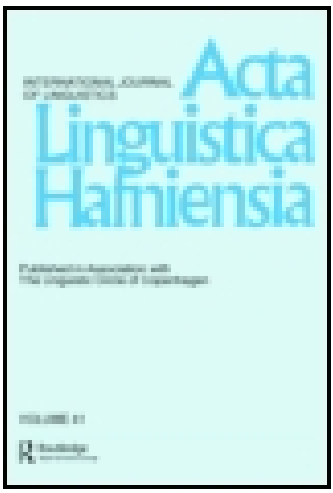

Acta Linguistica Hafniensia: International Journal of Linguistics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/salh20

The de-iconization and rebuilding of iconicity in spatial deixis

Niklas Johansson[a] & Gerd Carling[a]

[a] Centre for Languages and Literature, Lund University, SE-221 00 Lund, Sweden

Published online: 17 Apr 2015.

Acta Linguistica Hafniensia, 2015
http://dx.doi.org/10.1080/03740463.2015.1006830

The de-iconization and rebuilding of iconicity in spatial deixis

An Indo-European case study

Niklas Johansson and Gerd Carling*

Centre for Languages and Literature, Lund University, SE-221 00 Lund, Sweden

This paper investigates iconicity as a possible driving force behind the rebuilding of deictic systems and forms in individual languages. A comparison of a reconstructed Proto-Indo-European deictic system (based mainly on Beekes, *Comparative Indo-European Linguistics: An Introduction*, 1995) compared with the systems of attested Indo-European languages makes it clear that both systems and forms have undergone change, may it be through sound change, analogy, and/or semantic change. Based on the assumptions by Ultan (*Universals of Human Language 2, Phonology*, 1978), Woodworth (1991), Traunmüller (*Tongues and Texts Unlimited. Studies in Honour of Tore Jansson on the Occasion of his Sixtieth Anniversary,* 1994), and Johansson and Zlatev (Motivations for Sound Symbolism in Spatial Deixis: A Typological Study of 101 languages. *The Public Journal of Semiotics,* 2013), iconicity obviously plays a role in the synchronic systems of spatial deixis, which in turn indicates the iconicity has played a role on the process of change, both of the forms themselves and the systems as such. Data from 13 contemporary and 17 historical languages, belonging to 12 Indo-European branches was used. Vowels and consonants were divided into voiceless sounds as being more proximal, and voiced sounds being more distal (see the explanation below). The voiced sounds were divided according to the frequency of their f_2, with [i] and voiced palatal consonants as more proximal and [u] as more distal (Ohala, *Sound Symbolism*, 1994). Results were divided into *motivated* (fulfilling the expected relation between deictic form and sound value), *non-motivated* (arbitrary), and *reversed-motivated* (the reverse of motivated). Five strategies of rebuilding deictic systems and forms were identified. None of the languages investigated have used a system identical to the Proto-Indo-European reconstructed system. Mostly internal material from the Proto-Indo-European deictic system was used in the forms of the systems of the daughter languages. Generally, a statistically significant motivated support was found: 70.2% of the forms of the languages used were identified as motivated, 9.2–10.4% were non-motivated and 19.4–20.7% reversed-motivated. Due to the different strategies of rebuilding systems and forms, generative explanations for the motivated support should be excluded. Hence, iconicity seemed to be reintroduced after the decay, by means of language change, of a former (motivated) deictic system. Therefore, it turned

*Corresponding author. Email: gerd.carling@ling.lu.se

out as a very likely conclusion that iconicity has been and is involved in the rebuilding of deictic material, relating to the systems as such.

Keywords: iconicity; sound symbolism; spatial deixis; language history; Indo-European; motivatedness

1. Introduction

Language is always in a process of change. The types of changes possibly affecting language are plenty in number, ranging from sound change and analogy to semantic change and lexical diffusion.

However, there has been a discussion in linguistic literature concerning motivated relations between sound and meaning for more than 2000 years; from Plato's *Cratylus* (arguing in favor of such a relationship) to the Confucianists (arguing against) via Herder (1772), with the idea of language originating from imitations of sounds of natural phenomena, to the twentieth century Saussurean dogma of arbitrariness, which has dominated linguistic literature since then, with the exception of linguists such as Jespersen (1922), Sapir (1929), and Köhler (1929). However, during the last 30 years the notion of iconicity has resurged, repopularized, among others, by linguists such as Hinton, Nichols, and Ohala (1994) etc.

Many previous studies have had a synchronic typological focus and diachronic investigations have been put to the side, despite the fact that these phenomena often have been documented throughout history. In order to investigate the role of iconicity as one of the factors shaping language we have decided to look at spatial deixis in Indo-European. This for two reasons: (1) spatial deixis is a type of iconicity that is relatively easy to identify and detect: the systems are typically restricted (2–3 deictic forms) and the forms easy to identify (typically monomorphemic, with a consonantal skeleton and a vocal nucleus), (2) the Indo-European language family is probably the best testing-ground for a historical overview of deixis; there is a lot of historical data and the reconstructions are relatively safe.

Section 2 includes general background of spatial deixis and the Proto-Indo-European deictic system. This section also includes an overview of various strategies of decay and rebuilding of forms and systems, presenting also a theoretical background of the identification of iconicity. Section 3 describes the methodology used for the study and section 4 includes an overview of etymologies of the linguistic forms in the study, as well as results and discussion, which are summarized in the conclusion of section 5.

2. Spatial deixis

Not all languages distinguish between spatial demonstratives in the same way as English *this* and *that*. There are different strategies of conveying spatial reference, coding it into a monomorphemic lexeme or using for instance spatial

adverbials to carry some of the information (see e.g., Dixon 2010). However, it is likely that most languages are able to distinguish between the basic spatial deictic concepts of *here* and *there* (Diessel 2005, 223–247).

One of the two most typologically commonly occurring deictic systems (see WALS, Dixon 2010, 234ff.), is similar to the system of Modern English, i.e., a two-way distinction, present in the demonstrative pronouns *this* and *that*. These lexemes are expressing different distances, relative to a deictic center, and can therefore signify completely different positions in different contexts (Saeed 2003). The deictic center is usually the location of the speaker at the time of the utterance (Diessel 2005), which means that *this* refers to something or someone close to the speaker (the *proximal* term), while *that* is further away (the *distal* term).

The second most common deictic system is a three-way system, which can be of two kinds. If the system is *distance-oriented*, the three forms are a reflection of a two-way system but with the addition of another form, i.e., proximal (*this*), medial (*that*), or distal (*that yonder*). In a *person-oriented* three-way system the proximal form refers to something or someone near the speaker, the medial term refers to something or someone near the addressee and the distal term to something or someone away from both the speaker and addressee. As expected, there are yet more advanced deictic systems in other languages: Malagasy differentiates between six degrees of distance, Daga incorporates a horizontal dimension and Yup'ik has an especially complex system (see Table 1).

It is also possible for a language to be fully *distance-neutral*, which indicates that the demonstrative pronouns carry no deictic information, as in Standard Swedish *den*/*denna* and *det*/*detta* (the distinction being gender; *common* and *neuter*). The deictic distinction is then typically expressed by means of adverbials, such as Swedish *här* (proximal) and *där* (distal), accompanying the distance-neutral demonstratives.

Table 1. The system of deictic demonstratives in Yup'ik (Anderson and Keenan 1985 in Saeed 2003).

Extended	Restricted	Obscured	Meaning
man'a	una	–	this (near speaker)
tamana	tauna	–	that (near addressee)
–	–	imna	the aforementioned one
ukna	–	–	the one approaching the speaker
augna	ingna	amna	the one going away from the speaker
agna	ikna	akmena	the one across there
qaugna	kiugna	qamna	the one inland, inside upriver
qagna	keggna	qakemna	the one outside
un'a	kan'a	camna	the one below, towards river
unegna	ugna	cakemna	the one downriver, by the exit
paugna	pingna	pamna	the one up there, away from river
pagna	pikna	pakemna	the one up above

2.1 The Proto-Indo-European deictic system

As in every reconstruction of a proto-language, it is difficult to be fully sure of what the Proto-Indo-European system really looked like. The same goes for the demonstrative stems: the fact that the structure of demonstrative systems and forms in the daughter languages are very different complicates the matter even further. There are various strategies for dealing with this problem: either, one reconstructs a Proto-Indo-European demonstrative system with a large number of forms. Because we consider thus reconstruction improbable, we prefer an alternative reconstruction of a basic set of demonstrative stems, which can be combined with a few reconstructed affixes (cf. Beekes 1995, 202; Meier-Brügger 2000, 213f.). The differences between the daughter languages could then be explained by a combination of the reconstructed demonstrative pronominal stems with derivations from reconstructed particles or adverbs.

Here, the reconstructed set (following Beekes 1995, for alternative reconstructions and views see Szemerényi 1989, 215–218; Meier-Brügger 2000, 212–213) would consist of three deixis neutral demonstrative pronouns, meaning "this, that;" m. **so*, f. **seh$_2$* and n. **tod*, and a set of anaphoric pronouns, meaning "that, the (just named);" m. **h$_1$e*, f. **ih$_2$*, and n. **id*. Beekes (1995) considers this latter set to be used as also for the person pronouns (in third person). Finally, there are three particles/adverbs used for indicating deictic difference; **k̂i* "here," **h$_2$en* "there," and **h$_2$eu* "away, again." This reconstruction indicates that the system was deixis-neutral on the forms of the demonstrative and the third person singular pronouns; hence, deixis was marked by supplementary particles or adverbs.

3. Change of systems and forms

No attested language uses a system which is identical to the reconstructed system. Modern English, for instance, uses two deictic demonstrative pronouns *this*, *that*, with no inflection for gender and no deictic particles. Latin uses a three-way system of deictic demonstrative pronouns *hic*, *ille*, *iste*, all inflected for gender. Modern Armenian uses a three-way system of deictic demonstrative, *ays*, *ayd*, *ayn*, with no inflection for gender. Hence, despite the fact that spatial deictic words (pronouns, particles, adverbs) could be regarded as belonging to core vocabulary of any language, there is no, or at least no common, one-to-one relationship between Proto-Indo-European and its successors, which leads to the conclusion that a significant amount of both decay and rebuilding of systems and forms did take place in the individual languages.

The question that thus remains to be answered is how and to what extent iconicity has shaped the rebuilding of the various Indo-European deictic systems.

3.1 Paradigmatic change of forms: basic types involved

Languages change in a number of ways, the most prototypical being sound change, in which a sound changes, either in all instances (unconditioned change) or in a defined, conditioned number of circumstances (conditioned change).

Table 2. The spread of /i/, /j/ from the first person singular present form to the remainder of the paradigm in Swedish.

	1p sg. pres.	1p sg. pret.	1p pl. pret.	Supine
Old Swedish	biuþa	bøþ	buþu	buþinn
Modern Swedish	bjuda	bjöd	bjödo	bjuden

There are several possible changes that might be involved in the process of emergence of iconicity in spatial deictic systems, the most important being analogy, by Arlotto (1972, 130) described as "a process whereby one form of a language becomes more like another form with which it is somehow associated." Its basic role is to re-structuralize the irregularities created by regular sound change (Antilla 1989, 94) and there are several different types of analogy. Most relevant to our study is a very frequent type of analogy, *analogical leveling*, which usually affects the forms of grammatical paradigms. As an example of analogical leveling, one might take the Old Swedish forms for the verb "invite, offer" in first person singular present tense, first person singular preterit, first person plural preterit, and in the supine, in which an /i/ is present only in the first form. In Modern Swedish, the /i/, /j/ has spread throughout the paradigm (Kroonen et al. 2011, 218), see Table 2.

Another type of interest is lexical diffusion, a process by which a phoneme is altered within a restricted part of the lexicon spreading from one lexeme to another. As an example one might take the shortening of English /u:/, which originated in a specific environment; when /u:/ was preceded by a non-anterior consonant and followed by a non-anterior, non-coronal consonant as in *cook*. The shortening then spread to words whose /u:/ was preceded by a non-anterior consonant only, or followed by a non-anterior, non-coronal consonant only, as in *took* and *good* respectively (see Kiparsky 2004). The question whether lexical diffusion is a sound change or a type of leveling is however debated.

3.2 The process of rebuilding of systems and forms

After the decay of a system through linguistic change (e.g., by means of loss of affixes or endings), the system needs to be rebuilt in some way. The lack of distinctive terms for the singular and plural forms of second person pronouns in English has led to constructions such as *you guys*, *youse* (a pluralized form of *you*), and *y'all* (a contraction of *you all*) to fill the need for a distinction. The same applies to deictic terms in the need for a minimum distinction between the concepts *here* and *there* (Diessel 2005). As mentioned before, Proto-Indo-European and its daughter languages apparently do not utilize the same deictic system, as shown for instance by the systems of Modern English, Latin, and Armenian, where a distance neutral system with three deictic particles has become two- or three-way demonstrative pronoun system, with or without inflection for gender.

Decay and rebuilding of systems occurs also outside of the deictic systems. A prototypical example of decay and rebuilding of a system is the evolution of the Tocharian case paradigm, as studied by Carling (2000, 2008). The starting point for a re-building of a system was here apparently a sparse system containing only three grammatical core cases of Indo-European origin; nominative, oblique, and combined dative/genitive, a remnant of the Proto-Indo-European eight case system. The system was later rebuilt, first by adding secondary cases of local functions; locative, allative, and perlative. To the system the secondary affixes ablative and comitative, as well as instrumental for Tocharian A and causal for Tocharian B were then added, formed by grammaticalized adpositional elements.

Through the process of analogy, thematic vowels became part of affixes in Tocharian A and oblique plural became part of affixes in Tocharian B. Through phonological erosion, final syllables and phonemes at morpheme boundaries were lost. Semantic bleaching made the lexical content of the original adpositions become case affixes and were lost or replaced. After these processes were completed a new desemanticalization process expanded the functional content and syntactic use of the case markers, which gained new, abstract, and more grammatical functions. This led to a wider syntactic use of cases, though holding on to their basic local function and meaning.

In relation to this process of decay and rebuilding, in which the system first broke down and was later rebuilt, using lexical material or free morphemes within a restricted functional sphere, it would be possible to assume that a similar process could have happened to the deictic systems. First, there was an initial loss and breakdown and later the systems were rebuilt, using lexical material or free, grammaticalized morphemes to rebuild a renewed system, in which iconicity had a role similar to analogy: a distribution was secondarily restored, creating a renewed, motivated system in the individual languages.

4. Iconicity: a general overview

The aim of the current paper is to make a contribution to the study on how iconicity shapes language change. In many ways, iconicity is similar to analogy in that it opposes conventional sound change, though it is not supposed to be irregular or sporadic because there are obviously general, synchronic tendencies for motivations of certain sounds to be connected to certain semantic domains

Jespersen (1922) assumes that onomatopoetic words that undergo general sound change often are re-formed to their original shape, which is motivated by a form-meaning correlation. This is exemplified by the word *cut*, whose vowel has changed from [u] to [ʌ], but the same change has not affected the word for the bird *cuckoo* (named by the sounds of its vocalizations). Kaufman (1994) shows similar evidence of resistance toward sound change in present-day Huastec, where onomatopoetic roots contain phonemes that are almost non-occurring in the general vocabulary. A similar case is provided by Traunmüller (1994, 216) concerning the sound change in sound symbolic words: "English *tiny*

'impressively small' is an illustrative example. Due to the general vowel shift, most of the original symbolism of this word has been lost, but it has been restored in the informal variant *teeny-weeny*." Accordingly, sound change caused by *iconicity* or *sound symbolism* could partly be considered a *conditioned* change, only occurring with certain semantic domains, while it is phonologically *unconditioned* because it does not require a specific phonological context to occur.

The idea of frequency code as underlying motivated realizations on the form side has been proposed before (Ohala 1994, Diffloth 1994). The question whether this is a universal or culture/areal-related feature will not be further discussed here (cf. Diffloth 1994). The basic idea is that the frequency code relates to the attempt toward an imitation of large creatures' low frequency voice. Due to the larger resonance chamber of big animals, the frequency created by the vibrating membranes (the vocal cords of animals and the syrinx of birds) is dependent on the body size of the individual and therefore indicates how powerful or threatening that individual is. Through manipulating the voice quality and/or intonation, an individual can imitate a certain size as well as attitudes. Hence, a high and/or rising F0 could be used for indicating smallness but also associated characteristics such as deference, politeness, submission, lack of confidence, questions, familiarity, dependence, and narrowness, while a low and or falling F0 could be associated with largeness but also assertiveness, authority, aggression, confidence, threat, and dominance.

Investigations of iconicity have been performed by various authors and on various semantic domains compatible with the frequency code including form, weight, movement, size etc. Some of the most well-known examples include an experiment conducted by Sapir (1929) in which a large table and a small table were to be named by the two fictive words *mil* and *mal*. Over 80% of the participants considered *mal* to be more suitable for the large table and *mil* more suitable for the small table. Köhler (1929) conducted a similar experiment by matching an asymmetrical roundish shape and a pointy figure with two fictive words; *takete* and *maluma.* In total, 95% of the subjects connected *takete* with the pointy shape and *maluma* with the roundish shape.

4.1 Iconicity in spatial deixis

The existence of iconicity in spatial deixis includes the mapping of LARGE > LARGE DISTANCE and it is a type of iconicity that has been given special attention in recent literature.

Ultan (1978) found that 33.1% out of a sample of 136 languages associated the proximal term of their demonstrative systems with a closed, front, unrounded vowel. Woodworth (1991) found that 13 out of a sample of 26 languages used a vowel with a higher frequency F2 for the proximal form than with the distal form.

Traunmüller (1994) found that out of a sample consisting of 37 languages that 32 languages had proximal demonstrative pronouns containing vowels with a higher frequency second formant (F2) than the distal form. Traunmüller also

included an investigation of person deixis and iconicity due to a correlation between first-person personal pronouns and nasals, as well as second-person personal pronouns and stop consonants. Two hypotheses were formed both connected to the notion of using manner of pointing in order to refer to something at a distance. The first concerned oral closure and sustained voicing for proximal and oral pressure build-up and explosion for distal, which received 11 supporting cases and 3 counter-examples. The other hypothesis concerned sounds with tongue and lip protrusion connected to distal, while proximal was connected to the opposite i.e., closing of the lips, receiving nine supporting cases and seven counter.

Johansson and Zlatev (2013) expanded Traunmüller's investigations, examining six different possible motivations for iconicity in spatial deictic words in 101 areally and genetically diverse languages. The motivations were based on the sense of touch, sight, hearing; frequency of vowels, frequency of consonants and oral pointing gestures; lip protrusion; and air ejection. Each motivation was defined using expected scales of phonemes on the proximal-distal dimension and results were divided into *motivated* (according to a particular prediction), *reverse-motivated* (the reverse of a particular prediction), and if neither was the case *neutral*. The motivation concerning the frequency of vowels received the highest support with 55.6% motivated results compared to 22.3% for both the reverse-motivated and neutral results and 71.3% motivated results when compared to the reverse-motivated results only.

The notion of *reverse-motivated* could be reminiscent of the process described as *anti-Zipfian* effect by Carling, Lindell, and Ambrazaitis (2014, 89f.). Here, the shaping of cultural specialization or marginalization of certain speech communities can be seen in certain parts of language though *hypocoristic formations*, i.e., clipping, camouflaging morphology, or lexical manipulation. However, several explanations for manipulation or conscious change of language are also possible, such as Sprachbund effects producing "incorrect" or reversed relationships to expected or default marking, e.g., between forms of which the original distribution could have been blurred through language change and then rebuilt, but with the original distribution in a reversed position. According to Zipf's law frequently used words, such as determiners and particles ought to contain a small number of phonemes (cf. Sigurd, Eeg-Olofsson, and van de Weijer 2004). In mixed Romani dialects, many closed class words contain "too many" phonemes (Carling, Lindell, and Ambrazaitis 2014), e.g., Welsh Romani demonstrative pronouns *kadava* and *kodova*, compared with English *this* and *that*. This can be regarded as a case of *reverse-motivation*, more specifically *quantitative reverse-iconicity*. The social situation between the Romani people and non-Romani people (often the majority population), using reversed terms in order to keep their language indecipherable would be beneficial. Hence, the forms are not arbitrary (unmotivated); they are motivated to be, so to speak, "unnatural," for purposes of social cohesion, and keeping a distance to other communities.

In a similar manner, the relationship between certain sounds and certain meanings e.g., small size with high frequency sounds and large size with low frequency sounds, shown by Sapir (1929), would be the expected relationship. A *reverse* relationship however would map the meanings with the opposite sounds. For example, [i], containing a high frequency second formant is used in Georgian for the word "large" *didi*. The word meaning "small" *p'at'ara* contains three [a], usually associated with the meaning "large." Hence the [i]-[a] relationship may still be thought to be motivated, though it is reversed compared to the *motivated pattern*. This phenomenon has by most authors simply been judged as a case of arbitrariness. However, because this reversed motivation may not be simply a matter of chance, but at least in some cases actually represent the same relation as the motivated pattern, it is relevant to present its presence by separating it from the clearly unmotivated case, e.g., where there is no relevant contrast at all. In the same way as the distinct antonym pair of *big-small* is possibly affected by a reversed motivation in Georgian, the *here-there* distinction could be affected of this as well.

Because it is obvious that the deictic system of Proto-Indo-European reconstructed by Beekes (1995) has not survived in all of its daughter languages, a significant amount of decay and rebuilding must have taken place. And as shown by Ultan (1978), Woodworth (1991), Traunmüller (1994), and Johansson and Zlatev (2013), iconicity exists, at least in some form, in spatial deictic systems; could this mean that the actual rebuilding and evolution could be influenced by iconicity? Does the contrast between deictic terms reappear? In order to provide an answer two research questions are posed;

- How are deictic systems rebuilt?
- Does iconicity appear and reappear throughout history?

5. Method

5.1 Lexemes used in this study

This study was based on primarily deictically coded demonstrative pronouns, unmarked forms (nominative singular), including all genders. In cases of various pronouns being combined with a deictic adverbial in order to constitute an equivalent meaning, the deictic adverbials were the object of study.

5.2 Languages

In order to be able to detect the possible decay and rebuilding of iconicity, the investigation of several steps of language change was needed. Therefore, historical and present languages were treated in the same way, making a synchronic-diachronic evaluation from the earliest reconstructed form of a language to the most contemporary form, e.g., from Proto-Indo-European through Proto-Germanic and Old English to Modern English.

The chosen languages all belong to the Indo-European language family because it is thoroughly studied and provides an abundance of historical data. Languages chosen were reconstructed Proto-Indo-European as well as languages from 12 branches with at least one language from each branch, adding up to a total of 30 languages, with 13 contemporary and 17 historical/reconstructed predecessors or near relatives, as seen in Table 3. Because the branches have been described to varying detail, it was necessary to compensate for the low number of languages in the less-well documented branches, while selecting fewer representative languages for the branches with more languages. Thus, from the Italic branch only two languages were selected, which are Portuguese and Romanian in order to account for the geographical spread of Vulgar Latin. In this way a few branches would not totally dominate this language sample simply due to the fact that they are better documented. Reconstructed stages, here Proto-Indo-European, Proto-Germanic, Proto-Romani, and Early Romani were not used in the statistical analyses. Still, they were used for making conclusions about motivated change in the deictic systems.

5.3 Sound classification

The deictic classification of sound values was based on their relation to the frequency code. In order to classify the phonetic values of the deictic forms in terms of iconicity, a scale consisting of all cardinal vowels and all pulmonic consonants were used.

The first distinction made was between voiceless and voiced sounds. The voiceless group was considered the most proximal because these type of sounds generally have more energy on higher frequencies and no fundamental frequency. The voiced group consists of both vowels and voiced consonants. The 21 featured vowels were simply put on a row from highest F2, [i], to lowest F2, [u]. The voiced consonants positions in the scale were determined by the frequency in which most accumulation of energy is found, though some researchers would call this the second formant, which is dependent on the place of articulation; labials ranging between 500 and 1500 Hz, velars between 1000 and 1500 Hz, dentals between 1400 and 1800 Hz and palatals exceeding 2000 Hz, explained in Lindblad (1998). This leads to some degree of overlapping compared to the vowels, meaning that several vowels were treated as having the same proximal-distal value as e.g., velar voiced consonants.

To summarize; the most proximal sounds were the voiceless sounds, followed by voiced sounds with a high frequency F2. The most distal sounds were voiced sounds with a low frequency F2, as shown in Table 4.

5.4 Coding

Two segments were distinguished: *binding segment* and *deictic defining segment (s)*. This is, in English *here – there*, /ere/ was defined as binding segment and

Table 3. Indo-European languages in which spatial deictic demonstratives were investigated for the purpose of this study.

Origin	Historical [reconstructed]		Contemporary	Branch
[Proto-Indo-European]	Hittite		–	Anatolian
	Tocharian A		–	Tocharian
	Tocharian B		–	
	–		Albanian	Albanian
	Old Irish		Modern Irish	Celtic
	Latin		Portuguese	Italic
			Romanian	
[Proto-Germanic]	Old English		Modern English	Germanic
	Gothic		–	
	Old Norse		Icelandic	
	Classical Greek		Modern Greek	Greek
	Classical Armenian		Modern Armenian	Armenian
	–		Latvian	Baltic
	Old Church Slavonic		Russian	Slavic
	Old Persian		Modern Persian	Iranian
Sanskrit	–		Hindi	Indic
	[Proto-Romani]	[Early Romani]	Kalderaš Romani	

Notes: Languages in square brackets represent reconstructed stages. These were investigated but not included in the statistics in Tables 6, 7, and 8. The total number of languages investigated was 30, 26 of these formed basic data for the statistics in Tables 6, 7, and 8.

Table 4. Realization of the frequency code: the first distinction made was between voiceless and voiced sounds.

	Voiceless	**Voiced**																						
f_2 frequency	–	2000–Hz			1500–2000 Hz						1000–1500 Hz						500–1000 Hz				500> Hz	
Vowel quality	–	i	y	e	ɛ	ø	æ	ɨ	a	œ	ɘ	œ	ɐ	ʉ	ʌ	ɤ	ɑ	ɒ	ɯ	ɔ	o	u
Consonant quality	Consonants	Palatal consonants																				
							Dental consonants															
											Velar consonants											
											Labial consonants											
		More proximal																		More distal		

Notes: The voiceless group was considered the most proximal because they generally have more energy on higher frequencies and no fundamental frequency. The voiced group consists of both vowels and voiced consonants. The 21 featured vowels were simply put in a row from highest F2 [i], to lowest F2 [u]. The voiced consonants positions in the scale were also determined by their F2, which is dependent on the place of articulation. The most proximal sounds were thus the voiceless sounds, followed by voiced sounds with a high frequency f_2. The most distal sounds were voiced sounds with a low frequency F2 (Johansson and Zlatev 2013, Ahlner and Zlatev 2011).

/h/ – /th/ as deictic defining segments. The binding segments were not considered in the analysis. However, the defining segments were classified as distinguishing (1) proximal-medial deixis, (2) proximal-distal deixis, or (3) medial-distal deixis.

Due to the results of Johansson and Zlatev (2013) vowels were always considered before consonants, regardless of the consonants' phonetic values, e.g., [ij] was considered more proximal than [as] because on the vowel-level [i] belongs to a more proximal group than [a], disregarding that [j] belongs to a group more distal than [s]. When the difference between two terms was made up by one or more phonetic values contrasting with nothing (except the *binding segment*), the *binding segment* was considered being a ground level which the term with the phonetic value or values were determined by, i.e., if the phonetic value was more proximal than the ground level, the term was considered proximal. For example, [a-Ø] was judged more proximal than [a-u] because [a] was acting as the ground level and [u] belongs to a group more distal than [a]. The vowels before consonants rule was also applied here; hence, the stem's vowels were determining the ground level as long as vowels were present.

Based on the coding of Johansson and Zlatev (2013), the results were divided into *motivated* (supporting the motivations, i.e., proximal = higher frequency, distal = lower frequency) while results not fulfilling the *motivated* conditions were divided into *non-motivated* (not supporting the motivations, arbitrary) and

Table 5. Examples of hypothetical proximal and distal deictic forms and their non-motivated and reversed- motivated relationships of both vowels and consonants.

		Motivated		Non-motivated		Reversed-motivated	
		Proximal form	Distal form	Proximal form	Distal form	Proximal form	Distal form
THIS VS. THAT	V.	j**e**	b**a**	t**i** **e**	k**i** **iu**	b**a**	j**e**
	C.	**j**	**b**	**t** **j**	**k** **t**, **m**	**b**	**j**
THAT VS. THAT-THERE	V.	(b**a**) Ø	(ba) **u**	(b**a**) Ø	(ba) **a**	(ba) **u**	(b**a**) Ø
	C.	(**l**) Ø	(l) **m**	(**l**) Ø	(**l**) n	(l) **m**	(**l**) Ø

Notes: The table shows the vowel first-rule, the difference between the more common deictic systems of coding distance in all forms, e.g., *this* vs. *that*, and coding only one form by using a particle or adverbial, e.g., *that* vs. *that* (*there*). "()" marks the *binding segment* (such as a particle or adverbial) used as *ground level*. The *motivated* forms are fulfilling the expected proximal-distal relation by having a higher f_2 in the proximal forms than in the distal forms while the *reverse-motivated* have higher f_2 in the distal forms. The *non-motivated* forms on the other hand have defining segments too similar to be judged distinctive in either direction.

reversed-motivated (proximal = low frequency, distal = high frequency) in order to at least be able to reveal possible reversed but motivated relationships. However, in the statistical analysis the motivated cases are compared with all other results, i.e., *non-motivated* and *reversed-motivated* are both considered as negative results.

Deictic defining segments falling under the type *arbitrary* were typically found under the same frequency value group, e.g., voiceless consonants, as in Portuguese *este–esse*, or having no deictic marking what so ever. The same applies to if the phonetic value of the defining segment of word A was in between the two phonetic values of word B or vice versa. Examples presented in Table 5.

5.5 Quantification

Languages with proximal, medial, and distal forms were judged by contrasting proximal with medial, proximal with distal, and medial with distal in order to treat person-oriented and distal-oriented deictic systems equally.

After coding a language, the values were divided by the total number of values for that specific language. For example, a language with proximal, medial, and distal forms and masculine, feminine, and neuter gender forms respectively has 9 values, of which 4 are motivated, 3 are non-motivated, and 2 are reversed-motivated. Divided by 9, this would yield; motivated 0.45, non-motivated 0.33, and reversed-motivated 0.22, yielding 1 in total. In cases of more distal forms than proximal forms or vice versa, the fewer forms were contrasted twice with the

other forms. For example, in languages with one proximal form and two distal forms the proximal form was first compared with the first distal form, and then again with the second distal form. The values were then divided by 2, yielding a total of 1 for the language in question.

6. Results and discussion

6.1 Etymologies

The three contrasting deictic particles; proximal *$\hat{k}i$, medial *h_2en, and distal *h_2eu, were judged having motivated values for all levels; contrasting proximal-medial *i and *e, proximal-distal *i and *eu, medial-distal *n and *u.

Hittite has a rather common three-way deictic system; the proximal form *kāš* (ka-a-aš) was derived in a straight line from the proximal particle *$\hat{k}i$ (Beekes 1995, 202) cf. Luvian *zā-*, *zi-*, resulting in /k/ being its defining segment. The medial form *apāš* (a-pa-a-aš) $<$ *$apāt$ $<$ Proto-Anatolian **Hobó* seems to be made up of *e/*o + *b^hei (probably the same component as in Lat. $ibī <$ *h_1e + *b^hei and found in the Proto-Indo-European instrumental case, Puhvel 1997, 86; de Vaan 2008, 90; Kloekhorst 2008, 191, 220). The distal form aši (a-ši) which is derived from Proto-Indo-European *$h_1$ós combined with the masculine anaphoric pronoun -i $<$ *h_1e- (Kloekhorst 2008). The *h_1o- stem is found in the other Indo-European languages; hence, Kloekhorst suggests that it was created within Anatolian in analogy to the vowel (-*u*) alterations of *kāš*, *apāš*. Contrasting proximal /k/ and medial /p/, which both belong to the voiceless group, gives a non-motivated value. Motivated results were found for both proximal-distal /k/ and /i/, and for medial-distal /p/ and /i/. Although, because the phonetic value of /p/ is uncertain, possibly being /b/, the results would be motivated results for proximal-medial and proximal-distal, and reverse-motivated for medial-distal.

In all forms in the shared Tocharian two-way deictic system, the proximal and distal forms in all genders, based on the Proto-Indo-European demonstrative pronouns **so*, **seh$_2$*, and **tod* (Pinault 1989), the source of the defining phonemes remains obscure, with the exception of the latter part of Tocharian B distal form, m. *samp*, f. *somp*, n. *tamp*, from the adverb *om(p)* "there, at that place," the apocopated variant of *ompe* of uncertain origin (Adams 1999). In Tocharian A, the proximal forms m. *säs*, f. *sās*, n. *täs* contrast with the distal forms *saṃ*, *sāṃ* and *taṃ*. The masculine and neutral forms' vowels /ɨ/ and /a/ have a motivated contrast. The feminine forms have the same vowel; however, the remaining two sounds, the consonants /s/ and /n/ have a motivated contrast instead. Tocharian B uses two proximal and one distal set of deictic demonstratives; proximal 1 m. *se*, f. *sā*, n. *te*, proximal 2 m. *seṃ*, f. *sāṃ*, n. *teṃ* and distal m. *samp*, f. *somp*, n. *tamp*. The distal forms' /a/ had a lower frequency than the /e/ in the proximal forms. The two feminine proximal forms contained /ā/, though the distal counterparts contained /o/, resulting in all forms being judged motivated.

Albanian has a two-way deictic system as well. The proximal forms m. *ky* and f. *kjo* are both based on the Proto-Indo-European proximal particle *$\hat{k}i$, combined with anaphoric set masculine and feminine forms, *h_1e and *ih_2. The distal forms m. *ai* and f. *ajó* are constructed in the same way, but using the Proto-Indo-European distal particle *h_2eu instead (Kortlandt 1987, 223–224). As for values, the masculine pair *ky*/*ai* was considered non-motivated contrasting [y] with [ai], while the feminine pair *kjo*/*ajó* was considered motivated, contrasting [k] with [a].

Old Irish has a two-way deictic system which is gender neutral. The proximal form *-so* is directly derived from the Proto-Indo-European masculine demonstrative **so*, found in Brythonic as derivatives only e.g., Welsh *hwn* "this" (m.) *hon* (f.) and Middle Breton *ho-n* "your" (Matasović 2009, 37, 299). The distal form *-sin*, reflected as Old Welsh *hinn*, Gaulish *sinde*, *sindas*, is constructed from Proto-Celtic **sindo-*, from *s* + *ih_2m (Proto-Indo-European feminine accusative demonstrative) combined with the particle **do*, a variant of **de* cf. Old Latin *en-do* "in" and perhaps OCS *do* "to, until." The contrasting /o/ with /i/ were judged reverse-motivated.

The Modern Irish system developed into a three-way system. The proximal from *seo* from Old Irish *-so*, the medial from sin from Old Irish distal *-sin*, and the distal form *siúd* from the shortened form of *ucut*, i.e. *úd*, as well as an analogical *-s*. *ucut* is the second person singular of the preposition *oc* "by," meaning "by you" from Proto-Celtic **onko-tu*, also attested by Middle Welsh *wnc*, *wng* "near." The preposition **onko-* is derived from the verbal root **anko-* "reach," also found in OIr. *-ic* (*ro-ic* "reaches," *do-ic* "comes"), MW *rynghu* "reach," Lat. *nanciscor*, and so forth, ultimately from Proto-Indo-European *$h_2ne\widehat{k}$- "reach, attain," in which **-nk-* regularly developed to /g/ in Goidelic (Matasović 2009, 37, 299). The proximal-medial [o:] and [i] was judged reversed-motivated, while proximal-distal [o:]– [u:] and medial-distal [i]– [u:] were judged motivated.

Latin, like Hittite, has a three-way deictic system. The proximal forms m. *hic*, f. *haec*, n. *hoc* seem to have been constructed by combining the Proto-Indo-European stems *g^ho- and *g^heh_2 (Baldi 1999, 342–344) with the emphatic particle *c* (*e*) < *$\hat{k}i$ (de Vaan 2008, 102, 284). The medial forms m. *iste*, f. *ista*, n. *istud* were formed by combining the Proto-Indo-European masculine anaphoric form *h_1e with the strengthening enclitic connective *-te*, as well as inflection for gender. The distal forms m. *ille*, f. *illa*, n. *illud* were probably formed from the stem **ol-* (connected to *ul-terior* "farther," *ul-trā* "beyond"), combined with *-*no-*, from the Proto-Indo-European demonstratives **so*, *seh_2, and **tod*. The values for Latin were quite diverse. The masculine deictic set was considered motivated at all levels; proximal-medial and proximal-distal /h, k/ and /e/, and medial-distal /st/ and /ll/. The feminine proximal-medial and proximal-distal sets, contrasting /e/ with /i/ yielded reverse-motivated values, while proximal-distal /st/ and /ll/ were judged motivated. The neuter proximal-medial and proximal-distal sets, contrasting /o/ with /i, u/ yielded non-motivated results, the proximal-distal set, as the feminine version, contrasting /st/ with /ll/ were judged motivated.

Portuguese inherited the three-way system from its mother language. The proximal set m. *este*, f. *esta*, n. *isto* was derived from the Latin medial set *ist-*, seen clearly in the neuter form. The medial set m. *esse*, f. *essa*, n. *isso*, was derived from Latin *ipse*, *ipsa*, *ipsum*, which was derived from the Proto-Indo-European masculine anaphoric form $*h_1e$ combined with the first demonstratives $*so$. The distal set m. *aquele*, f. *aquela*, n. *aquilo* were constructed by combining *accu* (ultimately from $*h_1e + *\hat{k}i$) with the Latin distal set *ill-* (Azevedo 2005, 159). Due to the more consistent Portuguese forms compared to the Latin forms, the defining segments for each distance distinction had the same values of all genders; proximal-medial contrasting [st] with [ss] yielding a non-motivated distribution; proximal-distal contrasting [st] with [a] and medial-distal, contrasting [ss] with [a], which were all found to be motivated.

The Latin three-way system has collapsed into a two-way system in Romanian, inflected for masculine and feminine. The proximal forms m. *acesta* and f. *aceasta* are, similar to the Portuguese distal forms, formed from *ecce* ($<ecce < *eke- < *eko- < *h_1e + *\hat{k}i$) (Lindsay 2010, 617) and the Latin medial form *ist-*. The distal forms are formed in the same manner, from *ecce* ($<*\hat{k}i$) and the Latin distal set *ill-*. The non-motivated values are gone, the masculine set was judged motivated, contrasting [st] with [l] and the feminine set judged reversed-motivated, contrasting [a] with [e].

Proto-Germanic, just like all Germanic languages in this investigation, has a reduced deictic system, using proximal and distal forms. The proximal forms m. (*h*) *iz*, f. *sī*, n. (*h*) *it* derive from the Proto-Indo-European anaphoric set $*h_1e$, $*ih_2$, $*id$ and the distal forms m. *sa*, f. *sō*, n. *þat* from the other Proto-Indo-European demonstrative set $*so$, $*seh_2$, $*tod$ (Ringe 2006, 288–289). The $*i$-phonemes of the proximal forms contrast with all of the distal forms in a motivated way, which contain $*a$ and $*ō$.

The Old English proximal forms m. *þes*, f. *þēos*, n. *þis* are constructed through merging the Proto-Germanic masculine and neuter distal forms $*sa$ and $*þat$ (Prokosch 1939, 271–272). The distal forms m. *sē*, f. *sēo*, n. *þæt* are directly derived from the Proto-Germanic distal set (Baldi 1999, 342). The masculine and feminine deictic sets were judged non-motivated because they were contrasting proximal /θ/ with distal /s/, both voiceless consonants. The neuter set was clearly motivated, contrasting proximal /i/ with distal /æ/.

The Modern English gender neutral system, proximal *this*, distal *that*, is derived from the only motivated set, Old English neuter forms *þis*/*þæt*. Like in the set in Old English, the Modern English deictic pair was judged motivated, contrasting [I] with [æ].

The Gothic system is reduced even further to a distance-neutral system, m. *sa*, f. *sō*, n. *þata* based on the Proto-Germanic distal forms. In order to distinguish between proximal and distal, an adverb/adjective is added to the distal form, contrasting with the proximal form which is expressed by plain m. *sa*, f. *sō*, n. *þata* (Baldi 1999, 342). The addition m. *jains*, f. *jaina*, n. *jainata* are based on the Proto-Indo-European medial particle $*h_2en$ (Prokosch 1939, 272). All deictic

pairs were judged reversed-motivated; in the masculine and neuter pairs the /i/ in the adverb/adjective was more proximal then the ground level /a/ in *sa* and *þata*. The feminine pair was even clearer because all vowels in *jaina* were considered more proximal than the /o/ in sō.

The Old Norse proximal form m., f. *sjá* was constructed from PGm. **sa* combined with *-si* (ultimately from the same root), with an insertion of *-j-*. This form was later replaced by *þessi*, from PGm. *þat* and *sa*. The neuter form *þetta* was directly derived from PGm. **þat*. The distal forms m. *sá*, f. *sú*, n. *þat* were derived from PGm. **sa*, **sō*, **þat* (Prokosch 1939, 271–272). All deictic pairs were considered motivated. The /j/ of *sjá* contrasts with the ground level /a/ in *sá* and *sjá*, while the /a/ of *sjá* contrasts with /u/ in *sú*. The /e/ of *þessa* contrasts with the ground level /a/ in *sá* and *þessa*, and it contrasts with /u/ in *sú*. Lastly, /e/ contrasts with /a/ in *þessa* and *þat*.

Like Gothic, Modern Icelandic has reduced its system to a distance-neutral system, m., f. *þessi* and n. *þetta* was directly derived from Old Norse *þessi* and *þetta* (Jóhannesson 2006, 34). Two adverbs are used for distinguishing between proximal and distal; *hérna* and *þarna*, which were judged motivated, contrasting [e] with [a].

Classical Greek as of the fifth century BCE has become a two-way system, and the proximal forms m. *houtos*, f. *hautē*, n. *touto* were constructed by combining the Proto-Indo-European distal particle *h_2eu, in zero grade *h_2u, with the demonstratives **so*, **seh$_2$* and **tod* (Beekes 1995, 202). The distal forms m. *ekeinos*, f. *ekeinē*, n. *ekeino* are probably created by combining the Proto-Indo-European proximal particle *$\hat{k}i$ with the medial particle *h_2en transformed into **eno-*, cf. Gk. *enē* "the third day," preceded by the anaphoric masculine pronoun *h_1e; *(h_1e-) $\hat{k}e$-*y*-h_1*eno*- (Beekes 2009; Waanders 1997: 273). The values were judged reversed-motivated contrasting /o̯u̯/ with /e, e̯i̯/.

The Modern Greek system is almost identical to the one of Classical Greek, with the exception of leveling in the proximal form by generalizing *tu-* initially; m. *tutos* < *houtos*, f. *tuti* < *hautē*, n. *tuto* < *touto*. The distal forms are basically the same m. *ekinos*, f. *ekini*, n. *ekino*. Both pairs have just like Classical Greek been judged reversed-motivated, though the generalizing has changed the contrasting, defining segments to [u] and [e, i].

Classical Armenian became a gender neutral three-way deictic system with one anaphoric set and one "real" demonstrative set. The proximal form of the anaphoric set is *so-* derived from the Proto-Indo-European proximal particle *$\hat{k}i$, the medial form *do-* is derived from the Proto-Indo-European neuter demonstrative **tod*, and the distal form *no-* is derived from the Proto-Indo-European medial particle *h_2en. The "real" demonstratives *ays*, *ayd*, and *ayn* are formed in the same way, though preceded by the Proto-Indo-European masculine anaphoric form *h_1e (Schmitt 1981, 120). On the proximal-medial and proximal-distal level the contrasting voiceless /s/ with voiced /d/ and /n/ were judged motivated, though the medial-distal level was judged non-motivated because /d/ was contrasted with /n/.

Like in the Greek branch, Modern Armenian is basically a mirror of its classical form, though with one difference; *sa* < *so-*, *da* < *do-*, and *na* < *no-* have shifted from anaphoric usage to becoming the "real" demonstrative, while *ays* < *ays*, *ayd* < *ayd*, and *ayn* < *ayn* are used as attributes (Tragut 2009, 129–135). The values were found to be identical to Classical Armenian; proximal-medial and proximal-distal motivated and medial-distal non-motivated.

Latvian has a two-way deictic system, inflected for masculine and feminine. The proximal forms m. *šis* and f. *ši* derive directly from the Proto-Indo-European proximal particle **k̂i* (Beekes 1995, 202). The distal forms m. *tas* and f. *tā* derive from the neuter Proto-Indo-European demonstratives **tod* (Kortlandt 1983, 312, 316). Both the masculine and feminine sets were judged motivated, the proximal forms [i(:)] contrasting with the distal forms [a(:)].

The Old Church Slavonic system was reduced to a two-way system, still inflected for all three genders; m. *s*, f. *si,* n. *сe*, all derived from the Proto-Indo-European proximal particle **k̂i* (Kortlandt 1983b, 313). The distal forms m. *onŬ*, f. *ona*, n. *ono*, are all derived from the Proto-Indo-European medial particle **h₂en* (Beekes 1995, 202). There was no binding segment for any of the gender pairs, though all were judged with motivated values, contrasting /ĭ/ with /o, Ŭ/, /i/ with /o, a/ and /ɛ/ with /o, ɔ/.

The Russian system shares no similarities with the Old Church Slavonic system. Both the proximal forms m. *èmom*, f. *èma*, n. *èmo* and the distal forms m. *mom*, f. *ma*, n. *mo* are related to the Old Church Slavonic distance-neutral demonstratives *mŬ*, *та* and *то*, derived from the Proto-Indo-European demonstratives **so*, **seh₂*, and **tod* (Beekes 1995, 204; Vasmer 1958, 465; Schmalstieg 1995, 62–64). The proximal initial ∋′- is derived from the masculine form of the Proto-Indo-European anaphoric form **h₁e*. Although having a completely different system compared to its older cognate, all three sets were judged motivated as well, contrasting [ɛ] with Ø, though contrasting through the ground levels made up by [o] and [a] in common for both deictic terms ([ɛ] being more proximal than [o] and [a]).

Old Persian has a two-way system, the masculine and feminine proximal form *iyam* is constructed by combining the Proto-Indo-European masculine anaphoric form **h₁e* and the particle *-am*. The neuter form *ima* derived from the Proto-Indo-European singular masculine accusative demonstrative **im*, also combined with *-am*. The distal forms m. *ava*, f. *avām*, n. *avaš-ciy* are all derived from the Proto-Indo-European distal particle **h₂eu*. There is also another proximal term *aita-*, constructed from Proto-Indo-European anaphoric forms **h₁e*, **id*, and **tod* (Kent 1953, 68–69). All three pairs for the first proximal forms and the distal forms were judged motivated, contrasting /i/ with /a/. In contrasting *aita-* with the distals, all pairs were judged motivated, through contrasting the masculine and feminine proximal /i/ with the ground level /v/ and /m/, and contrasting /t/ with /v, j/ in the neuter pair.

The Modern Persian system is quite simplified, contrasting proximal *in* (etymology not found) with distal *ān* < Avestan *anya*, cf. OPers., Av. *ava-* < *h_2eu* in zero grade *h_2u*. The pair was judged motivated, contrasting [i] with [a:].

The Sanskrit two-way system consists of proximal m. *ayám*, f. *iyám*, n. *idám*, constructed through combining the Proto-Indo-European anaphoric set *h_1e*, *ih_2*, *id* with the particle *-am* (cf. Old Persian, Thumb 1959, 135–148; Kent 1953, 68–69). The distal forms m., f. *asáu*, n. *adáḥ* are constructed by combining the Proto-Indo-European distal particle *h_2eu* with the other Proto-Indo-European demonstrative set **so* (for masculine and feminine) and **tod* (Beekes 1995, 202), Sanskrit also had a second proximal set m. *eṣás*, f. *eṣá*, n. *etát*, created from *e first combined *h_1e*, *ih_2*, *id* and then **so*, *seh_2*, **tod* (Kent 1953, 69). According to Kellogg (1938, 214), the vernacular language could have had a corresponding distal set to this set; **oṣás*, **oṣā́*, **otát*. All deictic pairs were judged motivated; *ayám* and *asáu* contrasting /j, m/ with /u/, *iyám* and *asáu* contrasting /i/ with /a, u/, *idám* and *adáḥ* contrasting /i/ with /a/, *eṣás* and *asáu* contrasting /e/ with /a, u/, *eṣá* and *asáu* and contrasting /e/ with /a, u/ and *etát* and *adáḥ* and contrasting /e/ with /a/. The motivated results would also apply to the hypothetical distal set.

Hindi uses a simple gender neutral two-way system, proximal *yah* and distal *vah*. The proximal form was derived from Sanskrit *eṣás* (Kellogg 1938, 214) and the distal form from the proposed corresponding Sanskrit distal form **oṣás*. Just like its predecessor the values were motivated, contrasting [e] with [a].

Proto-Romani has a similar system to Hindi, though with masculine and feminine gender; proximal m. **ata*, f. **ati* and distal m. **ota*, f. **oti*, though with no known origin. The **a* in the proximal forms contrasting with distal defining segment **o* yields motivated values for both genders.

The Early Romani system distinguishes between proximal plain m. *adava* and f. *adaja*, proximal specific (precisely here) m. *akava* and f. *akaja*, distal plain m. *odova* and f. *odoja* and distal specific (precisely there) m. *okova* and f. *okoja*. The forms were created by combining the deictic particles *adaj*, *akaj*, *odoj*, and *okoj* (with unknown origin) with *ava* < **ata*, *aja* < **ati*, *ova* < **ota*, and **oja* < **oti* (Matras 2002, 103–112). Because the defining segments are, just like Proto-Romani, proximal **a* and distal **o*, the forms were judged motivated.

Kalderaš Romani used a new system, with three deictic points; proximal m. *kako*, f. *kakya*, medial m. *kodo*, f. *kodya*, and distal m. *kuko*, f. *kukya*. The proximal forms are derived from the Early Romani proximal specific forms, the medial forms from the Early Romani distal plain forms and the distal forms from the Early Romani distal specific forms (Lee 2010, 54). The deictic triplets were both judged motivated because the vowel qualities were increasingly lower in frequency according to distance; [a], [o], [u]. All lingusitic forms summarized in Table 6.

Table 6. Linguistic forms for all languages featured in the study.

Language	Type	Proximal	Medial	Distal
PIE	–	*ḱi	*$h_2$en	*$h_2$eu
Hittite	–	kāš	apāš/abāš	aši
Tocharian A	M	säs	–	saṃ
	F	sās	–	sāṃ
	N	täṣ	–	taṃ
Tocharian B	M	se/seṃ	–	samp
	F	sā/sāṃ	–	somp
	N	te/teṃ	–	tamp
Albanian	M	ky	–	ai
	F	kjo	–	ajó
Old Irish	–	–so	–	–sin
Modern Irish	–	seo	sin	siúd
Latin	M	hic	iste	ille
	F	haec	ista	illa
	N	hoc	istud	illud
Portuguese	M	este	esse	aquele
	F	esta	essa	aquela
	N	isto	isso	aquilo
Romanian	M/N	acesta	–	acela
	F	aceasta	–	aceea
Proto-Germanic	M	(h)iz	–	sa
	F	sī	–	sō
	N	(h)it	–	þat
Old English	M	þes	–	sē
	F	þēos	–	sēo
	N	þis	–	þæt
Modern English	–	this	–	that
Gothic	M	sa	–	sa jains
	F	sō	–	sō jaina
	N	þata	–	þata jainata
Old Norse	M/F	sjá/þessa	–	sá
	N	þetta	–	þat
Icelandic	Adv.	hérna	–	þarna
Classical Greek	M	houtos	–	ekeinos
	F	hautē	–	ekeinē
	N	touto	–	ekeino
Modern Greek	M	tutos	–	ekinos
	F	tuti	–	ekini
	N	tuto	–	ekino
Classical Armenian	–	so	do	no
Modern Armenian	–	ays	ayd	ayn
Latvian	M	šis	–	tas
	F	šī	–	tā
Old Church Slavonic	M	sĭ	–	onŭ
	F	si	–	ona
	N	se	–	ono
Russian	M	ètot	–	tot

(*Continued*)

Table 6 – *continued*

Language	Type	Proximal	Medial	Distal
	F	èta	–	ta
	N	èto	–	to
Old Persian	M	iyam/ aita-	–	ava
	F	iyam/aita-	–	avām
	N	ima/aita-	–	avaš-ciy
Modern Persian	–	in	–	ān
Sanskrit	M	ayám/eṣás	–	asáu
	F	iyám/eṣā	–	asáu
	N	idám/etát	–	adá?
Hindi	–	yah	–	vah
Proto-Romani	M	*ata	–	*ota
	F	*ati	–	*oti
Early Romani	M1	adava	–	odova
	F1	adaja	–	odoja
	M2	akava	–	okova
	F2	akaja	–	okoja
Kalderaš Romani	M	kako	kodo	kuko
	F	kakya	kodya	kukya

6.2 Summary of results

An overwhelming majority (70.2%) of the values of the investigated languages were found to be motivated, $t\ (25) = 2.80$, $p = 0.010$, far above the statistical range of the margin of error (see Table 7).

Here one could argue that the results could have been due to the genetic relationship, i.e., the individual languages could have inherited forms from Proto-Indo-European which possibly could have been motivated, and the iconicity would have been preserved in a "frozen" form into the daughter languages. However, if the various strategies of rebuilding the deictic systems in individual languages, described below, are taken into consideration, this argument no longer has validity. None of the languages investigated have derived their terms in a straight line from Proto-Indo-European. However, one would not expect all languages to be motivated at any given time because the process of decay of systems makes some of the languages to be in a state with no or at most vaguely motivated associations, before being rebuilt once again.

Table 7. Distribution of forms as being motivated (+), non-motivated (0), or reversed-motivated (−).

Motivated +	Non-motivated 0	Reversed-motivated –	Total
18.24	2.38–2.71 7.76	5.05–5.38	26 languages
70.2%	9.2–10.4% 29.8%	19.4–20.7%	100%

Table 8. Distribution of strategies for rebuilding deictic system in the studied languages.

Origins of deictic components			
PIE deictic system continued	Anaphoric/Personal Pronouns	6.4%	55.2%
	Demonstrative Pronouns	20.9%	
	Deictic Particles	27.9%	
System-external Proto-Indo-European material used		5.1%	
Unknown origin of forms or affixes		16.9%	
Mergers (mainly system-internal)	Two Elements	17.6%	20.9%
	Three Elements	3.3%	

The reversed-motivated values were slightly more represented than the non-motivated values: 19.4–20.7% against 9.2–10.4% (depending on whether the Hittite medial form contains /p/ or /b/), possibly due to a part of reversed-motivated values being motivated, having the motivated distinction realized in a reverse structure, though still maintaining the actual contrast between the terms. Furthermore, if the four reconstructed languages would have been included into the statistics, even stronger motivated results would be found; 74.1% motivated, 7.9–9% non-motivated, and 16.8–17.9% reversed-motivated.

When looking at the origins of the investigated words, excluding the inflection origins, only counting the deictic part of the words, five major strategies for rebuilding the deictic systems were found, summarized in Table 8. The origins were relatively evenly distributed, 55.2% were derived from the Proto-Indo-European deictic system; however, because it consists of anaphoric/personal and demonstrative pronouns which both are not coded deictically as well as the deictic particles, no origin actually exceeds 30% of the total ways of rebuilding and only 27.9% of the deictic forms are directly derived from the Proto-Indo-European deictic particles. The various strategies for rebuilding the system can be summarized as follows:

- *Proto-Indo-European deictic system continued:* this indicates that the deictic segments of the forms were derived directly from the two Proto-Indo-European reconstructed demonstrative sets of deictic particles. This was the case for the majority of the forms in the investigated languages.
- *System-external Proto-Indo-European material used:* this indicates that the deictic segments of the forms were derived from outside the reconstructed Proto-Indo-European deictic system, e.g., Old Irish *siúd* (distal) from Proto-Indo-European *$*h_2ne\hat{k}$-.
- *System-internal merger of form:* this indicates that two or more (in this investigation two and three, and almost exclusively from the Proto-Indo-European deictic system) forms or roots were used, combined to create new deictic segments. The possible deictic value of the original forms or roots used did not belong to the same deictic values as the recreated terms in

Table 9. Distribution of forms as being motivated (+), non-motivated (0), or reversed-motivated (−) in relation to deictic relations within the system.

Proximal-medial			Proximal-distal			Medial-distal		
+	0	−	+	0	−	+	0	−
6–7	4–5	2	60	4	10	10–11	2	0–1
	6–7			14			2–3	
46.2–53.9%	30.8–38.5%	15.4%	81.1%	5.4%	13.5%	76.9–84.6%	15.4%	0–7.7%
	46.2–53.9%			18.9%			15.4–23.1%	

every instance. Examples would be Sanskrit *eṣás*, *eṣā́*, *etát* (proximal) from **e* + **h₁e* + **so*, **e* + **ih₂* + **seh₂*, **e* + **id* + **tod* (cf. Kent 1953; Mayrhofer 1986–2001, 272).

- *Ø used as a system contrast:* Gothic uses a distance-neutral demonstrative set combined with zero to denote proximal (*sa*, *sō*, *þata*), and the same set combined with an adverb/adjective to denote distal (*sa jains*, *sō jaina*, *þata jainata*).
- *Unknown origin of forms or affixes:* affixes or whole forms with disputable or uncertain origin occurred, quite well represented, e.g., Tocharian B *seṃ*, *sāṃ*, *teṃ* from **so*, **seh₂*, **tod* + *-ṃ*.

When looking at the results oppositional word pair by oppositional word pair for each of the three different oppositional levels, proximal-medial, proximal-distal, and medial-distal, the support for the values of each level differ. On the proximal-medial level, no clear patterns were found, motivated 46.2–53.9%, non-motivated 30.8–38.5%, and reversed-motivated 15.4%. However, for the proximal-distal and medial-distal levels the motivated values were found rather clear 81.1% and 76.9 −84.6%, respectively, though due to the low number of cases for the latter only the proximal-distal level was judged significant (t (25) = 2.906, p = 0.010), as seen in Table 9.

The fact that distinction between the proximal and medial forms was unclear can possibly indicate a closer relationship between these concepts because both are clearly distinguished from the distal concept. According to Diessel (1999b: chap. 3 in Diessel 2005), person-oriented three-way systems tend to function in a similar way. For this kind of system there are two different origins; the domain of the speaker which contrasts with the domain of the hearer and the common domain of the speaker and hearer, contrasting with the distal concept.

Of the total 9 Latin values all values but one were non-motivated/reversed-motivated on the proximal-medial level, as well as for 2 out of 3 values on the proximal-distal level, the rest being motivated[1]. However, in Portuguese, the

1. The occasionally used alternative forms of *ille*, *olle*, and *ollus* (de Vaan 2008, 426) were not taken into consideration in this case, however, even if the two forms would be included the results would remain motivated.

Table 10. Latin and Portuguese results (motivated (+), non-motivated (0), or reversed-motivated (−)).

		Proximal-medial	Proximal-distal	Medial-distal
Latin	Masculine	+	+	+
	Feminine	−	−	+
	Neuter	0	0	+
Portuguese	Masculine	0	+	+
	Feminine	0	+	+
	Neuter	0	+	+

motivated values were found on the entire proximal-distal and medial-distal levels, leaving just the proximal-medial level being non-motivated, seen in Table 10. This points toward a prioritizing of needed contrasts between proximal-distal, before proximal-medial.

6.3 Primary supporting arguments

Here, a few examples from the data sets will be brought forward to demonstrate how iconicity could have functioned as a driving force in restructuring deictic systems.

The defining segments of the proximal forms of Classical Greek: /ho*u*t/, /ha*u̯*t/, /to*u*t/ (m. *houtos*, f. *hautē*, n. *touto*, cf. Allen 1987), have been leveled̂ out to Modern̂ Greek /tut/ in which the vowel's F2 of the proximal forms has become lower in frequency. In the same way the distal forms changed from Classical Greek /eke:n/ (m. *ekeinos*, f. *ekeinē*, n. *ekeino*) to Modern Greek /ekin/, in which the vowel's F2 was increased. This means that the opposing deictic terms' divergence from each other has made the contrast stronger, even though they diverged in the "wrong" direction. What was even more interesting was that the Greek proximal forms contain the material of the Proto-Indo-European distal particle, while the Greek distal forms are based on a combination of the Proto-Indo-European proximal and medial particles, which goes hand in hand with the possible reverse-motivated results. This contrast is normally explained through general sound changes (cf. Allen 1987), though it was convenient that the sound of the forms have radicalized in this manner. The overall picture is that despite rebuilding of the Proto-Indo-European deictic system resulting in the Greek version, iconicity might still be present due to the apparent relationship of high frequency-low frequency in the forms.

All Tocharian forms seem to have been built on the same material, the distance neutral Proto-Indo-European demonstrative pronouns **so*, **seh$_2$*, and **tod*. Despite this, the motivated distinctions were still maintained between their oppositional deictic terms, applying to all values of both of the languages (Tocharian A and B). Furthermore, the two languages' distal forms all contain /a/-phonemes except for the feminine form of Tocharian B. The vowel of **seh$_2$*

Table 11. The Tocharian system (motivated (+), non-motivated (0), or reversed-motivated (−)).

		Proximal	Medial	Distal	P–M	P–D	M–D
Tocharian A	Masc.	säs	–	sa?		+	
	Fem.	sās	–	sā?		+	
	Neut.	tä?	–	ta?		+	
Tocharian B	Masc.	se	–	samp		+	
	Fem.	sā	–	somp		+	
	Neut.	te	–	tamp		+	
	Masc.	se?	–	samp		+	
	Fem.	sā?	–	somp		+	
	Neut.	te?	–	tamp		+	

have resulted in *ā-* in all feminine forms, again, except of the Tocharian B distal forms in which it has become *o*. If the vowel of the Tocharian B feminine distal forms would have been analogous to the rest of the feminine forms or the rest of the distal forms, it would have been judged non-motivated (Table 11).

All values for Proto-Germanic were judged motivated, though in Old English the system had developed into only the neuter set maintaining its iconicity, the others being non-motivated. The Modern English deictic demonstratives, also motivated, were derived directly from the neuter and only motivated set of Old English (Table 12).

From the Proto-Romani's motivated two-way system (proximal, distal) the (also motivated) Early Romani four-way system (proximal plain, proximal specific, distal plain, distal specific) was created. The Kalderaš Romani three-way system (proximal, medial, distal), also motivated, decreasing f_2 from proximal, through medial to distal, was created (Table 13). This indicates that despite three stages of reorganizations the deictic systems and rigorous changes of the sounds of which the forms are built upon, all languages had motivated deictic forms.

In the Russian system, both proximal and distal forms are related to the Old Church Slavonic distance-neutral demonstrative system. Despite this, through the insertion of a prefix [ɛ-] in the proximal form, all forms were judged motivated.

Table 12. The Germanic system (motivated (+), non-motivated (0), or reversed-motivated (−)).

		Proximal	Medial	Distal	P–M	P–D	M–D
Proto-Germanic	Masc.	*(h)iz	–	*sa		+	
	Fem.	*sī	–	*sō		+	
	Neut.	*(h)it	–	*þat		+	
Old English	Masc.	þes	–	sē		0	
	Fem.	þēos	–	sēo		0	
	Neut.	þis	–	þæt		+	
Modern English		this	–	that		+	

Table 13. The Romani system (motivated (+), non-motivated (0), or reversed-motivated (−)).

	Proximal		Medial		Distal				
	Plain	Specific	Plain	Specific	Plain	Specific	P–M	P–D	M–D
Proto-Romani	*ata		–	–	*ota			+	
Early Romani	*adava	*akava	–	–	*odova	*okova		+	
Kalderaš Romani	kako		kodo		kuko		+	+	+

Icelandic demonstrative pronouns have become distance-neutral, instead the deictic information shifted to the adverbs *hérna* "here" and *þarna* "there," which were motivated. Also, the contrast between the former masculine and feminine pair proximal *sjá* and distal *sá* and *sú* in Old Norse became perhaps more obvious when *sjá* was replaced by *þessi*, yielding a vowel contrast between the proximal and distal forms.

7. Conclusion

No language in this investigation used an identical deictic system to the one reconstructed for Proto-Indo-European (see Beekes 1995), the closest system being the one found in Icelandic. Five major ways of rebuilding material of deictic systems were used: (1) Proto-Indo-European deictic system continued, (2) system-external Proto-Indo-European material used, (3) system-internal merger of forms: Ø used as a system contrast (Gothic only): unknown origin of forms or affixes.

This means that regardless of the reconstruction of the Proto-Indo-European deictic system being correct or not, the deictic systems and forms of the deictic terms have been altered and re-formed via several ways. Despite the various ways of creating new deictic forms, 70.2% of the values of the languages in the investigation were found to have motivated values. Genetic explanations, inherited phonetic forms of the deictic terms, for the high motivated support can be disregarded due to the diversity in rebuilding of forms.

It seems that iconicity is reintroduced in languages time after time after the decay of a former deictic system. More suitable, motivated, forms seem to be preferred when systems are simplified, shown by the motivated values of Proto-Germanic, which became motivated in only one out of three deictic set in Old English, which was the set to survive in Modern English. Using iconicity could be preferred when a deictic system is changed or when all forms of a system are replaced, shown in the origin of Kalderaš Romani and Russian respectively.

In vaguely motivated deictic pairs, the motivated contrast between the terms can be reinforced, as shown by the development from Classical Greek to Greek and Old Norse to Icelandic.

In sum, the contrast was reestablished diachronically again and again, using various strategies for rebuilding. Words which "fit" the meaning of the concepts in question might be unconsciously "chosen" by the speaker as the preferred word, or already existing words are tweaked in a way that they come to carry some of the phonemes that are associated with the semantic domain. Based on the results of this study it seems very likely that iconicity is involved in the rebuilding of deictic systems and forms in Indo-European languages, both contemporary and historically, and it is highly likely that this is the case for other language families as well.

Abbreviations: Alb., Albanian; Arm., Modern Armenian; CArm., Classical Armenian; CG, Classical Greek; Eng., Modern English; ER, Early Romani; Gk., Modern Greek; Goth., Gothic; Hin., Hindi; Hitt., Hittite; Ice., Icelandic; Ir., Modern Irish; Kald., Kalderaš Romani; Lat., Latin; Latv., Latvian; OCS, Old Church Slavonic; OE, Old English; OIr., Old Irish; ON, Old Norse; OPers., Old Persian; Pers., Modern Persian; PGm., Proto-Germanic; PIE, Proto-Indo-European; Port., Portuguese; PR, Proto-Romani; Rum, Romanian; Rus., Russian; Skt., Sanskrit; ToA, Tocharian A; ToB, Tocharian B

Acknowledgements

We would like to thank Joost van de Weijer for assisting in statistical analyses.

Notes on contributors

Niklas Johansson has an MA degree in general linguistics from Lund University, Sweden. His research mainly includes lexical sound symbolism in large-scale quantitative perspectives, dealing with both synchronic and diachronic data. He is the author of articles investigating how sound symbolism operates in various types of fundamental (near-universal) vocabulary.

Gerd Carling is Associate Professor in general linguistics at Lund University, Sweden. Her research deals with mainly historical linguistics, historical typology and syntax. She is the author of several dictionaries, among others on Tocharian and Romani Chib, and currently is research leader of a project on language, culture and geography in Eurasia and the Amazon.

References

Adams, Douglas Q. 1999. *A Dictionary of Tocharian B*. Leiden Studies in Indo-European 10. Amsterdam: Rodopi.

Allen, William Sidney. 1987. *Vox Graeca: A Guide to the Pronunciation of Classical Greek*. 3 [enlarged] ed. Cambridge: Cambridge University Press.

Anderson, Stephen R., and Edward L Keenan. 1985. "Deixis." In *Language Typology and Syntactic Description. Vol. III: Grammatical Categories and the Lexicon*, edited by Timothy Shopen, 259–308. Cambridge: Cambridge University Press.

Antilla, Raimo. 1989. *An Introduction to Historical and Comparative Linguistics*. 2nd ed. Current Issues in Linguistic Theory, 4. Amsterdam: John Benjamins.

Arlotto, Anthony. 1972. *Introduction to Historical Linguistics*. Boston: Houghton Mifflin.

Azevedo, Milton Mariano. 2005. *Portuguese: A Linguistic Introduction*. Cambridge: Cambridge University Press.

Baldi, Philip. 1999. *The Foundations of Latin*. Berlin: Mouton de Gruyter.

Beekes, Robert S. P. 1995. *Comparative Indo-European Linguistics: An Introduction*. Amsterdam: Benjamins.
Beekes, Robert S. P. 2009. *Etymological Dictionary of Greek*. Boston: Brill.
Carling, Gerd. 2000. *Die Funktionen der lokalen Kasus im Tocharischen*. Berlin: Mouton de Gruyter.
Carling, Gerd. 2008. "Phonological Erosion and Semantic Generalization: Notes on the Grammaticalization of the Tocharian Case Paradigm." Working Papers, 53. Department of Linguistics, Lund University.
Carling, Gerd, Lenny Lindell, and Gilbert Ambrazaitis. 2014. *Scandoromani: Remnants of a Mixed Language*. Leiden: Brill.
de Vaan, Micheal. 2008. *Etymological Dictionary of Latin and the Other Italic Languages*. Leiden Indo-European Etymological Dictionary. Leiden: Brill.
Diessel, Holger. 2005. "Distance Contrasts in Demonstratives." In *World Atlas of Language Structures*, edited by Martin Haspelmath, Matthew Dryer, David Gil, and Bernard Comrie, 170–173. Oxford: Oxford University Press.
Diffloth, G. 1994. i: big, a: small. In *Sound symbolism*, edited by L. Hinton, J. Nichols, J.J. Ohala, 107–114. Cambridge: Cambridge University Press.
Dixon, R. M. W. 2010. *Basic Linguistic Theory. Vol. 2: Grammatical Topics*. Oxford: Oxford University Press.
Jespersen, Otto. 1922. *Language: Its Nature, Development and Origin*. London: Allen and Unwin.
Johansson, N., and J. Zlatev. 2013. Motivations for Sound Symbolism in Spatial Deixis: A Typological Study of 101 languages. *The Public Journal of Semiotics [Online]*, Vol 5, No 1. Available at: http://journals.lub.lu.se/ojs/index.php/pjos/article/view/9668/8249
Jóhannesson, Kristinn. 2006. *Isländsk Grammatik, tredje versionen*. Institutionen for svenska språket. Avdelningen för nordiska språk. Göteborgs Universitet.
Kaufman, Terrence. 1994. "Symbolism and Change in the Sound System of Huastec." In *Sound Symbolism*, edited by Leanne Hinton, Joanna Nichols, and John J. Ohala, 63–75. Cambridge: Cambridge University Press.
Kellogg, S. H. 1938. *A Grammar of the Hindi Language*. London: Routledge & Kegan Paul.
Kent, Roland G. 1953. *Old Persian Grammar, Texts, Lexicon*. New Haven, CT: American Oriental Society.
Kiparsky, Paul. 2004. "The Phonological Basis of Sound Change." In *The Handbook of Historical Linguistics*, edited by Brian D. Joseph and Richard D. Janda, 313–342. Oxford: Blackwell.
Kloekhorst, Alwin. 2008. *Etymological Dictionary of the Hittite Inherited Lexicon*. Leiden: Brill.
Köhler, Wolfgang. 1929. *Gestalt Psychology*. New York: Liveright.
Kortlandt, Frederik. 1983. "Demonstrative Pronouns in Balto-Slavic, Armenian, and Tocharian." In *Studies in Slavic and General Linguistics 3: Dutch Contributions to the 9th International Congress of Slavists: Linguistics*, 311–322. Amsterdam: Rodopi.
Kortlandt, Frederik. 1987. "PIE *s in Albanian." In *Studies in Slavic and General Linguistics 10: Dutch Studies in South Slavic and Balkan Linguistics*, 219–226. Amsterdam: Rodopi.
Kroonen, Guus, Erika Langbroek, Harry Perridon, and Annelies Roeleveld. 2011. "Long ū in Germanic." In *Thi Timit Lof: Festschrift für Arend Quak zum 65*, 211–224. Amsterdam: Rodopi.
Lee, Ronald. 2010. *Romani Dictionary: Kalderash-English*. Toronto: Magoria Books.
Lindblad, Per. 1998. *Talets akustik och perception*. Göteborg: Göteborgs Universitet.
Lindsay, Wallace M. 2010. *The Latin Language: An Historical Account of Latin Sounds, Stems, and Flexions*. New York: Cambridge University Press.

Matasović, Ranko. 2009. *Etymological Dictionary of Proto-Celtic*. Leiden: Brill.
Matras, Yaron. 2002. *Romani: A Linguistic Introduction*. New York: Cambridge University Press.
Mayrhofer, M. 1986–2001. *Etymologisches Wörterbuch des Altindoarischen*. Heidelberg: Winter.
Meier-Brügger, Michael. 2000. *Indogermanische Sprachwissenschaft*. Unter Mitarbeit von Matthias Fritz und Manfred Mayrhofer. Berlin: Walter de Gruyter.
Ohala, John J. 1994. "The Frequency Code Underlies the Sound-Symbolic Use of Voice Pitch." In *Sound Symbolism*, edited by Leanne Hinton, Joanna Nichols, and John J. Ohala, 325–347. Cambridge: Cambridge University Press.
Pinault, Georges-Jean. 1989. *Introduction au tokharien*. *Lalies* 7. Actes de la session de linguistique d'Aussois (27 août - 1er septembre 1985). Paris: Presses de L'École Normale Supérieure.
Prokosch, Eduard. 1939. *A Comparative Germanic Grammar, William Dwight Whitney Linguistic Series*. Philadelphia: Linguistic Society of America, University of Pennsylvania.
Puhvel, Jaan. 1997. *Hittite Etymological Dictionary. Vol. 1: Words Beginning with A*. Berlin: Mouton Publishers.
Ringe, Don. 2006. *From Proto-Indo-European to Proto-Germanic*. New York: Oxford University Press.
Saeed, John. 2003. *Semantics*. 2nd ed. Oxford: Blackwell.
Sapir, Edward. 1929. "A Study in Phonetic Symbolism." In *Selected Writings*, edited by David Mandelbaum. Berkeley: University of California Press.
Schmalstieg, William R. 1995. *An Introduction to Old Russian*. *Journal of Indo-European Monograph Series* No 15. Washington, DC: Institute for the Study of Man.
Schmitt, Rüdiger. 1981. *Grammatik des Klassisch-Armenischen mit sprachvergleichenden Erläuterungen*. Innsbrucker Beiträge zur Sprachwissenschaft, 123. Institut für Sprachwissenschaft der Universität Innsbruck.
Sigurd, Bengt, Mats Eeg-Olofsson, and Joost van de Weijer. 2004. "Word Length, Sentence Length and Frequency: Zipf Revisited." *Studia Linguistica* 58 (1): 37–52. doi:10.1111/j.0039-3193.2004.00109.x.
Szemerényi, Oswald. 1989. *Einführung in die vergleichende Sprachwissenschaft*. 3, vollständig neu bearbeitete Auflage. Darmstadt: Wissenschaftliche Buchgesellschaft.
Thumb, Albert. 1959. *Handbuch des Sanskrit*. Heidelberg: Carl Winter Universitätsverlag.
Tragut, Jasmine. 2009. *Armenian: Modern Eastern Armenian*. Amsterdam: John Benjamins.
Traunmüller, Hartmut. 1994. "Sound Symbolism in Deictic Words." In *Tongues and Texts Unlimited: Studies in Honour of Tore Jansson on the Occasion of his Sixtieth Anniversary*, edited by Hans Auli and Peter af Trampe, 213–234. Stockholm: Department of Classical Languages, Stockholm University.
Ultan, Russel. 1978. "Size-Sound Symbolism." In *Universals of Human Language 2, Phonology*, edited by Joseph Greenberg, 525–567. Stanford: Stanford University Press.
Vasmer, Max. 1958. *Russisches etymologisches Wörterbuch*. Heidelberg: Carl Winter Universitätsverlag.
Waanders, Frederik M. J. 1997. "Particulars of Some Proto-Indo-European Particles." In *New Approaches to Greek Particles*, edited by Albert Rijksbaron. Amsterdam: Gieben.
Woodworth, Nancy L. 1991. "Sound symbolism in proximal and distal forms." *Linguistics* 29 (2): 273–299. doi:10.1515/ling.1991.29.2.273.

Online resources

[DEX '98] Dicţionarul explicativ al limbii române, ediţia a II-a, Academia Română, Institutul de Lingvistică "Iorgu Iordan", Editura Univers Enciclopedic, 1998.