

Human Uniqueness, Bodily Mimesis and the Evolution of Language

Jordan Zlatev[†]

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

[†] 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.¹ 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

¹ 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 onomatopoetic 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).

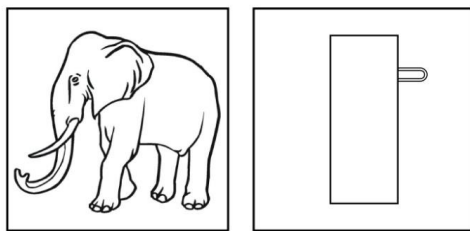


Figure 1. A primary vs. secondary iconic sign: the represented object in the first case is obvious, while in the latter it can be seen only after having been pointed out, e.g. a trombone sticking out from behind a door (from Ahlner & Zlatev, 2010, p. 306)

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 Caroline 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 bootstrapp 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 autonoesis: 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. Contier (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.

