptual analysis, and was often quite speculative, modern approaches of experimental (Galantucci & Garrod, 2010) and cognitive semiotics (Zlatev, 2012) are considerably more empirical. It is the combination of conceptual (intuition-based) analysis and experimental validation that makes semiotics so useful in addressing controversial topics such as the iconicity of gestures.

First of all, it is important to recognize that iconicity and conventionality (as well as the third type of expression-meaning relation known as indexicality, which is contiguity-based) do not stand in a mutually exclusive relation, as pointed out by several of the classics of the field:

> One of the most important features of Peirce's semiotic classification is … that the difference between the three basic classes of signs is merely a difference in relative hierarchy. It is not the presence or absence of similarity or contiguity between the signans and signatum, nor the … habitual connection between both constituents which underlies the division of signs into icons, indices and symbols, but merely *the predominance of one of these factors over the others*. (Jakobson, 1965: 26, my emphasis)

Furthermore, in his defense of the iconicity of pictures, Sonesson established a useful conceptual distinction between *primary iconicity*, where "the perception of an iconic ground obtaining between two things is one of the reasons for positing the existence of a sign function joining two things together as expression and content" and *secondary iconicity*: "the knowledge about the existence of a sign function between two things […] is one of the reasons for the perception of an iconic ground between these same things" (Sonesson, 1997: 741). The iconicity of a typical picture (Fig 1a) is primary, whereas that of a more abstract representation such as that shown in (Fig 1b) is secondary: first when we are told that this represents e.g. a man in a telephone booth playing a trombone can we see the resemblance.
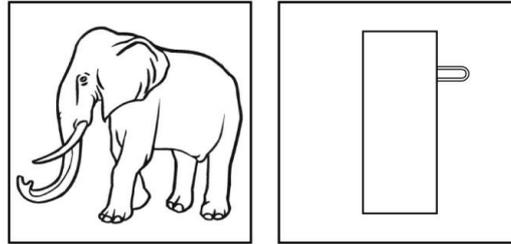
Figure 1. An example of (a) primary vs. (b) secondary iconicity (borrowed from Ahlner and Zlatev, 2010)

The question concerning gestures can now be reformulated along the lines of Jakobson (1965): does iconicity "predominate" over conventionality at least in some cases and in the style of Sonesson (1997): is it of the primary kind? A recent experimental study by Fay, Arbib and Garrod (2013) suggests positive answers to both questions. The researchers asked pairs of participants to play a game in which a "director" had to communicate 24 different concepts, divided in the categories Emotion, Action and Object, to a "matcher", without using language, by one of three means: vocalization, gesture or a combination of both. The results showed that in all cases matching was above chance, and that for the Emotion class, the vocalization-only group managed fairly well (ca. 70%). However, (pantomimic) gestures with or without vocalization, had a clear advantage, with success rates approaching ceiling level. The authors conclude that "gesture outperforms non-linguistic vocalization because it lends itself more naturally to the production of motivated signs" (ibid: 1). Since the game was played a number of times by each pair, a degree of simplification and conventionalization of the gestures occurred, but in no point did they become "arbitrary", or their iconicity purely secondary. On the other hand, the success rates for vocalization-only increased considerably with use, suggesting that conventionalization played a stronger role for successful communication in that medium. This leads to an important conclusion: while both the bodily/gestural and vocal modalities can be used for signs that are fully conventionalized, to the extent of losing all traces of iconicity and indexicality and thus becoming "arbitrary", the bodily/gestural modality is *intrinsically* more suited for motivated signs, while the vocal modality is less so. This difference is crucial to explain both why bodily mimesis and gesture are advantageous for establishing a sign system initially, and why with time there will be a shift towards the vocal modality, i.e. speech, as argued below.

3. But why speech?

The different kinds of evidence discussed in the previous section are supportive not only of the bodily mimesis hypothesis, but of gesture-first theories of language

evolution in general. The proposal of a "gestural stage" in language evolution has always been found appealing to some, but objectionable to others who have theorized about language origins. The major objection can be formulated tersely: why speech? Even authors who are very well aware of the importance of gesture in human communication find this objection (nearly) "fatal" or "insuperable":

> The gestural theory has one nearly fatal flaw. Its sticking point has always been the switch that would have been needed to move from a visual language to an audible language. (Burling, 2005: 123)

> Several different lines of evidence, then, can be added up to support the hypothesis that the first step in the evolution towards linguistic expression was taken with the employment of visible action, or gesture, for referential expression. Yet, as has often been pointed out, this seemingly attractive hypothesis faces […] an insuperable problem: Languages are overwhelmingly *spoken.* (Kendon, 2008: 12)

In his critical review of "gestural protolanguage theories", Fitch (2010, Chapter 13), argues convincingly that appealing to ecological factors is not sufficient to explain the transition to speech, since "each posited advantage can be paired with a similar selective force that would oppose them" (ibid: 443). Communicating in the dark may be beneficial, but silent gesturing is clearly safer in an environment of extensive predation. Speech may be "freeing the hands" for other purposes while communicating, but then it "burdens the mouth", making communication somewhat difficult, and even dangerous during a common communal activity: eating. Analogously, vocal communication may free visual attention, but it burdens auditory attention and furthermore in all cultures linguistic communication is predominantly conducted "face-to-face", involving multi-modal perception (Kendon, 2004).

As Fitch points out, Hewes (1973) did not appeal to such factors but rather to what he then thought to be certain linguistic disadvantages of signed languages compared to speech: having a limited vocabulary, lacking duality of patterning, i.e. the equivalent of phonemes, and being slower. However, such claims have been disproved since then. As even the currently popular praxis of parallel translation between spoken and signed languages shows, signed languages have the full linguistic functionality of spoken languages. This has made them a potent argument against an initial "gestural protolanguage": if everything that can be said can be just as easily signed, than why turn to speech? Furthermore, as recent studies of emerging signed languages show, modern human beings are capable of spontaneously constructing a signed language from the pantomimic kind of gestures typical of bodily mimesis over the span of a few generations (Senghas, Senghas, & Pyers, 2005) (Sandler, 2012).

The why-speech argument is indeed damaging to some proposals of gestural primacy, but not to all. On the one hand, proposals differ with respect to how

exactly the "gestural protolanguage" is conceived of. Corballis sees it as "a form of signed language similar in principle, if not in detail, to the signed languages that are used today by the deaf" (Corballis, 2003: 125). Arbib, it will be remembered, breaks up the evolutionary process in several stages, and preceding speech there is "proto-sign: a manual-based communication system, breaking the fixed repertoire of primate vocalizations to yield a combinatorially open repertoire […] elements for the formation of compounds which can be paired with meanings in a more or less arbitrary fashion" (Arbib, 2003: 195). Bodily mimesis, on the other hand, corresponds to neither: its virtue (as well as its ultimate disadvantage) is that the type of signs (in the semiotic sense) that it gives rise to are precisely *not* conventionalized, arbitrary and combinatorial (Zlatev, 2008a).

Furthermore, very few if any of the proponents of gestural primacy in evolution view the transition to speech as a discrete "switch", but rather as a process that was both gradual, and given the ubiquity of co-speech gesture, still remains only partial:

> While human primates must have been at first better at transmitting information through gesture than through voice, at some point voice became the preferred vehicle. But what if this "point" was a transitional period of over half a million years, say, from the appearance of *Homo erectus* to that of archaic *Homo sapiens*? And what if, during all this time, humans regularly communicated bi-modally, only gradually shifting from a code that foregrounded gesture to one that foregrounded voice?" (Collins, 2013: 136)

In general, the less pre-linguistic gestural communication is thought of as a "language", and the less modern spoken languages are conceived of as purely vocal, the less problematic the why-speech argument appears. While it is indeed damaging for scenarios that frame the transition as one "from hand to mouth" (Corballis, 2002), they are not if stated in the much less idiomatic "from body to mouth-and-body" (Zlatev, Donald, and Sonesson, 2010). That is, from whole-body communication supported by the human-specific capacity for bodily mimesis to the multi-modal system of linguistic communication which we use today, involving both speech and gesture.

Thus, the typical counter-argument against gesture-first theories is not in principle "fatal" or "insuperable" for the bodily mimesis hypothesis of human cognitive, and linguistic, origins. Still, a more explicit account of how and why the transition has taken place is due. In a recent doctoral dissertation, Brown sets herself this task precisely:

> A major step in the evolutionary process by which human communication could have emerged has been proposed in the bodily mimesis hypothesis. … This ability provides a foundation from which symbolic communication can arise, but how such a transition would

> have taken place has not been fully examined. This thesis examines
> the gap between bodily mimesis and symbolic communication.
> (Brown, 2012: 1)

Brown reviews different gesture-first theories of language origins and concludes, similarly to Fitch (2010) that those that posit some form of "switch" between an already conventionalized (proto) language and speech (e.g. Corballis, 2002; Arbib, 2005) fail to provide an adequate explanation for this switch. In addition to the issues discussed in Section 3, Brown argues that an intermediary stage of arbitrary gestures, e.g. corresponding to Arbib's notion of "proto-sign", would have minimized support for the stabilization of a conventional code: "the conventionalization process requires a rich and supportive communicative infrastructure in which novel arbitrary signs can be used … so that the intended form-meaning relationships could be correctly interpreted" (Brown, 2012: 81). This conclusion is supported by computational models of language evolution showing that the stabilization of a conventional language across a greater number of speakers requires factors such as extensive corrective feedback or restricted context – neither of which is characteristic of actual communication – or support from parallel non-arbitrary signals.

While theories that posit that "multimodal referential communication was a combination of arbitrary and non-arbitrary representation from inception" (ibid: 116), such as that of McNeill (2012), avoid the need to explain any switch, they face complementary problems since they both predict a stronger degree of speech-gesture unity that appears to be the case (cf. Section 2.4), and underestimate the degree of non-arbitrariness in speech.

By method of exclusion, Brown concludes that theories that propose a gradual and only partial transition from mimesis/gesture to speech (e.g. Zlatev, 2008b; Collins 2013) are most plausible, but objects that they "do not provide a reason why one modality is now predominantly symbolic and not the other" (Brown, 2012:120), i.e. why speech has undergone a greater degree of conventionalization, showing less iconicity, than gesture.

The answer proposed by Brown is both simple and ingenious: "the vocal modality would have become predominantly symbolic because its lower non-arbitrary capacity increases the likelihood that vocalizations are perceived as arbitrary" (ibid: 134).

This conclusion is supported by the methods of experimental semiotics (cf. Section 2.5), showing that the gestural modality carries more "communicative load" than the vocal modality when communication is restricted to non-conventional signaling, and furthermore that iconic gestures help the audience to interpret novel vocalizations as meaningful words, even when the latter are perceived as "arbitrary". Supported by a combination of semiotic experimentation and computational modeling, Brown concludes that in multi-modal gesture-vocalization communication, there will be an automatic pull towards increased

arbitrariness with the need to communicate a larger and more diverse set of concepts, and that this would take place in the vocal modality.

Taken along with the scenario suggested by Collins (2013) of a gradual shift of communicative load from gesture to speech over the duration of "over half a million years" thus gives a plausible answer to the why-speech question: Due to the diversification of hominin cultures, a less iconic (= more symbolic) code would have been beneficial, and since the vocal modality affords less iconicity than the manual/bodily one, it became naturally "recruited" to the task. The supposition that this took place from the emergence of *Homo erectus* at 1.5 MYA to *Homo sapiens* at 0.2 mya gives more than sufficient time for necessary biological adaptations necessary for increased vocal control to take place. The answer is consistent with evidence for bodily mimesis summarized earlier and with the increasing evidence for the partial non-arbitrariness of speech (Ahlner and Zlatev, 2010).

4. **Conclusions**

This chapter reviewed some of the confirming evidence for the bodily mimesis hypothesis, much of which can be also brought in favor of gesture-first theories of language origins. Unlike some recent and well-known proposals of a "gestural protolanguage", however, bodily mimesis is both a more general adaptation, since it concerns the volitional use of the body for other means than gestural communication as well, and less language-like. Hence, it was argued that it fares much better against the argument typically bought against gesture-first theories: how to explain the switch from a gestural (proto) language to a spoken one. It does so since (a) it emphasizes the non-conventionality and non-systematicity of bodily mimetic signaling, (b) it rejects the notion of a switch, and instead posits a long bio-cultural spiral of conventionalization and adaptation for speech and (c) insists that the "transition", which is possibly the wrong word, should be seen as only partial, given all the evidence for the adaptive role of gesture in language development and face-to-face communication.

What Brown's theorizing and evidence adds to this is a cognitive-semiotic explanation for why speech has during this process taken an increasingly higher communicative load: bodily movement and vocalization do not differ in their capacity to represent meaning purely conventionally, but vocalization is intrinsically less capable of doing so iconically. Given a multimodal gestural-vocal communicative signal, the vocal element is bound to be less iconic than the gestural, and thus to differentiate more clearly between an extensive set of concepts, even when their referents are visually similar.

In sum, the transition from communication based on bodily mimesis to relatively "arbitrary" speech was made possible by the multimodal character of human communication, through a prolonged process of increased articulation and conventionalization, but without language cutting off its bodily roots.

References

Alhner, F., and Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign System Studies, 38*(1/4), pp. 298-348.

Andrén, M. (2010). *Children's gestures between 18 and 30 months.* Lund: Media Tryck.

Arbib, M. (2003). The evolving mirror system: A neural basis for language readiness. In M. Christiansen, and S. Kirby, *Language Evolution* (pp. 182-200). Oxford: Oxford University Press.

Arbib, M. (2005). From moneky-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28*, pp. 105-168.

Bates, E., Benigni, L., Bretherton, I., Camioni, L., and Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York: Academic Press.

Brown, J. E. (2012). *The evolution of symbolic communication: An embodied perspective. PhD thesis.* Edinburgh: University of Edinburgh.

Burling, R. (2005). *The talking ape.* Oxford: Oxford University Press.

Call, J. (2001). Body imitation in an enculturated orangutan (Pongo pygmaeus). *Cybernetics and Systems, 32*, pp. 97–119.

Call, J., and Tomasello, M. (2007). *The gestural communication of apes and monkeys.* London: Lawrence Erlbaum.

Cela-Conde, C. J., and Ayala, F. J. (2007). *Human evolution, Trains from the past.* Oxford: Oxford University Press.

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.

Corballis, M. C. (2003). From hand to mouth: The gestural origins of language. In M. Christiansen, and S. Kirby, *Language evolution* (pp. 201-218). Oxford: Oxford University Press.

Custance, D., Whiten, A., & Bard, K. (1995). Can young chimpanzees (*Pan troglodytes*) imitate arbitrary actions? *Behaviour,* 132: 837–859.

Donald, M. (1991). *Origins of the moden 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.

Fay N., Arbib, M., Garrod, D. (2013) How to bootstrapp a human communication system. *Cognitive Science* 37, 7:1356-1367.

Fitch, W. T. (2010). *The evolution of language.* Cambidge: Cambridge University Press.

Galantucci, B., and Garrod, S. (2010). Experimental Semiotics: A new approach for studying the emergence and the evolution of human communication. *Interaction Studies, 11*, pp. 1–13.

Goldin-Meadow, S. (1998). *The development of gesture and speech as an integrated system.* San Francisco, CA: Jossey-Bass.

Gómez-Olivencia, A., Carretero, M. J., Arsuaga, L., Rodríguez-García, J. L., García-González, R., and Martínez, I. (2007). Metric and morphological study of the upper cervical spine from the Sima de los Huesos site (Sierra de Atapuerca, Burgos, Spain). *Journal of Human Evolution, 5*, pp. 6-25.

Hewes, G. (1973). Primate communication and the gestural origins of language. *Current Anthropology, 14*, pp. 5-24.

Hurely, S., and Chater, N. (2005). *Persepctives on imitation. From neuroscience to social science. Vol I and Vol II.* Cambridge, MA: MIT Press.

Hurford, J. R. (2004). Language beyond our grasp. In K. Oller, U. Griebel, and K. Plunkett, *Evolution of Communication Systems: A Comparative Approach* (pp. 297-313). Cambridge, MA: Cambridge University Press.

Iacoboni, M. (2008). *Mirroring people: the new science of how we connect with others.* New York: Farrar, Straus and Giroux.

Jakobson, R. (1965). Quest for the essence of language. *Diogenes, 13*, pp. 21-38.

Kendon, A. (1994). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kendon, A. (2009). Signs for language origins? *The public journal of semiotics, 2*(1), pp. 2-27.

Kita, S., and Özyurek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, pp. 16–32.

Liszkowski, U., Brown, P., Callaghan, T., Takida, A., and de Voc, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science*, pp. 1-16.

Lock, A., and Zukow-Goldring, P. (2012). Preverbal communication. In J. Bremner, and T. Wachs, *The Wiley-Blackwell Handbook of Infant Development* (pp. 394-425). Oxford: Willey-Blackwell.

MacLarnon, A. M., and Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. *Americal Journal of Physical Anthropology, 109*, ss. 341-363.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92*, pp. 350-371.

McNeill, D. (2005). *Gesture and thought.* Chicago: University of Chicago Press.

McNeill, D. (2012). *How language began: Gesture and speech in human evolution.* Cambridge: Cambridge University Press.

Murillo, E., and Belinchón, M. (2012). Gestural-vocal coordination. *Gesture, 12*(1), pp. 16-39.

Piaget, J. (1962). *Play, dreams, and imitation in childhood.* New York: Norton.

Pika, S. (2008). What is the nature of gestural communication in great apes? In J. Zlatev, T. Racine, C. Sinha, and E. Itkonen, *The shared mind: Perspectives on intersubjectivity* (pp. 165-186). Amsterdam: Benjamins.

Rizzolatti, G., and Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience, 362*, pp. 188-194.

Sandler, W. (2012). Dedicated gestures, and the emergence of sign language. *Gesture, 12*(3), pp. 265-307.

Senghas, R., Senghas, A., and Pyers, J. (2005). The emergence of Nicaraguan sign language: Questions of development, acquisition and evolution. In J. Langer, S. Parker, and C. Milbrath, *Biology and knowledge revisted: From neurogenesis to psychogenesis* (pp. 287-306). Mahwah, NJ: Lawrence Erlbaum.

Sikveland, R. O., and Ogden, R. (2012). Holding gestures across turns. *Gesture, 12*(2), pp. 166-199.

Slocombe, K., and Zuberbuehler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology, 15*, pp. 1779-1784.

Sonesson, G. (1989). *Pictorial concepts. Inquiries into the semiotic heritage and its relevance for the analysis of the visual world.* Lund: Aris/Lund University Press.

Sonesson, G. (1997). The ecological foundations of iconicity. In I. Rauch, and G. F. Carr, *Semiotics around the world: Synthesis in diversity* (pp. 739–742). Berlin : Mouton de Gruyter.

Streek, J. (2009). *Gesturecraft: The manu-facture of meaning.* Amsterdam: Benjamins.

Tallis, R. (2011). *Aping mankind. Neuromania, darwinitis and the misrepresentation of humanity.* Durham: Acumen.

Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard University Press.

Tomasello, M. (2008). *The origins of human communication.* Cambridge, MA: MIT Press.

Vaesen, K. (2012). The cognitive bases of human tool use. *Behavioral and Brain Sciences, 35*, pp. 203–262.

Weiner, S., Xu, Q., Goldberg, P., Liu, J., and Bar-Yosef, O. (1998). Evidence for the use of fire at Zhoukoudian, China. *Science, 281*, pp. 251-253.

Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*(7), pp. 701-712.

Volterra, V., Caselli, M., Caprici, O., and Pizzuto, E. (2005). Gesture and the emergence and development of language. In M. Tomasello, and D. Slobin, *Beyond nature-nurture: Essays in honor of Elisabeth Bates* (pp. 3-40). Mahwah, NJ: Lawrence Erlbaum.

Zlatev, J. (2007). Intersubjectivity, mimetic schemas and the emergence of language. *Intellectica, 46*(2-3), pp. 123-152.

Zlatev, J. (2008a). The coevolution of intersubjectivity and bodily mimesis. In J. Zlatev, T. Racine, C. Sinha, and E. Itkonen, *The shared mind: Perspectives on intersubjectivity* (pp. 215-244). Amsterdam: Benjamins.

Zlatev, J. (2008b). From proto-mimesis to language: Evidence from primatology and social neuroscience. *Journal of Physiology – Paris, 102*, pp. 137-152.

Zlatev, J. (2012). Cognitive semiotics: An emerging field for the transdisciplinary study of meaning. *Public Journal of Semiotics, 4*(1), pp. 2-24.

Zlatev, J. (2013). The mimesis hierarchy of semiotic development: Five stages of intersubjectivity in children. *Public Journal of Semiotics, 4*(2), pp. 47-70.

Zlatev, J., and Andrén, M. (2009). Stages and transitions in children's semiotic development. In J. Zlatev, M. Andrén, M. Johansson-Falck, and C. Lundmark, *Studies in Language and Cognition* (pp. 380-401). Newcastle: Cambridge Scholars.

Zlatev, J., Donald, M., and Sonesson, G. (2010). From body to mouth (and body). In A. Smith, M. Schouwstra, B. deBoer, and K. Smith, *The evolution of language* (pp. 527-528). London: World Scientific.

Zlatev, J., Racine, T., Sinha, C., and Itkonen, E. (2008). *The shared mind: Perspectives on intersubjectivity.* Amsterdam: Benjamins.