Sign Systems Studies 38(1/4), 2010

Cross-modal iconicity: A cognitive semiotic approach to sound symbolism

Felix Ahlner, Jordan Zlatev

Centre for Languages and Literature, Lund University
Box 201, 221 00, Lund, Sweden
e-mail: felix.ahlner@gmail.com, jordan.zlatev@ling.lu.se

Abstract. It is being increasingly recognized that the Saussurean dictum of "the arbitrariness of the linguistic sign" is in conflict with the pervasiveness of the phenomenon commonly known as "sound symbolism". After first presenting a historical overview of the debate, however, we conclude that both positions have been exaggerated, and that an adequate explanation of sound symbolism is still lacking. How can there, for example, be (perceived) similarity between expressions and contents across different sensory modalities? We offer an answer, based on the Peircian notion of *iconic ground*, and G. Sonesson's distinction between *primary* and *secondary iconicity*. Furthermore, we describe an experimental study, in a paradigm first pioneered by W. Köhler, and recently popularized by V. Ramachandran, in which we varied vowels and consonants in fictive word-forms, and conclude that both types of sounds play a role in perceiving an iconic ground between the word-forms and visual figures. The combination of historical conceptual analysis, semiotic explication and psychological experimentation presented in this article is characteristic of the emerging paradigm of *cognitive semiotics*.

1. Introduction

Is the relationship between the expression and content poles of the linguistic sign fundamentally arbitrary, as it is typically claimed, following the famous dictum of the "father of modern linguistics" (see Lyons 1968), Ferdinand de Saussure, or is there some kind of 'natural connection' between the two? This question dates back to antiquity, but — we would claim — is still not satisfactorily resolved. Even the most committed proponents of the arbitrariness dictum recognize the existence of onomatopoetic words, but tend to regard these as 'primitive curiosities', existing outside and alongside the language system proper. The alternative viewpoint, commonly known as 'sound symbolism' (Hinton *et al.* 1994), has gained ground during the last few decades. For example, a result that has been replicated a number of times is that when both adults and children (without autism) are given two fictive words like *bouba* and *kiki* and asked to decide which one denotes a roundish and which a pointy figure, they agree up to 95% that *bouba* suits best the roundish one (for example, Ramachandran, Hubbard 2001). How are we to explain this, and is it possible to find a dialectical synthesis that could help resolve the debate between 'arbitrariness' and 'sound symbolism'? This is the main question that we address in this article.

We proceed in three steps. First, we present a historical and theoretical overview of the debate, involving theoretical positions in linguistics and semiotics, and relate these to relevant empirical findings. This review shows surprising conceptual shortcomings in the discussion, including vague and inconsistent definitions of key terms such as 'arbitrariness', 'sound symbolism' and 'iconicity'. Hence, in a second step we endeavor to present a consistent semiotic analysis of the (possible) relation of resemblance between linguistic sound patterns and the objects that they denote, allowing us to make sense of results such as those mentioned in the previous paragraph. Such exercises of conceptual analysis and explication are necessary, but not

sufficient to understand the phenomenon in detail. Therefore, in the spirit of the emerging paradigm of cognitive semiotics (see Zlatev 2009), "integrating methods and theories developed in the disciplines of cognitive science with methods and theories developed in semiotics and the humanities"[1], we turn in our third step to the description of a psychological experiment aiming to decide whether the relevant (cross-modal) similarity between linguistic expression and content is found at the level of segments (vowels and consonants) or in the combination of segments, and possibly in whole "sound patterns".

While by no means claiming that this three-step exercise will provide the 'resolution' to the age-old debate, we will suggest that it casts it in a partially new light, with the potential for a synthesis between (aspects of) positions that have been usually regarded as incommensurable: the conventionality and motivatedness of lexical meaning. We will conclude by suggesting how our analysis, and in particular the notion of *cross-modal iconicity*, can be productive for future research.

2. Historical and theoretical background

2.1. Ancient Greece

As many 'big questions' in Western thought, the debate between the proponents of linguistic 'conventionalism' *vs.* 'naturalism' goes back to ancient Greece. In fact, language was not the primary focus of the ancients. The laws of society had for a long time been regarded as immutable and natural, but if they could be regarded as determined by convention, they could be questioned and debated, which was fundamental for the idea of democracy (see Lyons 1968; De Cuypere 2008). Since language, in the form of public discourse, constituted the

[1] www.cognitivesemiotics.com

means through which these discussions were carried out, it was inevitable that the debate would also come to involve the proper understanding of its essence.

Plato's famous dialogue *Cratylus*, from around 400 BC, is the principle reference for this early debate.[2] The two main stances that had developed up to that time were that language is either given by nature (*phýsei*) or by convention (*thései*). In the dialogue, these are represented by Cratylus and Hermogenes, respectively.[3] The former argues that a word's sound-form is somehow 'naturally' connected to its meaning. Therefore, a particular sound-pattern can be deemed more or less appropriate for expressing a particular meaning. The latter argues that words are man-created and passed on through generations, thus constituting the core of a conventional system of communication: language. As a supporter of *phýsei*, Cratylus presents examples of how certain sounds are better suited for certain meanings, since the movements of the tongue and mouth 'imitate' or resemble what the word means. For example, [r] is said to naturally express 'rapidity' and 'motion', and [o] 'roundness'. The proposed sound-meaning resemblances are then exemplified by 'appropriate' words, such as *góngylon* ('round'). Hermogenes then contests this by giving examples of words containing the same sounds, but whose meaning is quite different. In the end of the dialogue, the two debaters find themselves in a stalemate; some sound forms seem indeed to 'resemble' their meanings, while many others do not. They also agree that "in order to gain knowledge of reality […] one should study reality as such, instead of taking recourse to the study of language." (De Cuypere 2008: 14f)

Albeit far removed from the concerns of modern linguistics and semiotics, *Cratylus* shows that the basic positions of the nature-

---

[2] See Sedley (2003) for a recent comprehensive monograph on *Cratylus*.

[3] Cratylus and Hermogenes were historical persons, but the opinions expressed in the dialogue were ascribed to them by Plato.

convention debate were set early in Western thought. Several other characteristics also seem to have lived on during the centuries. Firstly, we can note the vague terms in which the *phýsei* position tends to be presented ('natural connection', 'imitation'). Secondly, the conventionality of sound-meaning pairings tends to be construed as an antithesis to there being any substantial resemblance (or other motivated relation) involved; the step from here to 'arbitrariness' is small. In fact, conventionality and motivatedness are far from antithetical. We return to this below, but let us here only mention the fact that approximately half of the signs of American Sign Language, which is undoubtedly a conventional semiotic system, are judged to be iconic, that is, their shapes resemble their meanings (Woll, Kyle 2004). Thirdly, the dialogue's non-conclusive outcome has become a recurrent theme; given the extremeness of the positions, combined with vague formulations it is hard to see how it could have been otherwise.

2.2. Saussure, structuralism and 'the arbitrariness of the linguistic sign'

As a key representative of both linguistic structuralism and semiotics, Ferdinand de Saussure strived to move away from a 'nomenclatural' view of language, according to which words are attached to things and ideas that exist independently of language and human thought. Instead, the linguistic sign was claimed to involve "not a link between a thing and a name, but between a concept and a sound pattern" (Saussure 1959[1916]: 66).[4] The linguistic sign thus constitutes one

[4] As well-known, Saussure's most famous publication *Cours de linguistique générale,* derives from his students' lecture notes based on Saussure's lectures at the University of Geneva 1906–1911. Current research shows that these may have misrepresented his thinking in various respects (Bouissac 2010).

"psychological entity" whose two parts are intimately conjoined and mutually dependent, each automatically evoking the other.

Saussure claimed that the "primordial principle" of the linguistic sign is its *arbitrariness*. By this, he meant that a sound pattern and a concept have no "natural connection" that motivates them to be linked together in a linguistic sign. This is to be shown by observing that sound-concept links differ between languages, for example, the concept HORSE is linked to the sound patterns [hɔːɹs] in English, [ʃəval] in French, and [kuːɲ] in Czech. Neither sound pattern seems to have any inherent 'equine' qualities. Saussure concluded: "Because the sign is arbitrary, it follows no law other than that of tradition, and because it is based on tradition, it is arbitrary" (Saussure 1959[1916]: 74). One may wonder about the validity of the second sentence: does cultural transmission ('tradition') necessarily imply arbitrariness? The conflation of the features 'conventionality' and 'arbitrariness' when it comes to linguistic signs has been commonplace, both before and after Saussure,[5] but this hardly guarantees its correctness.

Another basic tenet of structuralism was that signs are delimited and defined by their contrast with other signs in the specific language system. Saussure (1959[1916]: 115f) mentions how the English word *sheep* and the French *mouton* may have the same signification, but whereas *sheep* contrasts with *mutton* in English, in French *mouton* can denote both the living animal and its meat; hence *sheep* and *mouton* have different meanings. Such language-specific 'structure', even more prominent in grammar than in lexical semantics is indeed an important characteristic of human languages. But is it also 'arbitrary'? Saussure, and many in his footsteps seemed to believe so — but on what ground? Finally, Saussure acknowledged the existence of onomatopoetic words which seemed to contradict the principle of

[5] Compare: "Every existing form of human speech is a body of arbitrary and conventional signs for thought, handed down by tradition from one generation to another" (Whitney 1867: 32).

arbitrariness, but downplayed their role by stating that "such words are never organic elements of a linguistic system. […] Moreover, they are far fewer than is generally believed" (Saussure 1959[1916]: 69f).

Over the course of the 20th century, counterarguments from linguistic functionalists, based on a wealth of linguistic evidence, have made many of these claims difficult to maintain. However, before turning to these, it is important to remember that Saussure's point of view was in line with the linguistic consensus of the time. In 19th century Europe, linguistics as a field was to a large extent occupied with comparisons and reconstructions within the Indo-European language family. An exaggerated belief in the methods of comparative linguistics led linguists to believe that, through the application of sound changes, human language could be traced back to its very origin.

In this pursuit, onomatopoeia was considered too crude to have been able to give rise to the number and variety of related word roots existing across the Indo-European languages. Max Müller called onomatopoeia a "chaotic anarchy", and in his popular *Lectures on the Science of Language* he wrote that "[the] onomatopœic system would be most detrimental to all scientific etymology and no amount of learning and ingenuity displayed in its application could atone for the lawlessness which is sanctioned by it" (Müller 1864: 94). The historical changes under study were "the true natural growth of words", whereas onomatopoetic words were "[…] artificial flowers, without a root. They are sterile, and are unfit to express anything beyond the one object which they imitate" (Müller 1869: 361). Whitney drew the concise conclusion that "[there is no inner or] essential connection between idea and word […] in any language upon earth" (Whitney 1867: 32).

Not only did research focus almost exclusively on Indo-European languages, many also considered these to be superior to other languages. Inflecting languages such as Greek, German, Latin and Sanskrit were thought to be the best tools for expressing human thought, and they were thus placed on the highest stage of linguistic development (see Losonsky 1999). As an example, Crowther's gram-

mar of Yoruba deemed the lack of inflection in the Yoruba verb a "negative character" (Crowther 1852: 12). Joh. Bernhard Schlegel made an important contribution in his description of the Kwa language Ewe, but at the same time he noted that the language "still finds itself in an initial stage".[6]

These mutually reinforcing ideas led to what perhaps was the main problem, namely that the linguists of that time worked with a 'skewed database', in which the data came from only a fraction of the world's thousands of languages. The introduction to Saussure's *Course* states that one aim of linguistics should be "to describe and trace the history of all observable languages" (Saussure 1959[1916]: 6), but the book itself mentions only a handful non-Eurasian languages, none of which merit any illustrative examples.

The 20th century saw a wealth of new branches of linguistics, and most linguists have recognized the importance of a diverse database of languages. However, since much of this development has taken place within a Saussurean framework,[7] the question of sound symbolism "[has been] reduced to the status of curious but irrelevant oddities in modern linguistics in general" (Hamano 1998: 211).

2.3. Cross-modal sound symbolism

In opposition to the dictum (sometimes approaching a dogma) of 'the arbitrariness of the linguistic sign', the notion of *sound symbolism* was used during most of the 20th century to highlight the existence of a

[6] Schlegel 1857, quoted in http://ideophone.org/early-sources-on-african-ideophones-schlegel/.

[7] For example, in a recent textbook, Yule (2006: 10) boldly states "It is generally the case that there is no 'natural' connection between a linguistic form and its meaning. […] There are some words in language with sounds that seem to 'echo' the sounds of objects or activities and hence seem to have a less arbitrary connection. English examples are *cuckoo*, *CRASH*, *slurp*, *squelch* or *whirr*. However, these onomatopoeic words are relatively rare in human language."

diverse array of linguistic phenomena in which "a sound unit such as a phoneme, syllable, feature, or tone [can] go beyond its linguistic function as a contrastive, non-meaning-bearing unit, to directly express some kind of meaning" (Nuckolls 1999: 228), or a "direct linkage between sound and meaning" (Hinton *et al.* 1994: 1). This has, however, been a minority position, and as Hamano (1998: 3) remarks "[discussion] of sound symbolism in natural languages often trigger laymen's curiosity but linguists' skepticism".

The main argument in favor of sound symbolism is that resemblances between 'sound patterns' and meanings go beyond clear, but admittedly peripheral, cases of onomatopoeia such as the English words *meow* and *bang*. In this case there is indeed a degree of resemblance between the words' sound patterns and the sounds produced by certain animals or inanimate objects. This is uncontroversial since both expression and content involve the same sensory modality: audition. However, the number of non-arbitrary signs in the languages of the world would increase dramatically if similarities *across* modalities can also be considered, or what is sometimes called 'synesthetic sound symbolism' (Hinton *et al.* 1994). Let us consider in order two different kinds of phenomena that bare evidence for this: first existing expressions in the lexica of a variety of languages, and then experimental studies with fictive or unknown expressions, showing speakers' ability to be able to actively find such cross-modal correspondences.

2.3.1. Ideophones

A huge number of expressions that are claimed to display a non-arbitrary relation between expression and meaning are attested in languages on all continents, but are perhaps best known from West Africa and East Asia. Different terms are used in different scholarly traditions, but the most common ones are *ideophones* in Africanist literature (Childs 1994), *expressives* for Austroasiatic languages (Diffloth 1976; Svantesson 1983), and *mimetics* for Japanese and Korean

(Hamano 1998; Sohn 1999; Ivanova 2006). Such terminological profusion is confusing, since the phenomenon referred to appears to be the same, as illustrated below. To avoid confusion, we will consistently use the term 'ideophones' in the remainder of this article.

As examples of ideophones, consider the Japanese expressions *koro-koro* 'small object rolling continously' and *goro-goro* 'large object rolling continuosly' (Kita 2008). We may observe several typical features: (a) reduplication, complete or partial, is common though not necessary characteristic, (b) there are often, but again not always, contrasts in the sound-forms corresponding to differences in meaning, (c) to the extent that there is a resemblance between sound-patterns and meanings, this is at a rather 'abstract' level: it is not the sound of 'objects rolling' that is being imitated, and the contrast between [k] and [g] can hardly be said to correspond to an *auditory* feature related to size.

In fact, a class of Japanese ideophones (mimetics) sometimes called 'psychomimes' (Baba 2003) are even said to express mental states, and it is far from clear how this can be regarded as a relation of 'resemblance' or 'imitation' between sound patterns and meanings (see Ikegami, Zlatev 2007). (1) shows one such example, borrowed from Kita (2008: 31).

(1)	Taro-wa	*sutasuta*-to	haya-aruki-o	si-ta
	Taro-TOP	IDPH-COMP	haste-walk-ACC	do-PAST
	'Taro walked hurriedly'			

To consider a different language, (2) and (3) used by Dingemanse, show ideophones in the Niger-Congo language Siwu spoken in Ghana. The examples are taken from spontaneous discourse.[8]

[8] Dingemanse, Mark. 'How to do things with ideophones: observations on the use of vivid sensory language in Siwu'. Paper presented at the SOAS Research Seminar, June 3 2009, London.

(2) ɔ̀-ɔ̀-fiɛ mùnyɛ̀mùnyɛ̀
3SG-FUT-shine IDPH.sparklingly
'It will shine sparklingly'

(3) ɔ̀-to ɔ̀-tu kɔkɔ itì tsɔ̀kwɛ́tsɔ̀kwɛ́
3SG-PROG 3SG-cut fowl head IDPH.sawing
'He is cutting off the fowl's head (in a sawing or cutting manner)'

Again, we notice features (a) and (c) listed above (Dingemanse does not mention any contrasts in these particular examples), and we could perhaps add two more: ideophones are rather hard to translate, and do not lend themselves easily to paraphrases with other expressions. Dingemanse, for example, describes Siwu ideophones as "typically evoking a sensory event as a whole rather than describing just one aspect of it". For example, in (2) above, *mùnyɛ̀mùnyɛ̀* is not simply an intensifying word such as 'very', and it can only be used with verbs such as *fiɛ* ('sparkle', 'shine'). Finally, when speakers of Siwu are asked about their perceived function of such ideophones, they typically reply that "without them, speech is bland", and that they "make stories more interesting". Similar accounts of ideophones are also given by Japanese speakers.

Several dictionaries of Niger-Congo and Bantu languages contain thousands of entries for ideophones (Childs 1994). Diffloth (1976) reports similar numbers from Austroasiatic languages. There are dictionaries of Japanese especially devoted to ideophones, and they are also found in many ordinary dictionaries (Ivanova 2006). One preliminary conclusion is that ideophones are indeed *not* "far fewer than is generally believed", as Saussure claimed. Also the pure fact that there *could* be dictionaries of such items, shows, firstly, that despite what was suggested above, their meaning can be approximately rendered through other linguistic expressions, and they are not as "sterile, […] unfit to express anything beyond the one object which

they imitate" as Müller (1869: 361) claimed. Secondly and more importantly, their possible presence in a dictionary shows that they are truly *conventional*, that is, part of the common knowledge, shared by the speakers of the respective languages (see Itkonen 2003, 2008; Zlatev 2007). Since by all accounts they are not at the same time 'arbitrary', ideophones clearly show the mistake of conflating conventionality and arbitrariness.

At the same time, their *non*-arbitrariness is rather difficult to pinpoint. As shown above, the statement that there is a 'synesthetic' relation between expression and content 'across modalities' is not altogether clear. Which modalities? Is this a relation of similarity (resemblance) and if so, along which dimensions can such cross-modal similarity be established? Furthermore, since ideophones are as was pointed out *conventional*, it is not obvious that any potential similarity plays any cognitive role in their synchronic use. To address these issues we need to turn to psychological evidence.

2.3.2. Evidence for 'on-line' mapping between sound patterns and meanings

Since the 1920s, numerous experiments have been performed to show that when people are asked to match certain kinds of meanings with fictive or unknown word forms, the outcome is significantly higher than chance. Sapir (1929) investigated what he called 'phonetic symbolism' among 500 English speakers. He presented participants with fictive word pairs differing only in their vowel — for example, *mil* and *mal* — and two contrasting meanings to the words in each pair, for example, 'small table' and 'big table'. The results showed that in 80% of the cases the participants assigned the word containing [a] to the larger object, and [i] to the smaller one.

The same year, Köhler (1929) introduced the fictive words *takete* and *baluma* (later changed to *maluma*), which he predicted would be

matched with a pointy, angular figure, and a smooth, wavy figure, respectively. His own experiments, and later those of others confirmed Köhler's hypothesis with results showing that speakers of different languages do tend to match words and such figures far more often than chance predicts (Nuckolls 1999; Westbury 2005; Hunter-Smith 2007). The phenomenon has been called "shape symbolism".

The actual word pairs used in these experiments have varied, but in general, one word contains voiceless stops [p, t, k] and closed unrounded vowels [i, e], and the other contains voiced stops [b, d, g], sonorants [m, n, l, ŋ] and open rounded vowels [u, o, ɔ]. Firth (1957) mentions *kikeriki* and *oombooloo*, and Ramachandran and Hubbard (2001) introduced *bouba* and *kiki*, which have become the latest 'standard'. In their influential study, 95% of the participants matched *bouba* with the round, wavy figure, and *kiki* with the pointy, angular figure. The explanation of this phenomenon has, however, again been given in rather vague terms: "[…] the representation of certain lip and tongue movements in motor brain maps may be mapped in non-arbitrary ways onto certain sound inflections and phonemic representations in auditory regions" (Ramachandran, Hubbard 2001: 10). This description is in line with Ramachandran's theory of the phenomenon of *synaesthesia*, the involuntary linking of certain senses due to innate "cross-wiring". But as Ikegami and Zlatev (2007) observe, such "shape symbolism" is qualitatively different from true synaesthesia in not being involuntary: seeing certain figures is not accompanied automatically with certain sound perceptions, or vice versa. Furthermore, a condition for the success of such cross-modal mappings is that figures and sound patterns can be found to differ along the same qualitative dimension. In this particular case, this could be said to be the dimension ROUNDNESS (or conversely POINTEDNESS). In the studies performed by Sapir, it was SIZE.

In a related type of experiment, participants have been given two contrasting words (*small–big*, *round–flat*, etc.) in a familiar language and two words in an unfamiliar language, and asked to attempt to

match these. Studies of this kind have been conducted since the 1930s, involving several unrelated languages (Hunter-Smith 2007: 23f). Brown *et al.* (1955) asked English speakers to match word pairs in English with Chinese, Czech and Hindi. For all three unfamiliar languages, over half of the word pairs were matched correctly significantly better than chance, sometimes by over 90% of the participants.

Two conditions for the success of such experiments could be observed. First, participants are to match a pair of *familiar* words, with a pair of unfamiliar ones. Second, the familiar words should contrast along a given dimension (as noted above for "shape symbolism") and thus form antonym pairs, for example, *small–big*, *round–flat*, *bright–dark*. For example, when Maltzman *et al.* (1956) performed a study in which English speakers were asked to match antonyms in Croatian and Japanese (both of which were unknown to the speakers), they failed to perform this at levels higher than chance. When they were also given the antonyms in English, they performed significantly better than chance. In another study, Brackbill and Little (1957) did not use antonyms, but rather a list of highly frequent words in Chinese, Japanese and Hebrew. English speakers were asked to match given word pairs across languages, but failed.

In the next section, we offer an explanation of these findings, but let us first take stock.

2.4. From sound symbolism to cross-modal iconicity

As shown in this overview, the age-long debate on whether linguistic signs are fundamentally conventional or 'natural' has been troubled with conceptual unclarity and unnecessary polarization, both of which increased when conventionality was equated with arbitrariness at the end of the 19th century, most famously by Saussure. On the one hand, it is impossible to deny that linguistic signs are semiotic conventions, or even norms, since they are commonly known and there are public

criteria for their correct application, as shown by the possibility of describing them in dictionaries (Itkonen 2003, 2008). However, this does not exclude that their origin may be due, at least in part, to non-arbitrary criteria such as partial similarity between expression and content, and that this similarity is maintained, at least for certain (classes of) expressions and more so for some languages than others.

On the other side, proponents of sound symbolism, from *Cratylus* onwards, seem to have overstated their claims. Even if some form or degree of resemblance (between sound pattern and meaning) can be established for some expressions, and more so in some languages than others, such resemblance would not be sufficient for understanding what the sound patterns mean, if the pairings were not *also* conventional. Even the most sound-symbolic linguistic expressions are not like realistic drawings (pictures), were the resemblance relation is indeed sufficient, as we will explicate in the next section.

What seems to have been lacking is a consistent conceptual framework in which to analyze the non-arbitrariness of (some) linguistic signs. To this testifies the loose definitions of the phenomenon: it is not sufficiently explained in which ways sound patterns are supposed to 'imitate', 'directly express' or 'resemble' meanings, especially if this is to take place across sensory modalities.

In the next section, we propose to provide the outlines of such a conceptual framework, borrowing seminal concepts from the other "founding father" of semiotics, C. S. Peirce, (the first being Saussure) interpreted and further developed by G. Sonesson (1989, 2007, 2009b). In such a perspective, the term 'sound symbolism' becomes less adequate, since for Peirce a 'symbol' is a sign that expressly *lacks* any motivated relation ('ground') between expression ('representamen') and content ('object'). Since the professed relation between expression and content observed in ideophones and cross-modal mappings (reviewed in Section 2.3) is that of similarity (resemblance), a more appropriate term for the phenomenon appears to be *cross-modal iconicity*.

3. A semiotic analysis of cross-modal iconicity

3.1. Signs and their ground

Semiotics is usually characterized as "the study of signs",[9] but this is hardly adequate since semiotics studies meanings of all kinds — from those inherent in direct perception and sensorimotor experience, to cultural phenomena such as music and dance, to visually perceived representations like pictures and gestures, to language, and its variety of meanings, on the levels of prosody, lexis, grammar and discourse. If all of these are characterized as 'signs', the concept becomes extensively broad (see Sonesson 1989, 2007). Hence, we find it better to regard semiotics as "the systematic study of meaning" (Fuller 1997: 30; compare Sonesson 2009a; Zlatev 2009). Within this study, the concept of sign nevertheless retains a crucial place, and the question of how the capacity to use signs emerged in evolution, and develops in ontogeny is of vital importance (Donald 1991, 2001; Mandler 1996; Deacon 1997; Tomasello 1999, 2008; DeLoache 2004; Sonesson 2007, 2009a; Zlatev 2007, 2008, 2009).

Peirce defined his central concept many times, but the following is perhaps the most famous one:

> [a] sign, or *representamen*, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. The sign stands for something, its *object*. It stands for that object, not in all respects, but in reference to a sort of idea which I have sometimes called the *ground* of the representamen. (CP 2.229)

The sign (as a whole) can be said to involve at least three entities interacting in the process of semiosis (see Fig. 1): a *representamen* which stands for an *object*, "to somebody" (the *interpreter*). The

[9] For example, Chandler, *Semiotics for Beginners*, http://www.aber.ac.uk/media/Documents/S4B/

representamen gives rise to an effect in "the mind of that person", namely an *interpretant* (De Cuypere 2008: 30ff; Parker 1998: 144ff). It is crucial that the representamen and the object are related only "in some respect or capacity" and "not in all respects"; otherwise they would be identical and inseparable. This basis of relating the representamen to the object is a fourth entity in semiosis, the *ground*.

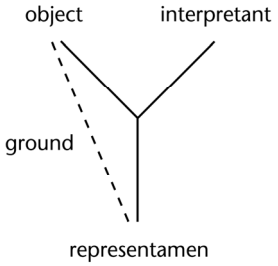

Figure 1. A schematic illustration of semiosis (N.B. in order to simplify the following figures, the interpretant will not be drawn).

Based on the nature of the ground — that is, in what way a representamen stands for an object — Peirce divided signs into three so-called "ideal types", the first of which is of central importance for our analysis. In an *iconic sign*, the ground is that of *similarity*, or more precisely when the representamen and object are found to share certain similar qualities independently of each other. The typical example is a picture and its visual similarity to that which it depicts. The second type is an *indexical sign*, where the ground is not based on similarity, but on contiguity in time and space. The third type is the *symbolic sign*. This differs from the earlier two by being based *only* on convention, for example, using $ as a sign for 'US dollar'. It is important to bear in mind that this taxonomy presents *ideal types*, and that real-world signs usually contain properties from more than one of the three, that is, more than one type of ground.

3.1.1. Primary and secondary iconicity

Since actual signs involve combinations of grounds, Sonesson (1997) introduced an important distinction between iconic signs. In *primary iconic signs*, the iconic ground is sufficient for establishing the sign, and any possible conventional ground plays a secondary role. The obvious example of this is again a drawing: while stylistic conventions of drawing differ enormously, knowledge of them is not required in order to recognize that a particular drawing depicts, say, a human face. On the other hand, if the sign is established by other means — by convention, by pointing to a referent while uttering its name (its representamen) or by simply telling the interpreter what "something means" — and any possible iconic ground is found by the interpreter only later (if at all), this would be a case of a *secondary iconic sign*.

> Primary iconicity: *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. 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, our emphases)

Sonesson illustrates primary iconicity with pictures, and secondary iconicity with so-called droodles: simple drawings whose ground becomes evident only once it is pointed out (Fig. 2). But the distinction is perhaps even more relevant in discussing iconicity in language.

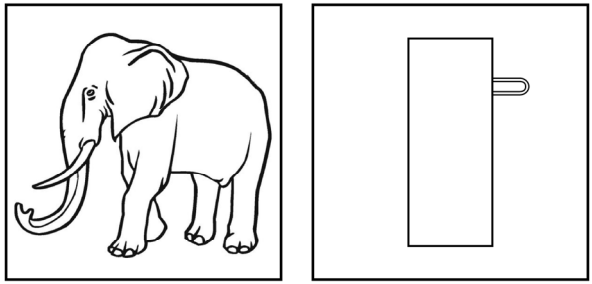

Figure 2. A primary iconic sign: a picture of an elephant, and a secondary iconic sign (a 'droodle') showing (for example) a person playing trombone in a wardrobe, or perhaps a paper clip placed under an envelope.

3.1.2. Iconicity in language

Jakobson (1971[1965]) first introduced the Peircean concept of iconicity into linguistics, and over the last 30 years, it has become a popular term in the cognitive-functional school of linguistics, whose representatives have applied it in analyses of syntax and morphology (Haiman 1980; Croft 2003; Itkonen 2004, 2005). At the same time, there has been persistent disagreement on how to apply the notion. For example, Croft (2003) claims that the principle "one expression — one meaning", that is, the tendency to limit polysemy (one expression — many meanings) and synonymy (several expression — one meaning) is a case of linguistic iconicity. Itkonen (2004) justifiably objects that this is an overextension of the concept, since expressions and meanings do not exist *independently of one another* (which, as pointed out above, is a requirement for iconicity), but constitute the two sides of the linguistic sign; else we would be back to the nomenclature view of language criticized by Saussure.

Already Jakobson suggested that plural endings illustrate an iconicity of quantity: longer expressions correspond to a larger quantity. On the other hand, Haspelmath (2008) objects that the phenomenon

is better explained through frequency of use: more commonly occurring types of word (the ones in singular) have shorter forms since this is simply more economical. There are problems to this analysis, but if we accept it, we could use Sonesson's distinction between primary and secondary iconic signs to explicate it: to the extent that singular/plural forms of nouns can be viewed as iconic to lower/higher quantity, such an iconic sign is only of the secondary type; it is established by convention, and the 'similarity' between expression and meaning (found by linguists) is epiphenomenal: it plays no role either in synchronic interpretation or in diachronic origin of the plural morphemes. Sonesson himself has made suggestions along these lines: "As for iconicity in language and in music, it most of the time seems to be secondary" (Sonesson 2009b: 51). De Cuypere (2008) argues for a similar position.

However, if this were altogether true, it would effectively devalue iconicity as an explanatory concept in linguistics. It would, for example, make the iconicity of ideophones analogous to that of droodles: possibly vicarious similarities that can be perceived only after the signs have been established by convention. However, there is something deeply unsatisfactory to this latter claim. On the one hand, the sheer number of ideophones in at least some languages seems to speak against it. Even more, it would make the 'on-line' cross-modal mappings that have been experimentally attested inexplicable. Therefore, Ikegami and Zlatev proposed that there is indeed more than secondary iconicity in language:

> We would like to suggest that the distinction between primary and secondary iconicity is more of a cline, defined by the degree to which the "sign function" (i.e. knowing what an expression represents) is necessary for perceiving the similarity involved in iconicity. From this perspective, the bouba/kiki phenomenon is somewhat intermediary in the cline. (Ikegami, Zlatev 2007: 270)

As it stands, however, this formulation is also unsatisfactory: it is not really clear how phenomena such as "shape symbolism" can be

regarded as lying "between" primary and secondary iconicity, since the two are defined in a complementary way. In the following, we provide a possible resolution to this dilemma.

3.2. Cross-modal iconicity as a combination of primary and secondary iconicity

Let us again consider the two types of experiments described in Section 2.3.2, those involving the matching between fictive words and figures, and those involving the matching between words in a familiar language and an unfamiliar one — and attempt to provide a semiotic explication of their findings in terms of the notions discussed in the previous sub-section.

3.2.1. Matching fictive words and figures

In the experiments carried out in the tradition inaugurated by Köhler and made famous by Ramachandran and colleagues more recently, there are always two representamina (the fictive words such as *bouba* and *kiki*) and two perceptual objects exhibiting some kind of contrast (round *vs.* pointy figures, small *vs.* big objects, etc.). The interpreter will succeed in performing the matching to the extent that he or she can discern a similarity, that is, an iconic ground, between the representamina and objects in order to create two signs. Schematically, this can be presented as a sequence of three steps (see Fig. 3). Step 1 shows the initial conditions in the experiment with two separate sets of representamina (R1, R2) and two objects (O1, O2). Step 2 shows that two *parallel* grounds are discerned (dashed lines), constituting a

composite analogous ground: R1 is to O1 as R2 is to O2.[10] On this basis in Step 3, the linking of representamina and objects is established, creating two (presumed) signs.

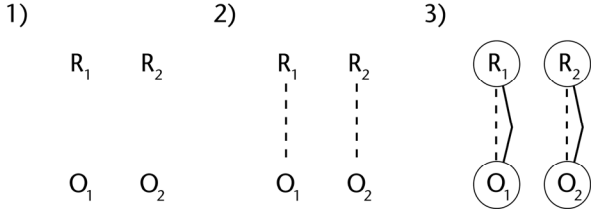

Figure 3. Matching fictive words to figures by finding a composite analogous ground.

Is this to be regarded a matter of primary or secondary iconcity? We propose that it is not one or the other, and not something "in between" (as suggested by Ikegami and Zlatev (2007)), but rather a combination of the two, as follows:

First, the interpreter is told that a combination of sign relations exists between the pairs of representamina R1 and R2 and perceptual objects O1 and O2. This is a precondition for discerning the composite analogous ground in Step 2. That is, knowledge of the existence of sign relationships, "is one of the reasons for the perception of an iconic ground", that is, conforming to the definition of *secondary iconicity*, given earlier.

Once this analogous ground is perceived, however, it serves as the basis for positing *specific* sign relations between R1–O1 and R2–O2 in Step 3. That is, quoting again Sonesson's definition of *primary*

[10] One R-O ground would be, strictly speaking, sufficient for performing the analogy if there are only two (presumptive) signs involved, as an anonymous reviewer points out.

iconicity (see above): "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".

Observe that the precondition for the success of the second step is that the contrastive relationship between the two expressions (representamina) on the one side is found to correspond to an analogous contrastive relationship between the two objects. This kind of *second-order* iconicity is characteristic of analogy (Itkonen 2005), that is, similarity of relations rather than objects. It also presupposes the characteristic feature of sign-to-sign contrast emphasized by structuralism, though not the tenet of "the arbitrariness of linguistic sign". Therefore, we may conclude that the primary iconicity involved in the process is *doubly conditioned* by, on the one hand, prior knowledge that there are (yet unspecified) sign relations (secondary iconicity), and the principle of linguistic contrast, on the other hand. We will return to this in what follows.

3.2.2. Matching familiar antonyms with ones in an unfamiliar language

In the case of matching two antonymic words R1 and R2 in a familiar language (for example, *large* and *small*) to corresponding words R3 and R4 in an unfamiliar language, we have no perceptual objects, but since the first two words are familiar, the interpreter will have recourse to their meanings via the sign relationship and can imagine, for example, a LARGE and SMALL object. This is shown as Step 1 in Figure 4. In Step 2, due to secondary iconicity, an analogous ground is discerned between representamina and objects. In Step 3, a similar analogous ground is found, R3–O1 and R4–O2, and on this basis a sign relationship is posited to exist between these (that is, primary iconicity) in Step 4. Thereby R1 is matched to R3, and R2 to R4.

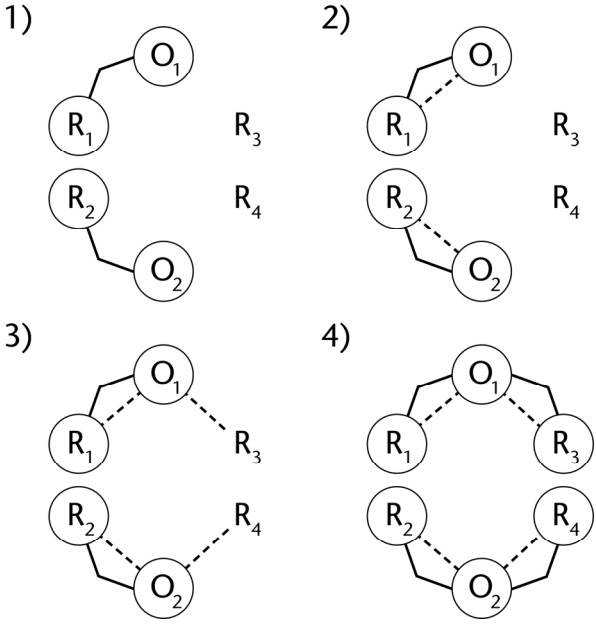

Figure 4. Matching familiar antonyms with those in an unknown language, by finding a higher-order analogy between two composite analogous grounds.

The rate of success in this type of experiments was typically lower than those in which fictive words were matched to figures,[11] which can be explained by the following two factors. First, the matching is performed across sets of composite grounds, that is, a third-order relation. Given this complexity, it is nevertheless remarkable that such experiments meet success at all! Second, all word-forms (representamina) are actual existing words which may or may not (for example, due to

[11] Some antonyms give very high degrees of correct matching, while others show results identical to chance. In some cases, different languages appear to have 'reversed' grounds, leading to results significantly *worse* than chance (Hunter-Smith 2007).

historical sound change) have a degree of iconic ground characteristic of fictive forms, explicitly chosen like *bouba* and *kiki* or *mil* and *mal* to exhibit (cross-modal) similarity with corresponding figures. On the other hand, this testifies that iconic grounds are in many cases preserved in languages, and many signs are indeed not arbitrary. This also suggests that cross-modal iconicity in language, far from being epiphenomenal, can under appropriate conditions have cognitive consequences, for example, in foreign language learning.

Again, however, one of the "appropriate conditions" is prior knowledge that what we have in the two languages are synonymous linguistic signs: remember that experiments were unsuccessful when two sets of representamina in unfamiliar languages were to be matched. Therefore secondary iconicity (in Step 1) was a precondition for the operation of primary iconicity (in Step 4). Also as pointed out in 2.3.2, participants were able to solve the task only when O1 and O2 were contrasted on the same quality dimension (SIZE, ROUNDNESS etc), but not when expressions were picked on factors such as frequency. Finally, the principle of inter-sign contrast was assumed here as well, as part of the design of the experiment, and the instructions given to participants.

3.3. Summary and further questions

In this section, we showed how moving beyond the vague formulations of "sound symbolism" and adopting a semiotic framework distinguishing between primary and secondary iconicity allowed us to make sense of experimental results, and to confirm the cognitive and linguistic reality of a degree of primary iconicity (in combination with secondary iconicity and inter-sign contrast along uniform dimensions). Furthermore, discerning the (composite) iconic grounds necessary for performing tasks such as those that were here explicated presupposes the ability to find similarities across the modalities of hearing (and speaking) and at least the modality of vision.

Many questions, however, remain. How is such cross-modal mapping achieved? Do speakers perceive the form of their vocal apparatus, haptically and propioceptively, for example as "round" in forming the sound [o] when mapping to a roundish figure, or is the matching performed on a deeper, more abstract or even 'amodal' level? Is the matching performed on the basis of segments (such as phonemes), or even distinctive features, such as voiced/unvoiced, or is it rather a matter of transitions between segments and whole sound patterns? If it is a matter of segments, do vowels or consonants play a more determining role for establishing what was here called the iconic ground?

The conceptual explication offered here cannot help us to answer these questions, and we need to turn to so-called empirical methods. Since we are not aware of studies which have been designed explicitly to address questions such as those given above, we performed a case study of our own, which we describe in the following section.

4. A case study of cross-modal iconicity

4.1. Goals and method

We performed an experiment in the tradition of Köhler (1929) and Ramachandran and Hubbard (2001), described in Section 2.3.2, asking Swedish participants to match fictive words and visual figures. In contrast to earlier studies, we decided to utilize a larger range of word pairs, and a more systematic comparison of different types of sounds: vowels, consonants and combinations of these. The purpose of this was to attempt to discern the relative importance of vowels and consonants, as well as the patterns of these taken as a whole, for perceiving an iconic (cross-modal) ground that would allow the participants to 'correctly' match the fictive words and figures, as analyzed in Section 3.

First, we selected maximal phonetic contrasts between consonants and between vowels, based on the 'synaesthetic' properties that have

been ascribed to such sounds in previous studies, and on the basis of our own intuitions.

Consonants:

voiceless obstruents ('hard', 'sharp', 'pointy')
vs.
voiced sonorants ('soft', 'smooth', 'heavy'):
[p, t, k, tʃ] *vs.* [m, l, n, ŋ]

Vowels:

front close unrounded ('sharp', 'small')
vs.
back open ('round', 'large'):
[i] *vs.* [u]

Second, we defined four pair-types of fictive words forming simple reduplicative patterns of the form CVCV, using these phonetic contrasts in different ways. These constituted four different conditions. Conditions (a) and (b) contrasted individual sounds only, while (c) and (d) contrasted combinations of these, as follows:

a) Two words with different vowels, but the same sonorant consonant, for example, *lili* vs. *lulu*.
b) Two words with different consonants, but with the same vowel [i], for example, *kiki* vs. *nini*.
c) Incongruent combination: a word with a 'hard' consonant and 'round' vowel was contrasted with a word with a 'soft' consonant and 'sharp' vowel, for example, *tutu* vs. *lili*.
d) Congruent combination: a word with a 'hard' consonant and a 'sharp' vowel was contrasted with a word with 'soft' consonant and 'round' vowel, for example, *titi* vs. *lulu*.

The actual pair-types used in the experiment are given in Table 1 in Section 4.3. The task for the participants was to match the 'words' in each pair of figures such as those shown in Figure 5, which were similar to those used by Ramachandran and Hubbard (2001), that is, a sharp pointy 'star'-like figure, and a roundish 'amoeba'-like figure.

Figure 5. The contrastive figures used in the experiment: the 'star' and the 'amoeba'. All figures were of these types, but with slight variation, so that no participant was shown two identical figures.

4.2. Hypotheses

We formulated a set of three conditional hypotheses, on the basis of the contrasts between the four types of word pairs (conditions).

H1. If the iconic ground between sound patterns and figures is above all discerned on the basis of contrast between 'soft' and 'hard' consonants, then conditions (b), (c) and (d) should give rise to a high rate of matching of the words with these consonants to the 'amoeba' and 'star' figures, respectively, while condition (a) should lead to a lower rate, or to no effect higher than chance.

H2. If the iconic ground is above all due to contrasts between 'round' and 'sharp' vowels, then a higher rate of 'correct' matching should be found in conditions (a), (c) and (d), and a lower rate or no effect at all in condition (b).

H3. If the iconic ground is due to the combination of consonants and vowels, (and possibly the transitions between them), then the congruent condition (d) should lead to the highest rate of 'correct'

matching, followed by the cases in which there was contrast in only vowels (a) and consonants (b), and least of all in the incongruent condition (c).

Our expectation was that H3 would find the strongest support.

4.3. Participants and procedure

In the experiment, 20 participants (9 female, median age: 25) were recruited among students at Lund University, and compensated with a small sum, corresponding to a lunch ticket. The task was administered through the software *E-prime*. The 20 participants were divided into four groups of 5 (numbered 1 to 4 in Table 1). Each participant in the 4 groups was given a specific version of the software, which presented the participant with one instance of each of the 4 word pairs conforming to the conditions (a), (b), (c) and (d), without hearing the same fictive word twice, as shown in Table 1.

Table 1. The spoken stimuli used in the experiment, ordered by word pair type (A to D) and participant group (1 to 4).

	A. Vowel C_2V_1 vs. C_2V_2		B. Consonant C_1V_1 vs. C_2V_1		C. Incongruent C_1V_2 vs. C_2V_1		D. Congruent C_1V_1 vs. C_2V_2	
1	lili	lulu	tʃitʃi	mimi	pupu	ŋiŋi	kiki	nunu
2	mimi	mumu	pipi	ŋiŋi	kuku	nini	titi	lulu
3	ŋiŋi	ŋuŋu	kiki	nini	tutu	lili	tʃitʃi	mumu
4	nini	nunu	titi	lili	tʃutʃu	mimi	pipi	ŋuŋu

When the target figures (see Figure 5 above) were presented on the screen, randomized for left-right order, a pre-recorded voice was

heard saying (in Swedish) a sentence conforming to the schema: "Let's say that one of these figures is called <Word X> and the other <Word Y>, which one would you call <Word X/Word Y>?" The fictive words for the variables <Word X> and <Word Y> were taken automatically from a list consisting of the word pairs given in Table 1, in a carefully arranged way. The order in which the elements of the word pair were played was randomized, so that the 'harder' sound pattern would sometime come first, and sometime second. Correspondingly, the question of which sound pattern was to be explicitly matched with which figure (and which left to be matched to the other by logical inference), half of the time concerned <Word X>, and half of the time <Word Y>. Participants had the possibility to repeat the spoken stimuli as many times as they wanted, by pressing a certain key. In our debriefing questions, however, none of the participants reported that they had done this.

The order in which the conditions (a), (b), (c) and (d) were presented was also randomized, and the figures were slightly varied in each of the four conditions, but always having a 'star'-like figure and an 'amoeba'-like figure. Their relative left-right position was also randomized, to fully eliminate any possible bias that could arise, for example, from the left-to-right direction of writing in Swedish. The response was to be given by pressing one of two keys on the keyboard.

Interspersed between these key tasks were 4 'filler tasks'. One consisted of the participant hearing a short sound clip in a foreign language, and then choosing from three alternatives which language it was. The other tasks were 'logical' questions, similar to those used in IQ tests.

The purpose of avoiding presenting a given sound pattern (fictive word) twice was that once the participant had matched a particular sound to a figure, this was likely to affect the decision on a subsequent trial, in which the fictive word would have participated in a different contrast. For example, if a participant's first word pair were *lili* vs. *lulu*, and *lili* was matched with the 'star'-like figure, and then the second

pair were *lili* vs. *kiki*, it is likely that the participant would have been disinclined to give the name *lili* to the 'amoeba' figure, despite the fact that *kiki* may sound 'sharper' to them in this context.

Each participant performed the experiment in a closed room, sitting alone at the computer. On completion, participants were asked a few 'debriefing questions' in a short interview carried out by the first author, in order to attempt to understand something about their reasoning when performing the matching task. These interviews were recorded, and key portions were subsequently transcribed.

4.4. Results

In order to evaluate the three hypotheses, the quantitative data was analyzed by pooling the replies from all 20 participants for each of the four conditions. The results are shown in Table 2.

Table 2. Results from the word-to-figure matching tasks, divided by conditions.

Condition	'star' figure	'amoeba' figure
A. Vowel contrast	[i] – 90% [u] – 10%	[i] – 10% [u] – 90%
B. Consonant contrast	[p, t, k, tʃ] – 80% [m, l, n, ŋ] – 20%	[p, t, k, tʃ] – 20% [m, l, n, ŋ] – 80%
C. Incongruent combination of V and C	[u]+[p, t, k, tʃ] – 65% [i]+[m, l, n, ŋ] – 35%	[u]+[p, t, k, tʃ] – 35% [i]+[m, l, n, ŋ] – 65%
D. Congruent combination of V and C	[i]+[p, t, k, tʃ] – 90% [u]+[m, l, n, ŋ] – 10%	[i]+[p, t, k, tʃ] – 10% [u]+[m, l, n, ŋ] – 90%

Statistical analysis (binominal distribution) of the results showed that the results for conditions (a), (b) and (d) corresponded highly significantly to the expected ('correct') matching (obstruents and [i] matched to the 'star' figure, sonorants and [u] to the 'amoeba' figure) ($p < 0.01$). In the case of (c) there seemed to be a certain bias for matching obstruents to 'star' and sonorants to 'amoeba', despite the fact that the vowel contrast should have 'pulled' in the opposite direction, as seen by the results in condition (a), but the bias for (c) was markedly lower ($p = 0.13$), and not significantly higher than chance.

Concerning H1, stating that the iconic ground would be perceived above all on the basis of the consonant contrast, the results from condition (b) and (d) were supportive, but those in (a) and (c) contradictive. Hypothesis H2 was most clearly contradicted by the results in both conditions (b) and (c), despite the fact that (a) showed that even a contrast in the vowels alone was sufficient for a very high rate of expected matching (90%). The strongest evidence that it was not *either* vowels *or* consonants that serves as the ground for matching to the figures was (c) — the only condition in which the rate of expected matching fell below significance level.

The results did not correspond entirely to H3, since the rate of expected matching in condition (d) was identical to that in (a) and nearly so in (b). But the finding that vowels and consonants play a role on their own, and that they significantly lowered the matching rate in our 'incongruent combination' (c), allows us to conclude that H3 was the hypothesis most clearly supported among the three. Therefore, we may conclude that the results showed that *both* vowels and consonants independently, and in combination, contribute to establishing the iconic ground in cross-modal iconicity of the type investigated in the study.

4.5. Discussion

While our relatively low number of participants seemed to give rise to a ceiling effect of sorts — leveling out the possible differences between conditions (a), (b) and (d) — it is not implausible to conclude that the iconic ground was due to the whole sound pattern gestalt, rather than to individual segments. Consider one example of word pairs from the congruent condition (d). The sound patterns [kiki] and [mumu] contain sounds whose articulations differ greatly: [k] requires non-vibrating vocal chords, and a blocking of the vocal tract, followed by a release of air. The vowel [i], on the other hand, requires a steady flow of air from the lungs, passing through vibrating vocal chords. These differences result in an abrupt transition between sounds in the sound pattern [kiki]. This can be seen in the first spectrogram in Figure 6. Inversely, the sounds [m] and [u] are very similar to each other, leading to a less abrupt transition. Both [m] and [u] are pronounced with vibrating vocal chords, and [m] involves a closing of the lips, interfering minimally with the tongue's position for [u]. In [mumu], the lips are rounded and repeatedly pressed against each other, adding to a haptic-proprioceptive 'softness' of the sound pattern.

A broad qualitative analysis of the debriefing interviews showed that participants had some awareness of the cross-modal similarities between figures and sound patterns, but (unsurprisingly) did not formulate this with sufficient meta-linguistic knowledge to allow us to judge whether they focused on segments (vowels and consonants) or on the whole patterns. The majority of the replies were such as those given in (4)–(7) (translated from Swedish), in which the participants mention "sounds" more generally, while only (8) mentions "consonants" and (9) is rather indeterminate.

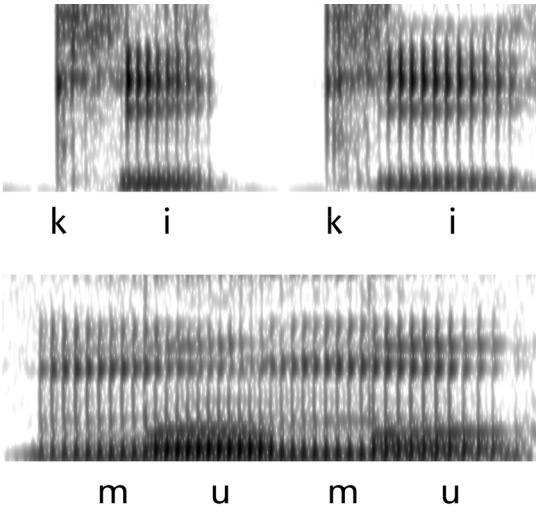

Figure 6. Spectrograms of the fictive words [kiki] and [mumu], as used in the experiment. The white part in the middle of [kiki] is silence (that is, a blocking of the vocal tract before a release of air). These spectrograms show 0–5,000 Hz and are 400 ms in duration.

(4) "I associated certain sounds with certain shapes."

(5) "One of the words sounded like something soft, and the other as something hard."

(6) "The figures looked like the sounds in their names."

(7) "I felt how the words felt and sounded."

(8) "Hard consonants are sudden sounds."

(9) "Hard sounds are shorter. Drawn out sounds are softer."

Interestingly, few of the participants spontaneously mentioned "vowels", despite the fact that in Swedish schools, for reasons of spelling rules, the vowels [i], [e], [y], [ɛ], and [ø] are usually called 'soft vowels' (*mjuka vokaler*), whereas [a], [u], [ʉ], and [o] are called 'hard vowels' (*hårda vokaler*). We were concerned that this fact could affect

the participants' performance, and in the debriefing we explicitly asked them whether they had thought about these terms during the experiment, but none claimed to have done so.

When asked to exemplify what they meant by 'sharp' and 'soft' sounds, participants replied with both consonants and vowels, in an expected manner. Voiceless obstruents were given as examples of 'hard' and 'pointy' sounds, whereas the voiced sonorants were described with adjectives such as 'soft' and 'round'. Although the vowels [i] and [u] were the only ones used in the experiment, participants sometimes offered other vowels to express the difference between 'sharp' and 'soft' sounds. These were vowels that usually followed a back/front division, so that for example, [o] and [ɔ̟] were considered 'soft', 'round' vowels, and [e] and [y] were considered 'pointy', 'hard' vowels.

Concerning the still more difficult question how the participants managed to establish the similarity between figures and sound patters, the descriptions of the participants were not detailed enough to be able to judge, though the reply given in (7) suggests that haptic (or proprioceptive) perception of the vocal tract during actual pronunciation was important. Indeed, some of the participants reported to have pronounced the words to themselves before making their choices. For the time being, we cannot offer more to the 'how' question than the somewhat speculative account offered by Ikegami and Zlatev, describing the study performed by Ramachandran and Hubbard (2001) using the expressions bouba and kiki:

> If we start with the shapes, the cross-modal mapping between vision and touch would allow them to be perceived as "soft" and "sharp" [respectively], motivating the use of these quasi-synaesthetic metaphors as a natural way to describe these figures. From the side of the expressions the production of the velar stop /k/, even more so combined with the front, unrounded vowel /i/ involves obstructions and narrowings in the vocal tract, which can similarly be perceived as "sharp" and "edgy". On the other hand, the shape of the vocal tract and the lips in the production of /u/ in bouba are quite literally "roundish" and

> the passage of air is "soft". The mappings between the senses Vision-Touch-Proprioception-Sound in internal meaning space thus provides for a correspondence between the shapes and the labels that would be impossible otherwise. A robot, or a Martian with a very different kind of body […] would not be able to perceive the iconicity involved. (Ikegami, Zlatev 2007: 225)

5. Overall summary and further research

We began this article by asking whether the cognitive semiotic approach — combining methods and theoretical concepts from the humanities and cognitive sciences — could help provide "a dialectical synthesis that could help resolve the debate between 'arbitrariness' and 'sound symbolism'. The three 'steps' that we took to address this question are rather typical for this approach. First, to take a historical perspective, looking at how the problem has been framed over the centuries, as well as more recently, and what kinds of arguments and evidence have been considered relevant. Second, to attempt to reframe the problem, in (preferably) clearer concepts, and thirdly, to make this reframing amenable to empirical investigations, in order to enter a cycle of cross-fertilization with more pronouncedly philosophical perspectives, such as those of semiotics and phenomenology (see Gallagher, Zahavi 2008; Sonesson 2007, 2009c; Zlatev 2009).

In the process, we found reasons to propose not a Grand Dialectical Synthesis, but several more minor ones, on all of these three levels. Let us here summarize our findings, relating, where appropriate, to questions deserving future research.

5.1. The historical perspective

As pointed out when summarizing Section 3, both extreme sides in the age-long (and continuing) debate have been in error: those claiming that language is *essentially arbitrary* (at least on the level of simple

signs), and those claiming something similar to what the early Wittgenstein famously stated: that language "is a picture of reality". The error of the first group, who we may (non-arbitrarily) call *arbitrarians* was due in part to the conflation of the concepts of convention and arbitrariness, and in part due to a 'skewed database' of linguistic evidence, disregarding non-European languages, and of course, the signed languages of the deaf, which were only recognised (in some quarters still grudgingly) as 'true languages' some 40 years ago. The error of the second group, *the sound symbolists*, was again due in part to lack of adequate concepts: just remember the profusion of terms used to describe the phenomenon: 'sound symbolism', 'shape symbolism', 'phonetic symbolism', 'mimetics', 'expressives', 'ideophones', 'psychomimes', 'synaesthemes' etc. in which sounds 'imitate', 'directly express', or 'resemble' meanings. Many sound symbolists also show clear traces of a binary logic (that can still be observed in some cognitive-functional quarters nowadays), contrary to that of the arbitrarians: since language is not arbitrary, it must be *essentially motivated*, with resemblance, or iconicity as a chief factor.

There have of course been many exceptions to such binary opposition between the two extremes (for example Givón 2002), some of these quite old, as can be seen in the following quote from Jespersen (1964[1922]: 397):

> Yes, of course it would be absurd to maintain that all words at all times in all languages had a signification corresponding exactly to their sounds, each sound having a definite meaning once for all. But is there really much more logic in the opposite extreme, which denies any kind of sound symbolism and sees in our words only a collection of wholly accidental and irrational associations of sound and meaning?

But such voices have somehow been dwarfed over the ages, it seems. With the widening of the linguistic database during the 20th century, Saussure's statement that sound-symbolic words "are never organic elements of a linguistic system", became increasingly hard to maintain. Hinton *et al.* (1994) contains several descriptions of sound symbolic

systems in languages from all over the world, including less 'exotic' languages including English (Rhodes 1994; Oswalt 1994). But we sense a new possible error: when sound symbolism can no longer be denied, there are tendencies to present it as an ordinary linguistic 'parameter', like word-order, or an 'areal feature', like the presence of noun classifiers or click consonants. For example, Svantesson (1983) classifies sound symbolic words in the Austroasiatic language Kammu as a distinct 'word class', with its own morphological, syntactic and intonational features. But why should symbolic words constitute a unique word class in languages like Kammu and Japanese, but not in English or Spanish? To regard it simply as a typological/areal feature is hardly explanatory.

To some extent, there is indeed evidence that (clearly) sound symbolic expressions are treated differently from those that are not, by their speakers and by processes of language change. Dingemanse (2009: 840f), for example, describes Siwu ideophones as displaying "deviant phonotactic patterns". Hamano (1998: 86) points out that Japanese sound symbolic words have retained the word-initial sound [p], which, through sound change, has otherwise become [h] in Japanese. This can be seen as the result of a perceived connection between the sound [p] and its sound symbolic connotations of 'an abrupt and explosive movement'. Even Jespersen (1964[1922]: 406) mentioned that English *cuckoo* [kʊku] has not undergone the regular sounds changes that would have resulted in [kʌku].

Kita (1997, 2001) has attempted to provide a theoretical account of such a 'dual code' system. On the basis of considerable empirical basis, he argues that the meaning of Japanese ideophones (mimetics) consists of an 'affect-imagistic dimension', where "language has direct contact with sensory motor and affective information" (Kita 1997: 320), while non-mimetic expressions such as quantifiers, logical operators, and the members of ordinary noun classes have their meanings in an 'analytic dimension' (compare Ikegami, Zlatev 2007).

There is a degree of appeal to such accounts, but again without additional explanations, it is mysterious why certain languages would rely heavily on the 'affect-imagistic' dimension compared to others. Furthermore, a strict separation between sound-symbolic and non-sound symbolic expressions would seem to be contradicted by results such as those reviewed in Section 2.3.2, where the meaning of 'ordinary' antonyms in unknown languages was in many cases 'guessed' correctly (when presented with corresponding familiar antonyms, and asked to match the two pairs).

What is much more plausible, we propose, is an evolutionary explanation. Iconicity is a key factor in the *emergence* of new expressions — as can be seen in signed languages (where new signs emerge from gestures), or on the basis of iconicity not between sound and meaning, but between the meanings of different expressions, as *(computer) mouse*. At the same time, with conventionalization, the role of iconicity diminishes — as again seen in signed languages, where 'arbitrary' signs are deemed to be at least as many as 'iconic' signs (Woll, Kyle 2004). Centuries of writing and language standardization in European languages would have further contributed to this, and language change provoked by language contact — perhaps even more. Many of the languages heavy on ideophones lack one or both of these characteristics: writing/standardization and extensive language contact. From this perspective, there is a 'substratum' of sound-symbolism in all languages, and the degree to which this is to be found may be due to factors such as those here suggested.[12] This is clearly a hypothesis that would deserve further investigation.

---

[12] Drawing from Ecuadorian Quechua, Nuckolls suggests another possible factor, namely that "ideophones thrive within animistic cultures" (Nuckolls 2004: 132f). Her proposal is largely that people in animistic cultures use ideophones to relate to the surrounding world much more profoundly than in Western non-animistic cultures: "I argue for a view of ideophones as a type of cultural discourse through which speakers *align* themselves with nonhuman life forms and forces of nature. This alignment is suggested by the special performative

5.2. The cognitive semiotic explication

By utilizing a cognitive semiotic framework, and in particular Sonesson's (1997) notions of primary and secondary iconic signs, we believe to have clarified the phenomenon of sound symbolism even more. But here, as well, we were lead to a sort of dialectics. On the one hand, we agree with Sonesson that out of context, sound symbolism would appear to be a matter of secondary iconicity — since few could understand the meaning of even the most transparent ideophone on first mention, without the help of context to "unlock its key". But at least in some cases — in particular when provided with two contrastive expressions and two contrastive meanings — this secondary iconicity becomes the ground for understanding previously unknown signs, that is, becomes transformed, so to speak, into primary iconicity.

This analysis helped us make sense of the successful results of studies of matching fictive words and visual figures, and of matching known and unknown antonyms, and also to interpret the reasons behind some of the unsuccessful ones. Somewhat surprisingly, the analysis vindicated one of the key notions of structuralism, the contrastive relation between linguistic signs. Just as meanings and sounds can be defined by what they are *not*, the key condition for 'unlocking' the iconcity of fictive and real words (transforming it from secondary to primary, as formulated in the previous paragraph) seems to be to allow them to stand in a contrastive relationship, along a single semantic dimension, as shown in Figure 7.

---

properties of ideophones, which collapse the distinction between a speech event and a narrated event, thus compelling a speaker to *become* an action, event, or process, in order to communicate about it" (Nuckolls 2004: 131, our emphases).

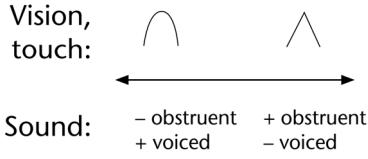

Figure 7. The dimension of SMOOTHNESS, with the two extreme ends 'smooth' and 'sharp', constituting the basis for matching visual and phonetic contrasts.

This could explain why there is less primary iconicity, even with the help of context, in (spoken) language than what one could assume from the hypothesis laid out at the end of the previous subsection. Far from all words in a language can be placed on one or another dimension, as in Figure 7. Adjectives, or other expressions of quality, are most susceptible to this, whereas nouns like *horse* obviously much less so. Hence, the prediction would be that even if such a non-gradable expression were initially strongly sound symbolic (iconic), this iconicity would not be preserved with conventionalization and historical change.

Indeed, Hinton and colleagues list a number of features that are likely to be expressed by sound symbolic words, namely "salient characteristics of objects and activities, such as movement, size, shape, color, and texture" (Hinton *et al.* 1994: 10). Note that all of these are qualities that are more or less gradient, that is, can be placed on dimensions such as SPEED, SIZE, ROUNDNESS, SMOOTHNESS etc. Dingemanse points out that sound symbolic words typically involve characteristics that are discernable by the six senses (sight, touch, hearing, interoception, taste, smell), as well as emotion, and — as we would expect from our emphasis on cross-modality, combinations of these.[13]

[13] Dingemanse, Mark 'What do we really know about ideophones?' Paper presented at the 6th World Congress of African Linguistics, August 21, 2009, Köln.

Typological studies such as these, however, have their limitations, since the processes in which the iconicity was established was sometime in the (indefinite) past, and we have no way of knowing whether it is 'alive' enough in the minds of speakers, or only a historical relic. This led us, as many others since Sapir (1929) and Köhler (1929) to experimentation.

5.3. Experiment

We found it lacking, especially from a linguistic perspective, that previous experiments with fictive word matching to figures simply picked two expressions that seems to their authors to be 'soft' vs. 'hard' such as *bouba* and *kiki*. Hence, we chose our sound patterns carefully, on the basis of systematic contrasts between vowels and consonants. Thus, we could formulate the question, utilizing the semiotic terminology introduced in Section 3: is it contrasts in vowels or in consonants that mostly lead to the perception of an iconic ground between sound patterns and figures? The dialectic answer was: both, and (most likely) their combination. Still, it was rather surprising that when consonants and vowels were combined 'incongruently' from a sound-symbolic perspective, there was a certain bias in the direction of the consonants. Perhaps this is not surprising, since it is the obstruent consonants that introduce the most salient contrast to all other segments, both the vowels and the sonorant consonants (see Figure 6). If we had instead used a less salient contrast for the consonants (for example, voiceless vs. voiceless obstruents) as the basis for a possible ground for matching to 'sharper' vs. 'softer' figures, the bias in favor of consonants would probably have been diminished. Such studies are an obvious further extension of our research.

Other extensions include the use of other dimensions such as SIZE (close vowels to small figures), and more challengingly, COLOR (different vowels are known to give rise to different colors in some

forms of synaesthesia, and the French poet, representative of the school of 'symbolism', Rimbaud (1986) claimed a particular mapping: "I invented the colors of vowels: A black, E white, I red, O blue, U green". But as we pointed out, synaesthesia is both involuntary, and known to give rise to 'arbitrary' associations due to 'cross-wiring' between the senses, which can be more or less idiosyncratic. COLOR is not a single dimension but at least two dimensions: HUE and LIGHTNESS. Whereas we would predict intersubjective agreement in a task involving matching sounds to the LIGHTNESS dimension (open vowels — 'dark', close vowels — 'light') it is much less obvious what the ground would be for perceiving similarities between sounds and dimension of HUE (the colors of the rainbow).

This discussion actualizes the point we made at end of Section 4: it is still not completely clear how the ground for cross-modal iconicity is perceived, that is, how we are able to establish similarities across different modalities. Figure 7 gives the impression that there is an abstract 'amodal' dimension, in this case of SMOOTHNESS, 'between' the senses of vision and sound, but this is doubtful, to say the least. We would prefer an account of step-by-step matching between senses, as suggested in the quotation from Ikegami and Zlatev (2007), but as admitted, this is still rather speculative. Cognitive semiotics, aided by neuroscience, would hopefully make progress in this direction with time. It is known that persons with lesions in the angular gyrus, known to be involved in cross-modal integration, have difficulties with performing the fictive words-to-figures matching task (Ramachandran, Hubbard 2001), and more recently it has also been strongly suggested that this also applies for people with autism (Oberman, Ramachandran, in press).

Lastly, we should mention the factor of writing. In designing the experiment, we took two steps to minimize any potential role of written language: the sound forms were presented orally, and their order in the pairs randomized, as well as that of the figures, to avoid a possible bias from the left-to-right direction of writing. But we can

nevertheless not exclude that (Swedish) writing may have played at least some role in performing the task,[14] since spelling is not completely 'arbitrary' either. The letters in the Latin alphabet for voiceless obstruents are indeed quite 'pointy', for example, <k> and <t>. Similarly for the letter for the front close vowel <i>. On the other hand, back open vowels tend to be 'round', <o> and <u>. Of course, the mappings are not completely systematic, since English also uses the arguably smoother <c> and <q> for the sound [k], which most Romance languages use as their standard. Still, we have no guarantee that such secondary iconicity could not be 'unlocked' in some contexts. Indeed, Westbury found an "interaction between visual form and phonology" (Westbury 2005: 16) in two experiments using Latin characters. It would be interesting to compare our results, performed with Swedish speakers, with results from languages using a script that does not show similar differences. In Burmese writing, for example, nearly all letters contain a semicircle, see Table 3.

Table 3. Table showing approximate representation of the sound sequences [bumo] and [kite] in different scripts.

Script	[bumo]	[kite]
Burmese	ဗုိမ်	ကိုင်တေ
Korean	부모	기데
Gujarati	બૂમો	ખીથે
Georgian	ბუმო	ქითე
Hiragana	ぶも	きて

[14] For example by visualizing the words spelled when hearing them. In one of the debriefing questions we asked whether the participants had done so. None reported to have done so, but… mysterious are the ways of the brain.

Another worthwhile comparison would be with children who have not yet learned the (Swedish) writing system. In a study of this sort, Maurer *et al.* (2006), found evidence for successful sound-figure matching even in pre-literate (English-speaking) toddlers, which is also what our cross-modal account would predict. But from what age the phenomenon appears still remains to be investigated.

To conclude, we hope to have shown that our cognitive-semiotic approach to sound-symbolism, and the notion of cross-modal iconicity can contribute to going beyond the debate with 'arbitrariness', and to help focus the discussion not so much on the existence of the phenomenon, but on its nature. In doing so, we are bound to move towards a more balanced and thus more adequate account of human language, shaped through the dialectics of convention and iconicity.[15]

References

Baba, Junko 2003. Pragmatic function of Japanese mimetics in the spoken discourse of varying emotive intensity. *Journal of Pragmatics* 35: 1861–1889.

Bouissac, Paul 2010. *Saussure: A Guide for the Perplexed*. London: Continuum.

Brackbill, Yvonne; Little, Kenneth B. 1957. Factors determining the guessing of meanings of foreign words. *Journal of Abnormal and Social Psychology* 54: 312–318.

Brown, Roger W.; Black, Abraham H.; Horowitz, Arnold E. 1955. Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology* 50: 388–393.

[15] We want to thank the research program at Centre for Cognitive Semiotics (CCS), Lund University, for providing the necessary intellectual environment for conducting this research, as well as for economic support for carrying out the empirical study. We wish also to express our gratitude to the Humanities Lab at Lund University for supplying equipment for conducting the experiment. Special thanks go to Joost van de Weijer for help with the statistical analysis. Finally, we gratefully acknowledge Göran Sonesson, Lars-Åke Henningsson, Timo Maran, and an anonymous reviewer for critical feedback on an earlier version of this article. We have followed their advice in many, but not all, respects and should therefore be held responsible for any remaining shortcomings.

Childs, George Tucker 1994. African ideophones. In: Hinton *et al.* 1994: 178–204.

CP = Peirce, Charles S. 1931–1958. *Collected Papers of Charles Sanders Peirce*. Cambridge: Harvard University Press. [Hartshorne, Charles; Weiss, Paul (eds.), 1931–1935; vols. 7–8. Burks, A. W. (ed.) 1958; In-text references are to CP, followed by volume and paragraph numbers]

Croft, William 2003. *Typology and Universals*. Cambridge: Cambridge University Press.

Crowther, Samuel Ajayi 1852. *A Vocabulary of the Yoruba language, Together with Introductory Remarks by the Rev. O. E. Vidal, M. A*. London: Seeleys.

De Cuypere, Ludovic 2008. *Limiting the Iconic: From the Metatheoretical Foundations to the Creative Possibilities of Iconicity of Language*. Amsterdam: John Benjamins.

Deacon, Terrance 1997. *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: Norton.

DeLoache, Judy S. 2004. Becoming symbol-minded. *Trends in Cognitive Sciences* 8: 66–70.

Diffloth, Gérard 1976. Expressives in Semai. *Austroasiatic Studies, Oceanic Linguistics Special Publication* 13: 249–264.

Dingemanse, Mark 2009. Ideophones in unexpected places. In: Austin, Peter K.; Bond, Oliver; Charette, Monik; Nathan, David; Sells, Peter (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory* 2. London: SOAS, 83–97.

Donald, Merlin 1991. *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge: Harvard University Press.

— 2001. *A Mind so Rare: The Evolution of Human Consciousness*. New York: Norton.

Firth, John Rupert 1957. *Papers in Linguistics 1934–1951*. London: Oxford University Press.

Fuller, Gillian 1997. Reading the readers: The role of semiotics in visitor research. *Visitor Behavior 12*(3–4): 30–31.

Gallagher, Shaun; Zahavi, Dan. 2008. *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. London: Routledge.

Givón, Tom 2002. *Bio-Linguistics. The Santa Barbara Lectures*. Amsterdam: Benjamins.

Haiman, John 1980. The iconicity of grammar: isomorphism and motivation. *Language* 56: 515–540.

Hamano, Shoko 1998. *The Sound-Symbolic System of Japanese*. Standord: CSLI Publications.

Haspelmath, Martin 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19: 1–33.

Hinton, Leanne; Nichols, Johanna; Ohala, John J. (eds.) 1994. *Sound Symbolism*. New York: Cambridge University Press.

Hunter-Smith, Sarah 2007. *Understanding without Babblefish: Reviewing the Evidence for Universal Sound Symbolism in Natural Languages*. Unpublished thesis from Swarthmore College.

Ikegami, Takashi; Zlatev, Jordan 2007. From non-representational cognition to language. In: Ziemke, Tom; Zlatev, Jordan; Frank, Roslyn M. (eds.), *Body, Language and Mind, vol 1: Embodiment*. Berlin: Mouton, 241–283.

Itkonen, Esa 2003. *What is Language? A Study in the Philosophy of Linguistics*. Turku: Turku University Press.

— 2004. Typological explanation and iconicity. *Logos and Language* 5: 21–33.

— 2005. *Analogy as Strucure and Process*. Amsterdam: Benjamins.

— 2008. The central role of normativity for language and linguistics. In: Zlatev, Jordan; Racine, Timothy P.; Sinha, Chris; Itkonen, Esa (eds.), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam, Philadelphia: Benjamins, 279–305.

Ivanova, Gergana 2006. Sound symbolic approach to Japanese mimetic words. *Toronto Working Papers in Linguistics* 26: 103–114.

Jakobson, Roman 1971[1965]. Quest for the essence of language. *Selected Writings II*. The Hague: Mouton, 345–359.

Jespersen, Otto 1964[1922]. *Language, its Nature, Development and Origin*. London: George Allen & Unwin Ltd.

Kita, Sotaro 1997. Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* 35: 379–415.

— 2001. Semantic schism and interpretive integration in Japanese sentences with a mimetic: a reply to Tsujimura. *Linguistics* 39: 419–436.

— 2008. World-view of protolanguage speakers as inferred from semantics of sound symbolic words: A case of Japanese mimetics. In: Masataka, Nobuo (ed.), *The Origins of Language, Unraveling Evolutionary Forces*. Tokyo: Springer, 25–38.

Köhler, Wolfgang 1929. *Gestalt Psychology*. New York: Liveright.

Losonsky, Michael (ed.) 1999. *Wilhelm von Humboldt: On Language: On the Diversity of Human Language Construction and its Influence on the Mental Development of the Human Species*. New York: Cambridge University Press.

Lyons, John 1968. *Introduction to Theoretical Linguistics*. London: Cambridge University Press.

Maltzman, Irving; Morrisett, L; Brooks L. O. 1956. An investigation of phonetic symbolism. *Journal of Abnormal and Social Psychology* 53: 249–251.

Mandler, Katherine 1996. *Language in Cognitive Development. The Emergence of the Mediated Mind*. Cambridge: Cambridge University Press.

Maurer, Daphne; Pathman, Thanujeni; Mondloch, Catherine J. 2006. The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental Science* 9(3): 316–322.

Müller, Max 1864. *Lectures on The Science of Language*. London: Longman, Green, Longman, Roberts & Green.

— 1869. *Lectures on The Science of Language, vol. 2*. New York: Charles Scribner and Company.

Nuckolls, Janis B. 1999. The case for sound symbolism. *Annual Review of Anthropology* 28: 255–282.

— 2004. To be or to be not ideophonically impoverished. In*: Proceedings of the Eleventh Annual Symposium about Language and Society — Austin, Texas Linguistic Forum* 47: 131–142.

Oberman, Lindsay; Ramachandran, Vilayanur S. (in press). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis. In: Pineda, Jaime A. (ed.), *Role of Mirroring in Social Cognition*. Totowa, New Jersey: Humana Press.

Oswalt, Robert L. 1994. Inanimate imitatives in English. In: Hinton *et al.* (eds.) 1994: 293–306.

Parker, Kelly A. 1998. *The Continuity of Peirce's Thought.* Nashville: Vanderbilt University Press.

Ramachandran, Vilayanur S.; Hubbard, E. M. 2001. Synaesthesia — a window into perception, thought and language. *Journal of Consciousness Studies* 8: 3–34.

Rhodes, Richard 1994. Aural images. In: Hinton *et al.* (eds.) 1994: 276–292.

Rimbaud, Arthur 1986. Collected Poems. Penguin Classics. [Oliver Bernard, ed.]

Sapir, Edward 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12: 225–239.

Saussure, Ferdinand de 1959[1916]. *Course in General Linguistics*. New York: The Philosophical Library.

Schlegel, Joh. Bernhard 1857. *Schlüssel der Ewesprache, dargeboten in den Grammatischen Grundzügen des Anlodialekts*. Stuttgart.

Sedley, David 2003. *Plato's Cratylus*. Cambridge: Cambridge University Press.

Sohn, Ho-Min 1999. *The Korean Language*. Cambridge University Press.

Sonesson, Göran 1989. *Pictorial Concepts*. Lund: Lund University Press.

— 1997. The ecological foundations of iconicity. In: Rauch, Irmengard; Carr, Gerald F. (eds.), *Semiotics Around the World: Synthesis in Diversity. Proceedings of the Fifth International Congress of the IASS, Berkeley, June 12–18, 1994*. Berlin & New York: Mouton de Gruyter, 739–742.
— 2007. From the meaning of embodiment to the embodiment of meaning: A study in phenomenological semiotics. In: Ziemke, Tom; Zlatev, Jordan; Frank, Roslyn M. (eds.), *Body, Language and Mind, vol. 1: Embodiment*. Berlin: Mouton, 85–127.
— 2009a. New considerations on the proper study of man — and, marginally, some other animals. *Cognitive Semiotics* 4: 133–168.
— 2009b. Prolegomena to a general theory of iconicity. Considerations on language, gesture, and pictures. In: Willems, Klaas; De Cuypere, Ludovic (eds.), *Naturalness and Iconicity in Language*. Amsterdam: John Benjamins, 47–72.
— 2009c. The view from Husserl's lectern. *Cybernetics and Human Knowing*, 16 (3–4): 107–148.
Svantesson, Jan-Olof 1983. *Kammu Phonology and Morphology*. Travaux de l'Institut de Linguistique de Lund 18. Lund: CWK Gleerup.
Tomasello, Michael 1999. *The Cultural Origins of Human Cognition.* Cambridge: Harvard University Press.
— 2008. *Origins of Human Communication.* Cambridge: MIT Press.
Westbury, Chris 2005. Implicit sound symbolism in lexical access: evidence from an interference task. *Brain and Language* 93: 10–19.
Whitney, William D. 1867. *Language, and the Study of Language: Twelve Lectures on the Principles of Linguistic Science*. New York: Charles Scribner's Sons.
Woll, Bencie; Kyle, James 2004. Sign language. *Encyclopedia of Language and Linguistics.* Oxford: Elsevier.
Yule, George 2006. *The Study of Language*. Cambridge: Cambridge University Press.
Zlatev, Jordan 2007. Embodiment, language and mimesis. In: Ziemke, Tom; Zlatev, Jordan; Frank, Roslyn M. (eds.), *Body, Language and Mind, vol 1: Embodiment*. Berlin: Mouton, 297–337.
— 2008. From proto-mimesis to language: Evidence from primatology and social neuroscience. *Journal of Physiology, Paris* 102: 137–152.
— 2009. The semiotic hierarchy: Life, consciousness, signs, language. *Cognitive Semiotics* 4: 169–200.

Межмодальная иконичность: когнитивно-семиотический подход к звуковому символизму

Все яснее становится, что положение Соссюра об «арбитрарном характере языкового знака» вступает в конфликт с широким распространением такого явления как «звуковой символизм». После обзора истории вопроса мы приходим к выводу о том, что обе позиции преувеличены и что удовлетворительного объяснения звукового символизма нет до сих пор. Как, например, можно ощутить сходство между выражением и содержанием в различных сенсорных модальностях? Наш ответ на этот вопрос опирается на понятие Пирса «основа иконичности» и на различение Г. Сонессоном первичной и вторичной иконичности. Далее мы описываем эмпирическое направление исследований, первым пионером которого был В. Кэлер и которое недавно вновь ввел в обиход В. Рамачадран. В этом исследований мы варьировали гласные и согласные в выдуманных словоформах и пришли к выводу, что звуки обоих типов играют существенную роль в восприятии основы иконичности между словоформами и визуальными образами. Комбинация исторического концептуального анализа, семиотических объяснений и психологических экспериментов, как это представлено в настоящей статье, является основой создаваемой парадигмы когнитивной семиотики.

Modaalsuste-vaheline ikoonilisus: kognitiivsemiootiline lähenemine häälikusümboolikale

Üha enam on mõistetud, et Saussure'i seisukoht „keelemärgi arbitraarsest loomusest" on vastuolus „häälikusümboolika" nime all tuntud laialt levinud fenomeniga. Pärast ülevaate esitamist selleteemalisest ajaloolisest debatist näitame, et mõlemad positsioonid on liialdatud ja rahuldav seletus häälikusümboolika nähtusele puudub jätkuvalt. Kuidas näiteks on võimalik tajuda sarnasust väljenduste ja sisu vahel erinevates sensoorsetes modaalsustes? Oma vastuses sellele küsimusele tugineme Peirce'i mõistele *ikoonilisuse alus* ja G. Sonessoni eristusele *primaarse* ja *sekundaarse ikoonilisuse* vahel. Lisaks sellele esitame oma empiirilist uurimust para-

digmas, mille esimeseks teerajajaks oli W. Köhler ja mida hiljuti on populariseerinud V. Ramachandran. Selles uurimuses varieerisime vokaale ja konsonante väljamõeldud sõnavormides ning jõudsime järeldusele, et nii konsonandid kui vokaalid mängivad olulist rolli sõnavormide ja visuaalsete kujundite vahelise ikoonilisuse aluse tajumisel. Ajaloolise kontseptuaalse analüüsi, semiootiliste seletuste ja psühholoogiliste eksperimentide kombinatsioon, nagu see on esitatud käesolevas artiklis, on aluseks kujunevale kognitiivse semiootika paradigmale.