PJOS 5(1), 2013

Motivations for Sound Symbolism in Spatial Deixis: A Typological Study of 101 Languages

Niklas Johansson[1] and Jordan Zlatev

Centre for Languages and Literature, Lund University, Sweden

We investigated possible motivations for sound symbolism in spatial demonstratives within 101 areally and genetically diverse languages. Six different predictions were formulated on the basis of factors such as (a) semiotic ground (iconic, indexical or combined), (b) speaker-centered, hearer-centered or both and (c) applicable to vowels, consonants or both. Each one of these six predictions resulted in different expected scales of phonemes on the proximal-distal dimension. Languages which conformed to these scales were regarded as *motivated* (according to a particular prediction). Languages which opposed it were treated as *reverse,* and if neither was the case, as *neutral*. The results showed significant motivated/reverse and motivated/neutral ratios only for the prediction based on vowel-frequency, motivated by a combination of iconic and indexical factors, and marginal support for the other predictions concerning vowels. The two predictions based on an assumed link between preverbal vocal pointing and demonstratives also found some, if weaker, support. The only prediction that was completely unsupported concerned the frequency of consonants. The conclusions are that a number of factors combine to motivate sound symbolism in spatial deixis, which appears to involve vowels more than consonants.

1. Introduction

The idea of the existence of a motivated, non-arbitrary relation between the sound patterns and the meanings of words, most often referred to as *sound symbolism*, has a long history, dating from Plato's famous dialogue *Cratylus.* It re-appeared in sound-imitation theories of language origins (Herder, 1772), and later as critical reactions (Jespersen, 1922; Jakobson, 1965) to the dominant thesis of the arbitrariness of the linguistic sign (Saussure 1916), with support from psychological experiments (Köhler, 1929; Sapir 1929). During the past two decades, the notion of sound symbolism has resurfaced in typological studies of phonesthemes and ideophones (Hinton, Nichols and Ohala, 1994; Dingemanse, 2012), as well as in a new flow of psychological studies (e.g. Ramachandran and Hubbard, 2001; Maurer, Pathman and Mondloch, 2006; Imai et al. 2008; Ahlner and Zlatev, 2010; Cuskley, this volume).

Yet, the factors that motivate sound symbolism are seldom made fully explicit and the empirical support for its existence is commonly regarded by skeptics as limited. Here, we attempt to address these reservations by analyzing the demonstrative pronouns in a sample of 101 languages with respect to six different predictions. These stem from suggestions made in previous research, that we believe have not been clearly distinguished. Typological studies of sound symbolism have on several previous occasions considered spatial deixis (Ultan 1978; Woodworth 1991; Traumüller 2000). We follow their path, but consider a wider range of possible motivations, and a larger language sample.

[1] Email: Niklas.Johansson@ling.lu.se, n.johansson@hotmail.com

In the literature on sound symbolism, terms such as "iconicity", "imitation", and "natural meaning" have been used interchangeably, often without providing clear definitions. Ahlner and Zlatev (2010) attempted to provide some clarifications, utilizing concepts from Peircean semiotics (e.g. Short 2007). Linguistic signs are semiotic conventions, i.e. they are commonly known in a speech community. However conventionality does not imply arbitrariness. Rather, conventions can be more or less *motivated*. A sign involves at least three interacting entities (see Figure 1): An *object* gives rise to a *representamen* which creates an *interpretant* in the mind of an interpreter. A forth notion, *ground* is what motivates the connection between the representamen and object.

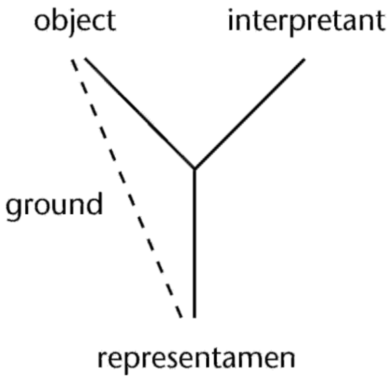

Figure 1. The basic concepts in Peirce's sign concept

Depending on the nature of ground, three "ideal types" of signs are commonly distinguished. In the *iconic sign* representamen and object share certain similar qualities independently of each other, e.g. a picture which depicts an object on the basis of visual similarity. The *indexical sign*'s ground is based on contiguity in time and space. Iconicity and indexicality may coexist in a particular sign: an animal's footprint is meaningful to a hunter on the basis of both kinds of ground. Finally, a *symbolic sign* is purely conventional. Linguistic signs are mostly symbols but also indices (linked to the situation of utterance) and may contain various types and degrees of iconicity.

In so-called "shape sound symbolism", researchers typically provide participants with two representamina, e.g. *takete* and *maluma* (Köhler 1929) or *kiki* and *bouba* (Ramachandran and Hubbard 2001) and two objects contrasting in shape, and ask the participant to make a match. The fact that people do this uniformly shows that they can indeed discern a similarity, or iconic ground, between the representamina and objects, even if they are given in different sensory modalities. Such cross-modal iconicity can proceed in several steps, as suggested by Ikegami and Zlatev (2007: 332).

> 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".

Furthermore, Ahlner and Zlatev (2010) manipulated systematically vowels and consonants in the representamina, showing that both sound-types contributed to mapping the novel words to the visual figures in a uniform way. Imai et al. (2008) showed that the phenomenon is more general than shapes, involving e.g. perceived similarities between the sounds and meaning of motion terms, and that this similarity is perceived by 2-3 year old children, and apparently utilized for language acquisition. A well-established dimension of similarity serving as the iconic ground in sound symbolism is that of SIZE (Sapir, 1929).

Another semantic domain likely to involve sound symbolism is spatial deixis. Demonstrative pronouns such as *this* and *that* are relative in distance to a so-called deictic center, which is usually the location of the speaker at the time of the utterance. Hence *this* is proximal, while *that* is distal. Three-way systems can either be distance-oriented or person-oriented. Person-oriented systems (e.g. Japanese) –where one of the terms is relative to the proximity of the speaker, the second to the proximity of the addressee and the third away from both – can also be seen as based on spatial distance (Diesel, 2005). Languages with systems containing more than three terms are usually person-oriented. Even more advanced distance-oriented systems are possible, such as Malagasy which has contrasts between six different degrees of distance. (Saeed, 2003: 185). Some languages do not code distance into their demonstrative pronouns at all and are called “distance-neutral”. Other languages make contrasts in whether the referent is visible or out of sight, at higher or lower elevation in relation to the deictic center or uphill or downhill.

A typological approach to spatial deixis (Diessel, 2005) suggests some generalizations. Over half of the 234 languages investigated use two-way systems, and over one third use three-way systems. Distance-neutral systems and systems containing more than three terms are fairly uncommon, about 8% of the world’s languages. Two-way and three-way systems are quite evenly spread across the world, while the less common systems are much less so (distance-neutral systems are found only in Africa, Europe and Mesoamerica; systems with more than three terms in North America, Africa and the Pacific region). Thus, despite considerable variation, distance appears to be the major semantic variable for spatial deixis.

Is such deictic distance expressed non-arbitrarily in the world’s languages? Ultan (1978) found that of 136 languages 33% exhibited sound symbolism in the sense that the most proximal term contained a closed, front, unrounded vowel [i], while more distal terms had vowels that were more open and back. Woodworth (1991) found a similar relation in his sample of 26 languages, where 50% of the languages had high frequency vowels in the proximal terms and low frequency vowels in the distal terms. (Two languages showed the opposite association and the remaining languages gave no clear results.) Traumüller (1994) considered 37 languages where there was a difference in vowel quality between the proximal and distal terms. He hypothesized that the proximal terms would have high (F2) frequency, and the distal terms low (F2) frequency. 32 of the languages in the sample followed this prediction.

Further, Traumüller proposed that the association between pitch and distance is based on a correlation stemming from the size of the objects referred to: large at a distance, small when close. Interestingly, heal so looked for non-arbitrariness beyond spatial deixis, based on the observation that 1p personal pronouns often contain a nasal [e.g. *me*], while 2p personal pronouns contain a stop consonant [e.g. French *tu*]. Thus, Traumüller hypothesized that 1p pronouns would tend to contain oral closure and sustained voicing while 2p pronouns would be characterized by oral pressure build-up and explosion. There were 11 supporting cases and 3 counterexamples for this hypothesis.

Thus, there are clearly suggestions for sound symbolism in spatial deixis, but the evidence is far from conclusive. First of all, the language samples used in the typological studies have been relatively small, and not sufficiently representative of the world's languages. The study in which a larger sample was used (Ultan, 1978), was also the one where the lowest proportion of languages supporting sound symbolism was reported (33%). Further, it is difficult to evaluate such studies quantitatively, since the exact predictions are seldom stated in advance. Finally, the motivations behind the (expected) non-arbitrary patterns are only rarely made clear, with some rare exceptions (e.g. Shinohara and Kawahara, to appear). As noted above, Traumüller assumed a key role for reference to distal/large, and proximal/small objects in establishing the mappings, but this is not the case for other researchers, and possibly not necessary.

In the study described in this article, we made an effort to improve on all these aspects, by using a larger, balanced language sample, and formulating testable predictions based on explicit motivations. In the next section, we formulate 6 different predictions for the patterning of vowels and consonants in spatial deictic systems, related to specific motivations. In the following, we test these predictions on a relatively representative language sample, following a clear methodology for language choice and quantification. As we show, some of the predictions are much better supported than others, leading us in the general discussion to some generalizations.

2. Predictions and motivations

As pointed out in the introduction, one of the drawbacks of the concept of sound symbolism has been its vagueness. Unless predictions, based on corresponding explanations, are made in advance of the empirical analysis, as customary in experimental psychology, it would be relatively easy to give post-hoc "motivations" of observed patterns. Hence, in this section we single out six predictions for proximal and distal terms differing with respect to vowels, consonants or both, that may be attributed to factors involved in speech-production, speech-perception or both, based on iconicity, indexicality or both. As we show, each of these predictions implies different scales of sounds corresponding to the dimension proximal-distal.[2]

2.1. *Felt size-to-distance (Haptic-Size-Distance)*

From the speaker's point of view, producing sounds by having the tongue close to the top of the oral cavity clearly feels smaller than when the tongue is low in the cavity and the canal through which the air passes during vocalization is wide (Figure 2a). This haptic sensation of SMALL vs. LARGE can be associated with the auditory perception of the corresponding sounds (mostly vowels), yielding a cross-modal associative mapping between haptic sense and sound (cf. the notion of "phonetic grounding" proposed by Shinohara and Kawahara, to appear). On the other hand, there is an iconic (similarity-based) mapping between SIZE and DISTANCE: small size = small distance, large size = large distance.

The prediction is therefore that sounds produced by having a relatively narrow vocal tract will be associated with *proximal* terms and sounds produced by having a greater distance between the tongue and the upper jaw would feel larger and therefore

[2] This idea was explored as early as Newman (1933).

associated with *distal* terms. Since all consonants are produced by obstructing the vocal tract, this prediction does not apply to them. It is true that approximants, especially semi-vowels, are considerably more open than e.g. nasals which are produced by completely closing off the oral cavity. However, the differences are not great enough to be systematized, and secondary articulations such as pharyngealization would have to be taken into account. Consequently, the prediction concerns only vowels, and gives rise to the scale shown in Figure 2b, predicting a sound-symbolic mapping between vowel quality and spatial distance, based on how open/closed the vowels are, irrespective of other factors such as vowel-roundedness.

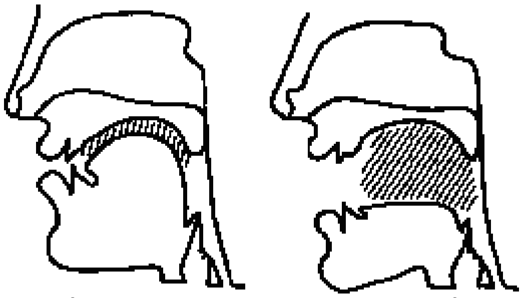

Figure 2a. The amount of open space while producing closed (left) and open (right) vowels

i y ɨ ʉ ɯ u	e ø ɤ o	ə	ɛ œ ʌ ɔ	æ ɐ	a ɶ ɑ ɒ

proximal ◄——————————————————► distal

Figure 2b. The more closed the vowel, the smaller the distance (Prediction 1)

2.2. *Seen mouth-size-to-distance (Visual-Size-Distance)*

From the standpoint of the listener, some sounds co-occur with the speaker opening his mouth extensively, while other sounds co-occur with smaller mouth-openings. This size of *visible* mouth opening could also be associated with spatial distance, according to an iconic mapping similar to that described above. How would the predicted scale look like? Closed vowels, regardless of place of articulation, would be perceived as “smaller” than open vowels. Thus, [i], [ɨ] and [ɯ] should be more proximal than [a] and [ɑ]. The same logic applies to the rounded counterparts of these vowels, leading to the prediction that [y], [ʉ] and [u] should be more *proximal* than [ɶ] and [ɒ].What about the relationship between rounded and unrounded vowels? There are different ways to reason here, but for present purposes we can make the assumption that rounded vowels would in general be perceived visually as smaller than unrounded vowels (see Figure 3a), with the neutral vowel [ə] as intermediate, giving rise to the scale in Figure 3b.

Figure 3a. Degrees of mouth openness: [u] (left), [i] (central) and [a] (right).

y ʉ u	ø o	œ ɔ	ɶ ɒ	ə	i ɨ ɯ	e ɤ	ɛ ʌ	æ ɐ	a ɑ

proximal ←→ distal

Figure 3b. Seen mouth-size-to-distance (Prediction 2)

2.3. *Predictions based on the frequency code*

A dog's growl is low in pitch and perceived as threatening while its whine is high in pitch and perceived as submissive. Ohala (1994) proposed that such correspondences are based on a universal "frequency code", grounded in facts of physiology. Due to the larger resonance chamber of the vocal apparatus of large animals, the frequency created by the vibrating vocal membranes will be dependent on the animal's body size. Thus it would be an *index* of the size of the animal, i.e. based on a correlation in space-time. Low pitch is naturally associated with large body-size, and is thus a so-called honest signal, though in some cases this can also be manipulated, leading to "deceptive behaviors" such as the lowering of the pharynx in red-deer during the mating period (Fitch, 2009).

The frequency code, which Ohala (1994) sees as one of the major factors behind sound symbolism in human languages, goes beyond the kind of indexicality described above, also involving an iconic ground. Low frequencies correlate not only with large animals (and objects in general), but also with greater distances; for example, they attenuate with distance less rapidly than high frequencies, so that low frequency sounds can be heard from much larger distances (Larom et al., 1997). This analogical relation between the frequency and the semantic dimensions of SIZE and DISTANCE (low-freq/high-freq = large/small = far/near) is what makes sound symbolism that is motivated by the frequency code both indexical and iconic.

2.3.1. *Vowel frequency-to-distance mapping (Frequency-Vowel)*

The second formant (F2) in the quality of a vowel is the most varying one, governing a great deal of the characteristics of vowels. Hence high F2 can be expected to correlate with *proximal* and low F2 with *distal*.[3] Traumüller (1994) predicted such a mapping for the five most common vowels in the world's languages, but this could be extended to include somewhat less common, but still not infrequent vowels such as closed, front, rounded vowels (e.g. [y]), closed, back, unrounded vowels (e.g. [ɯ]) and open, rounded vowels (e.g. [ɶ]). These can be placed on a relative high/low scale, as showed by Lindblad (1998), which we utilize as the prediction concerning proximal/distal deixis stemming from the frequency code for vowels (Figure 4).

i	Y	e	ɛ	ø	æ	ɨ	a	œ	ə	ɶ	ɐ	ʉ	ʌ	ɤ	ɑ	ɒ	ɯ	ɔ	o	u

proximal ←→ distal

Figure 4. Frequency code for vowels (Prediction 3)

2.3.2. *Consonant frequency-to-distance mapping (Frequency-Consonant)*

The association connected to F0 frequency can, according to Ohala (1994), also be plotted onto consonants. Voiceless consonants, which are produced by having a high velocity of airflow, have more energy on higher frequencies, while voiced consonants have more energy on lower frequencies (Silverstein, 1994). Hence voiceless consonants

[3] Other traits consistent with the frequency code are also possible: Tonal languages could associate high or rising tones with *proximal* and low or falling tones with *distal* due to the difference created towards the frequency of the modal voice, as in Yoruba, where the verbs for 'being small' *bírí* (high tone) and 'being large' *bìrì* (low tone), contrast only in tone (Ohala 1994).

may be associated with *proximal* and voiced consonants with *distal*. Furthermore, voiced consonants can be sub-divided into palatals, dentals, velars and labials with falling F0 frequency (Lindblad, 1998). Together, these observations give rise to the scale in Figure 5.

Voiceless	Voiced			
	Palatal	Dental	Velar	Labial

proximal ⟵⟶ distal

Figure 5. Frequency code for consonants (Prediction 4)

2.4. *Vocal pointing*

After extensive research during the past few decades, it is now generally recognized that there are close interrelations between speech and gesture. Spontaneous gestures are used in all human cultures, in which they develop similarly (Zlatev and Andrén, 2009; Liszkowski et al., 2012), and at least some gestures such as pointing emerge prior to speech and play a key role for the development of language (Lock and Zukow-Goldring, 2012). In some cultures pointing with the lips is used commonly, at least as much as manual pointing (Wilkins, 2003).

In a less-known study, Williams (1995) proposed that certain prelinguistic utterances containing the sounds [da], [d], [t] and similar variants are used by children with a deictic function. Evidence in support of this is that such vocalizations are primarily used when pointing to or touching an object, and that the origin of this usage is apparently not due to (parental) input. According to Williams there is a close parallel between exploring something with the tip of the finger and bringing the tip of the tongue in contact with the alveolar ridge or the teeth. Williams hypothesized that infants begin performing such "vocal pointing" at around 6 months, before manual pointing emerges around the end of the first year, since while at that age they still lack fine manual motor skills, adequate motor control in the tongue has already emerged through months of sucking and oral exploration. This possible universal connection between certain sounds (above all consonants) and pointing could therefore be a potential source of non-arbitrariness in the expression of spatial deixis. It is less obvious how on the basis of this conjecture to formulate predictive scales in a way similar to that done so far, but we offer the following two possibilities.

2.4.1. *Felt vocal gesture-to-distance mapping (Haptic-Vocal-Pointing)*

One dimension of consonant production that could correlate with "pointability", and thus distance, is whether the place of articulation is in the front part of the vocal apparatus, i.e. from the alveolar ridge forward, or in the back part. A second dimension is how "projecting" a sound is. As pointed out by Traumüller (1994: 223), when plosives and affricates are produced, an "oral pressure is built up and subsequently released in an explosion, [and] a speaker has the impression of suddenly projecting something outward and away from himself". Considering the prominence of this haptic experience given human embodiment, it is reasonable to expect that the projection dimension would have priority over the place of articulation dimension. In other words, a sound would "feel" more forceful, and therefore more distal, than a sound without such a projection, even if the first is produced in the back of the mouth, and the latter in the front. For example, [k] would be more similar to a pointing gesture (and hence more distal) than [n], even though [k] is a "back sound" and [n] is as a "front sound". Table 1 shows the four categories of consonants that are formed by crossing these two dimensions, and Figure 6 shows the predicted scale that follows from ordering the two dimensions as described above. It involves only consonants since vowel production does not project air from the mouth in any comparable manner.

Table 1. Four classes of consonants: (1) Back sounds (red), (2) Front sounds (blue), (3) Back plosives (orange) and (4) Front plosives (green)

	Place of Articulation											
Manner	**Bilabial**	**Labiodent.**	**Dental**	**Alveolar**	**Postalv.**	**Retroflex**	**Palatal**	**Velar**	**Uvular**	**Pharyngeal**	**Epiglottal**	**Glottal**
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ			
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʡ	ʔ
Fricative	ɸ β		θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	ʜ ʢ	h ɦ
Approximant		ʋ		ɹ		ɻ	j	ɰ				
Trill	ʙ			r					ʀ			
Flap/tap	ⱱ̟	ⱱ		ɾ		ɽ			ɢ̆		ʡ̆	
Lateral fricative				ɬ ɮ		ɭ̊˔	ʎ̥˔	ʟ̝̊				
Lateral approximant				l		ɭ	ʎ	ʟ				
Lateral flap				ɺ			ʎ̆					

Back sounds	Front sounds	Back plosives	Front plosives

proximal ◄——————————————————► distal

Figure 6. Haptic-Vocal-Pointing (Prediction 5)

2.4.2. *Seen vocal gesture-to-distance mapping (Visual-Vocal-Pointing)*

From the perspective of the audience, the *visibility* of the signal is more essential for the success of a pointing gesture than its felt strength. As mentioned earlier, lip-protrusion is a common form of pointing in some cultures (e.g. in Southeast Asia, the Americas, Africa, Oceania, and Australia). Thus, speech-sounds produced by pursing the lips may be perceived as more pointing-like, and hence more distal. In this category of "rounded sounds", we may include co-articulated labial consonants (like [m^w]), labialized consonants (like [w]), as well as rounded vowels (like [y]). To the opposite extreme would be positioned non-visually perceptible sounds (like [g]), and in between these, the group of visually perceptible (but not rounded) sounds [θ], giving rise to the scale in Figure 7.

Non-visually perceptible	Visually perceptible	Rounded sounds

proximal ◄——————————————————► distal

Figure 7. Vision-based Vocal-Pointing (Prediction 6)

2.5. *Summary of the predictions*

Table 3 summarizes the six predictions described in this section in relation to the features that differentiate them. As can be seen, the first three predictions concern only vowels, predictions (4) and (5) concern consonants, and the final one both sound types. In terms of semiotic ground, the predictions pattern differently: the first two are based on iconicity (similarity), the last two on degree of "pointability", an essentially indexical/directional characteristic, while predictions (3) and (4), based on the frequency code, were analyzed as relying on both indexical (associative) and iconic (proportional) principles. They are also the predictions that are most intersubjective, involving both the speaker's and hearer's perspective. Very likely, they are also the most multi-modal ones, given the following reasoning. Predictions (2) and (6) imply vision-to-vision correlations. (1) and (5) are cross-modal, implying correlations from felt (haptic) properties of sound production to (mostly) visually given information on distance. Predictions (3) and (4) link first sound (frequency) and object size, mostly visually given, and in a second step such sound/vision features to corresponding features in the environment related to distance.[4]

Table 2. The six predictions, with associated properties, and examples

Prediction	Semiotic ground	Sound type	Source senses	Perspective	Example	
					Proximal	Distal
(1) Haptic-Size-Distance	Iconic	Vowels	Haptic sense	Speaker's	i	a
(2) Visual-Size-Distance		Vowels	Vision	Listener's	u	a
(3) Frequency-Vowel	Indexical + Iconic	Vowels	Hearing + Vision	Speaker's + Listener's	i	u
(4) Frequency-Consonant		Consonants	Hearing + Vision	Speaker's + Listener's	t	m
(5) Haptic-Vocal-Pointing	Indexical/ Directional	Consonants	Haptic sense	Speaker's	h	t
(6) Visual-Vocal-Pointing		Vowels & Consonants	Vision	Listener's	a/h	y/w

3. Method

The method of testing the predictions consisted of three steps. First, a balanced language sample of 101 languages, with reliable data on spatial deictic terms, was created. Second, the terms to be compared and contrasted were carefully chosen, addressing issues of linguistic variation in the expression of the spatial deixis, as those mentioned in the Introduction. Third, a coding procedure for making the tests quantitative and amenable to statistic testing was devised.

Modern studies in linguistic typology underscore the importance of "sampling" from the world's languages in a motivated way in order to make sure that the language

[4] Shinohara & Kawahara (to appear) distinguish between "articulatory factors", corresponding to our "Haptic sense", and "acoustic factors", corresponding to our "Hearing" in a similar attempt to tease apart different possible motivations. However, since we also theorize the crucial role of vision, especially for cross-modal mappings, there is no clear correspondence between our respective typologies.

data that is used for making (or testing) generalizations is reasonably representative (Bybee, Perkins and Pagiuca, 1994; Veselinova, 2005). For this purpose, we used Ethnologue (http://www.ethnologue.com), a free and frequently updated linguistic database, containing information of approximately 6800 living languages. Hence, 68 different languages would correspond to 1% of the world's languages. The primary language families contain from approximately once (Mayan) to 22 times (Niger-Congo) that number (cf. Table 4), and were to be represented by a corresponding number of languages in the sample.

Language families represented by more than 1 language were divided into subgroups, with each sub-group represented appropriately. Language families containing less than 68 languages were grouped into five larger groups, according to the number of languages in the language families:(1) creoles, pidgins and mixed languages; (2) language families containing less than 7 languages, isolates and unclassified languages; (3) language families containing 7-20 languages; (4) language families containing 21-44 languages and (5) language families containing 45-67 languages. Languages representing these five groups were geographically spread out as much as possible. The final selection of languages was governed by available data, using reference grammars and *Compendium of the World's Languages, vol. I* and *vol. II* (Campbell, 1991), leading to the 101 languages presented in Table 4.

Table 3. The language sample, with % for language families/groups of families with less than 68 languages, the actual number of languages per category, and the names of the languages used

Language family	% of the world's 6800 languages	Represented by # in the sample	Languages
Afro-Asiatic	5.1	7	Beja, Hausa, Hdi, Margi, Oromo,Tamashek Tuareg, Tigrinya
Australian	2.2	3	Dyirbal, Nyamal, Wardaman
Austro-Asiatic	2.5	4	Khmer, Mon, Mundari, Nicobarese
Austronesian	17.8	18	Amis, Batak, Buginese, Bunun, Cham, Fijian, Hawaiian, Indonesian, Paiwan, Puyuma, Rapanui, Rukai, Saisyat, Seediq, Sundanese, Tagalog, Tahitian, Tongan
Dravidian	1.2	3	Brahui, Kannada, Telugu
Indo-European	6.2	9	Albanian, Armenian, Belorussian, Dutch, Greek, Irish, Latvian, Romanian, Sindhi
Mayan	1	2	Maya, Tzeltal
Niger-Congo	21.9	15	Akan, Efik, Etsako, Ewe, Fon, Fulani, Igbo, Kpelle, Mbili, Nkore, Swahili, Swazi, Vai, Wolof, Yoruba
Nilo-Saharan	2.9	4	Kanuri, Keliko, Nubian, Shilluk
Oto-Manguean	2.5	2	Chalcatongo Mixtec, Zapotec
Sino-Tibetan	6.4	9	Cantonese, Burmese, Garo, Kachin, Karen, Ladakhi, Limbu, Meithei, Tibetian
Tai-Kadai	1.3	2	Lao, Zhuang
Trans-New Guinea	6.9	4	Awyu, Korafe, Kewa, Tauya
(1) Creoles, Pidgins, Mixed	1.5	2	Haitian Creol, Welsh Romany
(2) <7 languages/family, Isolates, Unclassified	2.2	4	Ainu, Basque, Miskito, Onge
(3) 7-20 languages/family	2.9	2	Inuit West Greenlandic, Maricopa
(4) 21-44 languages/family	6	5	Abkhaz, Chipewyan, Hixkaryana, Komi, Nama
(5) 45-67 languages/family	5.6	6	Alamblak, Apurinã, Even, Guaraní, Nahuatl, Quechua
Total	~100	101	

For each of the languages in the sample, the demonstrative pronouns in their least marked form were extracted, and converted into the International Phonetic Alphabet (IPA) as far as possible. Given the wide spread of both two-way and three-way spatial deictic systems (see Introduction), both kinds were included in the sample, implying that both proximal-medial and proximal-distal contrasts were taken into consideration. Languages with systems using more than three terms were used only if the source clearly distinguished a dimension of (horizontal) distance, and could in that case be used as either a two-way or a three-way system. Distance-neutral languages were not included.

The demonstrative terms for all languages were coded for each of the six possible predictions, using the following categories: *Motivated*, *Neutral*, and *Reverse*. If a particular contrast followed the scales shown in Section 3, it received the value 1 for

Motivated and 0 for Neutral and Reverse. If a contrast was the reverse to the scales, it was coded so, while the other categories received 0. And finally, if it neither followed nor was reverse, it was coded for Neutral. In the case of three-way systems, the comparison was first made between the proximal and the medial form, and then between the proximal and distal form and then averaged (giving the possible values 0, 0.5 and 1 for the three categories).

4. Results and discussion

The six different predictions clearly differ in *granularity*, as the first three predictions list individual phonemes, or small sets of phonemes in each separate "cell" (cf. Figure 2b, 3b, 4), while the second three place large groups of sounds in each cell (cf. Figure 5, 6, 7). Clearly, the chance of both (or in the case of three-way systems, all three) demonstratives falling in the Neutral category is much higher in the latter than in the first case. Hence, the six predictions were primarily evaluated with respect to the Motivated > Reverse expectation. On a scenario of full arbitrariness, these values should be on average equal. Only as a second step, the expectation Motivated > Neutral was evaluated. The values for the six predictions, deriving from the coding procedure, are displayed in Table 5, including the results of statistical significance obtained by the following procedure: We computed odds ratios and confidence intervals for each of the six predictions; the confidence intervals were used to evaluate whether a ratio was significantly different from 1. The chart in Figure 8 shows the odds ratios of Motivated/Reverse and of Motivated/Neutral patterns, indicating that predictions 1, 3 and 6 were significantly supported as far as Motivated > Reverse (with predictions 2 and 5 on the margins). Predictions 2 and 3 were also supported for Motivated > Neutral.

Table 4. Values for the six predictions in "points" (for 101) and percent, and significance results (see Fig 8)

Prediction	Value			Mot > Rev	Mot > Neu
	Motivated	Neutral	Reverse		
1. Haptic-Size-Distance	47.5 (47.2 %)	33.5 (33.2 %)	20 (19.8 %)	*	?
2. Visual-Size-Distance	44.5 (44.1 %)	27 (26.7 %)	29.5 (29.2 %)	?	*
3. Frequency-Vowel	56 (55.6 %)	22.5 (22.3 %)	22.5 (22.3 %)	*	*
4. Frequency-Consonant	27.5 (27.2 %)	48.5 (48 %)	25 (24.8 %)	*n.s.*	*n.s.*
5. Haptic-Vocal-Pointing	35 (34.7 %)	45 (44.6 %)	21 (20.8 %)	?	*n.s.*
6. Visual-Vocal-Pointing	42,5 (42.1 %)	36.5 (36.1 %)	22 (21.8 %)	*	*n.s.*

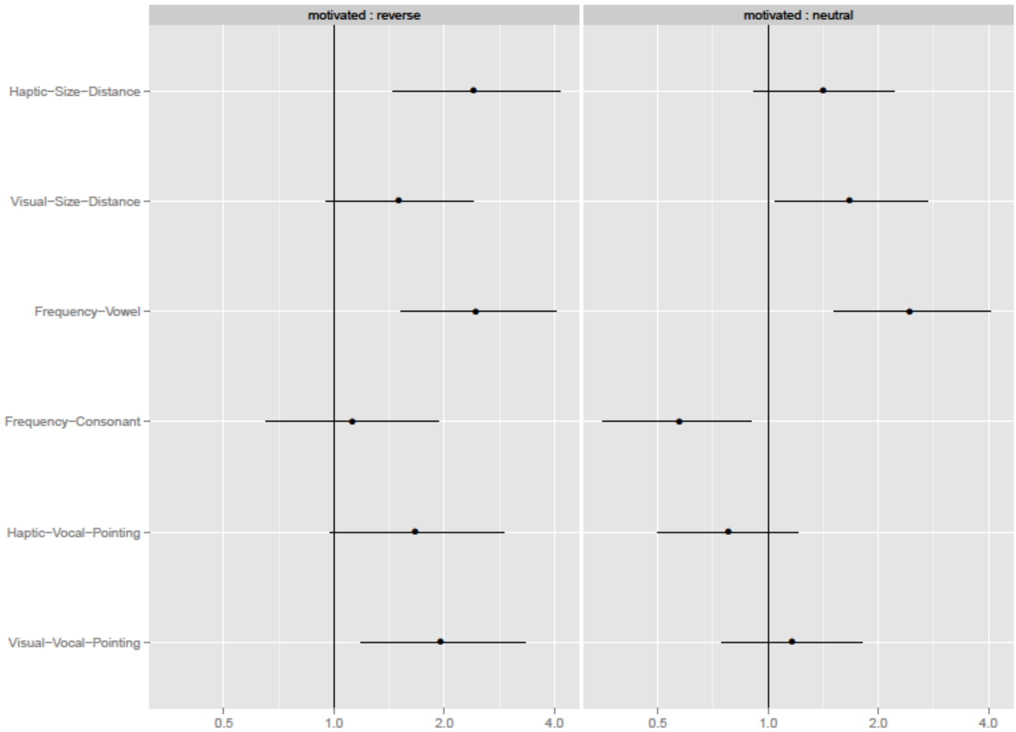

Figure 8. Odds ratios of motivated/reverse patterns (left panel) and of motivated/neutral patterns (right panel). The values of the odds ratios are represented by the dots and their confidence intervals are shown as lines extending from the dots. Odds ratios for which the confidence interval does not include 1 (indicated by the vertical lines) are significant ($p < .05$).

Thus, of the first three predictions, the first and especially the third were clearly supported. The second can be regarded as marginally supported. Even the last prediction, concerning vocal pointing, was supported, despite the high values for Neutral, which were probably due to the nature of the corresponding scales, as pointed out above.

It should be reminded that the first three predictions all concern *vowels*. Since the predicted proximal/distal scales for these bare a considerable degree of similarity (cf. Figures 2b, 3b and 4), it is possible that the most strongly supported of the three (Frequency-Vowel) determines the results for the other two. For example, [i] figures as one of the most proximal sounds for Frequency-Vowel and Haptic-Size-Distance and as relatively proximal for Visual-Size-Distance. Conversely, [a] is considered distal for all three predictions. There are, however, also clear differences: [u] is considered the most distal sound for Frequency-Vowel, but one of the most proximal sounds for the other two predictions. The relatively low results for the Visual-Size-Distance prediction were thus possibly due to an overriding effect from the Frequency-Vowel prediction. Thus, while the three motivations, corresponding to the respective predictions, may play partially independent roles for the presence of sound symbolism in spatial demonstratives, *the most potent of the three is the role of the frequency code for vowels.*

Some of the positive results for the Visual-Vocal-Pointing prediction could also be due to partial overlap with Vowel-Frequency: The only vowels classed as distal on its scale are rounded vowels, which are also lower in frequency. But this cannot explain the marginal support for Haptic-Vocal-Pointing which concerned only consonants. As mentioned in the Introduction, Traumüller (1994) found support for associating second person singular pronouns with "projective" sounds. Combined with the partial support for "vocal pointing" in our study and the arguments of Williams (1995), it seems plausible that plosive consonants and especially alveolar stops would tend to be used in demonstrative expressions.

The only one of the six predictions without any support at all was Frequency-Consonant. Its high value for Neutral can be explained similarly to those for Haptic-Vocal-Pointing: by the large sets of consonants in each cell in their respective scales. But unlike the latter, Frequency-Consonant did not show any difference between Motivated and Reverse values. In contrast with the strong support for Frequency-Vowel, this finding is quite noteworthy. One interpretation is that, while consonants may indeed have an iconic ground for sound symbolism with respect to *other* semantic dimensions (Ohala 1994; Ahlner and Zlatev 2010), they fail to do so in the case of spatial deixis. To the extent that consonants play a sound-symbolic role in demonstratives, they do so indexically, through vocal pointing. Prior to concluding so, however, a few methodological points need to be considered. First, by representing high frequency (= proximal) sounds by voiceless obstruents and low frequency (= distal) sounds by voiced sonorants (in the scale in Figure 3b), devoiced sonorants and voiced obstruents have been excluded. Second, with some more effort, consonants could have been included in some of the other predictions, such as Haptic-Size-Distance: sibilants, plosives and affricates could very well be perceived as *narrow* in relation to other consonants e.g. lateral and approximants. Thus, an iconic ground based on consonantal contrasts should not be ruled out yet.

Another complication is the following. So far, the value Reverse has been treated as the opposite of Motivated, suggesting that it is to be interpreted as *arbitrary*. This is not necessarily so, however. The motivations underlying the studied predictions are of such a character that they would facilitate the learning of the particular semantic contrasts (Imai et al., 2008; Kantartzis, Imai, & Kita, 2011). The "reverse" to this – to make a language deliberately harder to learn – could in some cases be functional too. It has, for example, been observed that several Romani dialects tend to "break" Zipf's law, according to which frequently used words are short. Instead, many closed-class words in such dialects are unexpectedly "long", possibly for purposes of social cohesion, and keeping a distance to other communities. We considered this possibility, and controlled for a correlation between high values for reverse in the languages of isolate/small group (Ainu, Basque, Miskito, Onge), but failed to find one. Still, the possibility should not be excluded that some of the demonstrative systems that should reverse values could also be motivated.

We investigated other possible correlations between predictions and specific language groups. By looking at the cases where one of the three categories (motivated, neutral and reverse) exceeded 50% of the total value of each language group (cf. Table 4) we found that families and groups with positive results for the Frequency-Vowel prediction also gave positive results for the first two supported predictions: Visual-Size-Distance and Haptic-Size-Distance, thus confirming the convergence between the three. The Frequency-Consonant prediction had neutral values for all language groups, except for Tai-Kadai, which interestingly had high reverse values.

As for areal distribution, the results shown in Table 4 were mostly spread out evenly across the world, apart from the following tendencies. Motivated values for the Frequency-Vowel and for the Haptic-Size-Distance predictions were very common in

Southeast and East Asia as well as in East Central Africa. Motivated values for the Haptic-Vocal-Pointing prediction were found to be uncommon in North America, while Reverse values were very uncommon in Africa. Somewhat surprisingly, no correlation was found between areas of reported high use of lip-pointing (Wilkins, 2003) and Motivated values for the Visual-Vocal-Pointing prediction.

In sum, three of the six predictions for different (if related) motivations for the presence of sound symbolism in the world's languages were clearly supported. The strongest of these was the prediction based on vowel-frequency. Characteristically, this was the most "multi-motivated" prediction: iconic and indexical, pertinent both for speaker and audience, and connecting sound, articulation, (bodily) size and distance. What role may such sound symbolism play for language?

Many studies have shown that irrespective of their first language, speakers are able to detect sound symbolism in appropriate contexts, and to perceive an iconic or indexical ground (Sapir, 1929; Köhler, 1929; Sereno, 1994; Ahlner and Zlatev, 2010). Other recent studies show that children, from the age of 2-3 years are able to use this in acquiring certain semantic distinctions. For example, not only Japanese children (Imai et al., 2008), but also English children are more successful in learning expressions denoting manner of movement when these include sound symbolic elements (Kita et al., 2010). A possible explanation is that our evolutionary history makes all children predisposed to use sound symbolic expressions, which some language families may have retained more than others (Ahlner and Zlatev, 2010). The presence and role of sound symbolism in spatial deixis is similar, we conjecture, and predict that an analogous developmental study would show that children find it easier to learn words like *this* referring to someplace near, and *that* to someplace far, rather than vice versa.

5. Conclusions

Our investigation was motivated by questions such as the following: Can we find support for sound symbolism in spatial deixis using a typological approach? What predictions can be formulated (in advance) and which would receive the strongest support, and why? What can the results tell us about the role of sound symbolism in language?

The answer to the first question must be affirmative. As summarized in the Introduction, Woodworth (1991) and Traumüller (1994) found some support for sound symbolism in spatial deixis, but used rather small and not quite representative language samples. Ultan (1978) used a larger sample of 136 languages, but offered much lower support (33%). In contrast, in our balanced sample of 101 languages, the most successful prediction (Frequency-Vowel) gave a result of 56% for motivated expressions.

Concerning the need to formulate clear predictions in advance, this was done only, to some extent, by Traumüller (1994). Of the six predictions which we spent some effort to motivate (and explain), three were strongly supported in the sense that Motivated values/ratios were significantly higher than both Reverse and Neutral. All these concerned vowels, and to some extent overlapped, without being completely redundant. Hence, the three first predictions must be viewed as mutually supportive. While less conclusively, some support was also found for a contributing role of pre-verbal vocal pointing. The negative results for the prediction related to consonant frequency could be interpreted as indicating a lower propensity for consonants to provide the basis for an iconic ground, though this conclusion should be regarded as tentative. Concerning the third and most difficult question, the present findings, combined with previous research, suggests functional reasons for the presence of sound symbolism in spatial deixis, though this too remains to be established experimentally.

As customary, our study has opened up as many new questions as it has answered. Our approach showed how dependent results are on the exact way in which sound symbolism predictions are formulated. Each one of the six predictions that we proposed could be formulated somewhat differently, and new ones could be proposed. For example, person-based deixis may not be best treated as a special form of distance-based deixis. In three-way systems, there are different approaches for grouping proximal, medial and distal (perhaps combining proximal and medial) terms. More complex spatial deictic systems, and even temporal ones, would also need to be analyzed.

Investigations such as these need be conducted, since they concern fundamental questions concerning the scope and limits of sound symbolism and arbitrariness in language. Answering these questions is paramount to understanding the principles by which sounds and meanings are linked, which is an important step in what Jakobson (1965) called "the quest for the essence of language".

Acknowledgments

We wish to thank members of the Centre for Cognitive Semiotics at Lund University and especially Joost van de Weijer for providing aid in the statistical analysis, as well as Arthur Holmer and Gerd Carling, and an anonymous reviewer for much helpful feedback.

References

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

Bybee, J., Perkins, R. and Pagiuca, W. (1994). *The evolution of grammar*. Chicago IL: The University of Chicago Press.

Campbell, G. (1991). *Compendium of the world's Languages, vol. I* and *vol. II*. London: Routledge.

Diessel, H. (2005). Distance contrasts in demonstratives. In M. Haspelmath, M. Dryer, D. Gil and B. Comrie (Eds.), *World atlas of language structures* (pp. 170-173). Oxford: Oxford University Press.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass, 6(10)*, 654–672.

Fitch, T. (2009). *The evolution of language*. Oxford: Oxford University Press.

Herder, J. G. (1772) [1969]. Essay on the origin of speech. In P.H. Salus (Ed.), *On language: Plato to von Humboldt* (pp. 147-166). New York: Holt, Rinehart and Winston.

Hinton, L., Nichols, J. & Ohala J. (1994). *Sound symbolism*. Cambridge: Cambridge University Press.

Ikegami, T. & Zlatev, J. (2007). From non-representational cognition to language. In T. Ziemke, J. Zlatev and R. Frank (Eds.) *Body, language and mind. vol. 1: Embodiment*, (pp. 241–283). Berlin: Mouton de Greuyter.

Imai, M., Kita, S., Nagumo, S. & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*, 54-65.

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

Jespersen, O. (1922). *Language - its nature, development and origin*. London: Allen and Unwin.

Kantartzis, K., M. Imai, and Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science 35(3),* 575-586.

Kita, S., Kantartzis, K. & Imai, M. (2010). Children learn sound symbolic words better: Evolutionary vestige of sound symbolic protolanguage. In A. Smith, B. de Boer and M. Schouwstra (Eds.) *Evolution of language. Proceedings of the 8th international conference (Evolang8)* (pp. 206-213). Singapore: World Scientific.

Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.

Larom, D., Garstang, M., Payne, K., Raspet, R. & Lindeque, M. (1997). The influence of surface atmospheric conditions on the range and area reached by animal vocalizations. *Journal of Experimental Biology, 200 (3),* 421–431.

Lindblad, P. (1998). *Talets akustik och perception*. Göteborgs Universitet.

Liszkowski, U., Brown, P., Callaghan, T., Takida, A., & C. de Voc. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science, 36,* 1-16.

Lock, A., and P. Zukow-Goldring. (2010). Preverbal communication. In *The Wiley-Blackwell Handbook of Infant Development,* eds. J. Bremner and T. Wachs, 394-425). Oxford: Willey-Blackwell.

Maurer, D., Pathman, T. & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science, 9(3)*, 316-322.

Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology, 45(1),* 53-75.

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols and J.J. Ohala (Eds.) *Sound symbolism* (pp. 325-347). Cambridge: Cambridge University Press.

Ramachandran, V. S. & Hubbard E. M. (2001). Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies 8,* 3–34.

Saeed, J. (2003). *Semantics, Second Edition*. Oxford: Blackwell.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology 12*, 225-239.

Saussure, F. (1916/1983). *Course in general linguistics*, Duckworth. London.

Sereno, J. A. (1994). Phonosyntactics. In L. Hinton, J. Nichols and J.J. Ohala (Eds.) *Sound symbolism* (pp. 263-275). Cambridge: Cambridge University Press.

Shinohara, K. & Kawahara, S. (to appear). A cross-linguistic study of sound symbolism: The images of size. *Proceedings of the Thirty-Sixth Annual Meeting of the Berkeley Linguistics Society.*

Short, T. L. (2007). *Peirce's theory of signs*. New York: Cambridge University Press.

Silverstein, M. (1994). Relative motivation in detonational and indexical sound symbolism of Wasco-Wishram Chinookan. In L. Hinton, J. Nichols and J.J. Ohala (Eds.) *Sound symbolism* (pp. 40-60). Cambridge: Cambridge University Press.

Traunmüller, H. (1994). Sound symbolism in deictic words. In H. Auli and P. af Trampe (Eds.) *Tongues and texts unlimited. Studies in honour of Tore Jansson on the occasion of his sixtieth anniversary* (pp. 213-234). Stockholm University.

Ultan, R. (1978). Size-sound symbolism. In J. Greenberg (Ed.) *Universals of human language 2: Phonology*, (pp. 527-568). Stanford: Stanford University Press.

Veselinova, L. (2005). *Suppletion in verb paradigms*. Amsterdam: John Benjamins.

Wilkins, D. (2003). Why pointing with the index finger is not a universal. In S. Kita (Ed.) *Pointing: Where language, culture and cognition meet* (pp. 171-215). Mahwah, NJ: Lawrence Erlbaum.

Williams, S. (1995). *Da and its variants in a bilingual context*. Center for Research on Bilingualism, Stockholm University.

Woodworth, N. L. (1991). Sound symbolism in proximal and distal forms. *Linguistics, 29*, 273-299.

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