Title
The Sound of Valence: Phonological Features Predict Word Meaning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors
Dinnissen, Karlijn
Louwerse, Max M

Publication Date
2015

Peer reviewed

The Sound of Valence: Phonological Features Predict Word Meaning

Karlijn Dinnissen (karlijn_dinnissen@hotmail.com)
Tilburg University, Warandelaan 2
5037 AB Tilburg, The Netherlands

Max M. Louwerse (m.m.louwerse@uvt.nl)
Tilburg Center for Cognition and Communication, Warandelaan 2
5037 AB Tilburg, The Netherlands

Abstract

Various studies have recently shown that the long-held claim that the relation between the sound of a word and its meaning is arbitrary needs to be revisited. In two computational studies we investigated whether word valence can be derived from sound features in English, Dutch and German. In Study 1, we identified the extent to which individual phonological features explained valence scores per language separately. In Study 2, we aimed to determine the optimal combination of cues that can predict valence scores across the three languages using two statistical classifiers and four machine learning classifiers. Our results showed that frequency and word complexity were the most reliable shared cues to predict valence for all three languages, obtaining a correct valence classification of about 60%. This percentage could be enhanced for individual or pairs of languages using additional relevant cues. These findings demonstrated that the claim that the relation between the sound of a word and its meaning is arbitrary is too strong.

Keywords: arbitrariness; sound-meaning; phonology; symbol grounding.

Introduction

In *Cratylus*, Plato reported a debate in which Socrates and two of his pupils consider whether words might be a natural reflection of the things being named, or whether language attaches form to content by convention, the relationship between word and meaning therefore being arbitrary. In this dialogue, Socrates tends to agree that natural reflection is superior to an arbitrary relation.

However, modern linguistics has generally accepted the contrary, at least until recently. De Saussure (1916) described a sign as an entity that consists of two abstract parts: the concept and the sound image. According to De Saussure, this system can only exist if the initial matching between concept and sound image happens in an entirely arbitrary fashion. In a similar vein, Hockett (1960) argued that the sound image has no connection to the object it represents, other than the initial matching (De Saussure, 1916; Hockett, 1960). Both scholars, however, worked with skewed databases containing small amounts of data. In light of larger databases, as well as stronger computational algorithms and a more advanced theoretical framework, the strong arbitrariness claim should be reconsidered. The purpose of the current paper is to add to the existing literature and invite cognitive scientists to not univocally adhere to the arbitrariness principle without further investigating the extent to which phonological cues relate to meaning.

Arbitrary sound-meaning relations

For a language system, to have arbitrary relations between a word's sound and its meaning is, to a certain extent, advantageous. Arbitrariness provides an increased economy of reference (Burling, 1999). It is common for both human and animal communication systems that over time iconic symbols (e.g., gestures) become arbitrary symbols (Corballis, 2002). Indeed, as the number of concept and sound image-pairs increases, Gasser (2004) proposed, the chance of two different concepts sharing the same sound-image increases as well, leaving less 'space' in a language for new words and meanings (Gasser, 2004).

Arbitrariness can help in the language learning process by keeping concepts with similar meanings separate for small languages. It allows context information to have a maximum impact on the mapping from sound image to meaning, while an iconic relationship between the two would impede learning (Gasser, 2004). The beginning of a word would then be most relevant for information about the unique word, while greater systematicity between word form and general category could be found in the second half of the word, a hypothesis that studies have indeed found support for (Gasser, 2004; Monaghan et al., 2011).

The arbitrariness notion also fits theories of embodied cognition. According to one embodied cognition account, meaning in language is conveyed by amodal, abstract, and – importantly – arbitrary symbols (Glenberg & Kaschak, 2002). Because the sound of a word cannot say anything about its meaning, meaning must be grounded outside the language system, in referents in the world or perceptual simulations of these referents.

In conclusion, the evidence favoring de Saussure's (1916) claim '*le signe est arbitraire*' is strong. But the claim itself is equally strong, leaving no room for non-arbitrary elements.

Non-arbitrary sound-meaning relations

A non-arbitrary relation (i.e., not entirely arbitrary) is in line with several fields of research, though we do not argue that the relationship between phonological cues and meaning is fixed. When the language system at large is considered, many non-arbitrary elements can be pointed out. For instance, binomials have a predictive order and there is a significant difference between one relation versus the other, so that we say here and there (Cooper & Ross, 1976), today and tomorrow (Louwerse, Raisig, Hutchinson & Tillman, under review), high and low (Louwerse, 2008), happy and sad, men and women, teacher and student (Hutchinson & Louwerse, 2013) and few and many (Hutchinson & Louwerse, 2014). Language users make use of pre-linguistic conceptualizations, so that phonological cues – akin to distributional semantic cues – are encoded in language (Benor & Levy, 2006; Louwerse, 2008).

For the relationship between sound and meaning within the word, there are patterns suggesting that claiming this relation is arbitrary is too strong as well. Sound symbolism argues that distinctive features in a word (e.g. a syllable or phoneme) bring out a correspondence in their meaning and in human emotional attitudes toward the word (Ahlner & Zlatev, 2010). This is most obvious in onomatopoeias: words that directly phonetically imitate or resemble what the source of the sound is, such as a certain animal ('meow') or an inanimate object ('boom') (Ahlner & Zlatev, 2010; de Saussure, 1916; Dingemanse, 2012). There are also categories in which this mapping is less obvious, such as ideophones. These are words that depict sensory imagery, ranging from depicting visual patterns to cognitive states, and even direct mimicking of sounds. This phenomenon is for instance present in Japanese, e.g. 'koro-koro' signifying a small object rolling, and in several African languages (Ahlner & Zlatev, 2010; Dingemanse, 2012).

Even beyond these categories, research has been conducted in which meaning was successfully mapped to sound. Sound was connected with size (Ohala, 1997), distance (Tanz, 1971) and shape (as in 'bouba'/'kiki') (Ramachandran & Hubbard, 2001).

Finally, for language acquisition there is evidence for, and there are advantages to, non-arbitrary relations between the sound of a word and its meaning. Even though arbitrariness supports individuation for words, especially words learned later in a learning process, words typically acquired early in the language learning process tend to be more non-arbitrary. Meaning in sound can enable early language learners to discover that words are representations of objects and concepts in the world around them (Monaghan, Shillcock, Christiansen, & Kirby, 2014). Arbitrariness becomes more useful when more complex meanings come into play later on (Gasser, 2004; Lewis et al., 2014; Monaghan et al., 2011, 2014).

It is important to note that while considering non-arbitrary sound-meaning relationships, neither sound nor meaning should be defined overly conservative. For instance, 'meaning' of the word can be the grammatical category of the word (state/event/process versus person/object) (Parsons, 1990). Monaghan et al. (2007) related phonological and prosodic properties of words to several grammatical categories for several languages. When predicting whether a word was a noun or a verb using stepwise discriminant analysis, they obtained an accuracy score of 67% (with 62% as a random base-line) for English, and 89% (with a baseline of 63%) for Dutch. This became higher when taking distributional cues into account, and the most effective when phonological and distributional cues were combined (Monaghan, Christiansen, & Chater, 2007).

Similarly, the 'sound' of a word can be translated into word complexity. Lewis, Sugarman, & Frank (2014) found that unknown long words were more likey to be paired with complex objects and short words with simple objects, and ascribe this to an innate 'complexity bias' that can even be retraceable to the naming of objects thousands of years ago: the more frequent an object was encountered, the more likely a short and easy name was chosen for it, as opposed to longer and more complex names for less frequent objects (Lewis, Sugarman, & Frank, 2014).

Based on the literature on the arbitrariness of the relation between sound and meaning, the conclusion to be drawn is that there is no straightforward answer to the question whether sound-meaning relations are arbitrary. To address both views as a division of labor may be the best way of dealing with both standpoints (Monaghan et al., 2011).

The sound of valence

The current study investigated whether the phonological features of a word can predict the valence of a word. Valence seems to be one of the core semantic features in language (Nielsen, 2011; Warriner, Kuperman, & Brysbaert, 2013) and evidence for the non-arbitrary relation between sound and meaning should therefore be expected on the positive-negative continuum of meaning as well.

There are some studies that point out that valence and linguistic features are related. For instance, the positivity bias entails that humans have the tendency to use positive words more frequently, as there are overall more positive events happening in the world than negative ones, and people generally experience mild positive emotions (Augustine, Mehl, & Larsen, 2011; Rozin, Berman, & Royzman, 2010). Additionally, negative adjectives tend to be longer, as they often consisted of their positive antonyms with a prefix (e.g. 'happy'/'unhappy', 'sincere'/'insincere') (Augustine et al., 2011; Rozin et al., 2010).

In Study 1 we aimed to identify which individual phonological cues are correlated with valence ratings in English, German and Dutch. In Study 2, a combination of these cues will be used to determine how well valence can be predicted on the basis of sound features across languages.

Study 1

Valence. Valence was operationalized using the ratings according to the *Affective Norms for English Words (ANEW)*-list originally created by Bradley and Lang (1999), which entails 1,034 words and their valence ratings and was recently extended by Warriner, Kuperman and Brysbaert (2013) to nearly 14,000 words (Bradley & Lang, 1999; Warriner et al. 2013).

Phonological cues. Phonological features were identified using CELEX (Baayen, Pipenbrock, & Gulikers, 1995), which provides morphological, phonological, syntactic and frequency information for 52.447 English, 51.728 German and 124.136 Dutch word types (lemmas).

Four categories of phonological cues were used, concerning 1) the whole word, 2) the word's vowels, 3) word onset (all consonants before the first vowel) and 4) first consonant. Unless stated otherwise, the total count of a cue per word was used for analysis. These variables were informed by (Monaghan et al., 2005; 2007) and derived from the International Phonetic Alphabet. They can be seen in Table 1, where cues used for onset and first consonant are printed bold. In order to avoid giving the impression that all of the features we looked at yielded significant effects, we included all considered features in Table 1. 33% of them ultimately had a significant relation with the valence score, as can be seen in the results.

The ANEW-list (which consists of English words) was translated in Dutch and German using Google Translate, as the translation accuracy for Google Translate has been satisfactory, especially among Western languages (Aiken & Balan, 2011; Balk et al., 2012). Any anomalies in the translations were corrected.

The English, Dutch and German word lists and their ANEW-scores were combined with the CELEX phonological representation and frequency, and discarded if there was none. For the two translated languages, words not available in the Google Translate-database would remain English. If this English version was not be present in the Dutch or German CELEX-database, it was filtered out for that language. This resulted in 91.3% of the original ANEW list for English (13,001 words in total), 88.1% (12,536 words) for Dutch, and 68.3% (9,718 words) for German.

Table 1. Phonological Cues

whole word	vowels
length in syllables	reduced vowels
length in phonemes	position: front
frequency	position: near front
vowels	position: central
consonants and affricates	position: near back
word complexity*	position: back
vowel density**	height: close
coronals	height: near close
voiced consonants	height: close mid
plosives	height: mid
nasals	height: open mid
fricatives	height: near open
approximants	height: open
bilabials	rounded
velars	nasal
alveolars	
palatals	
labials	
glottals	
dentals	

*: the percentage of consonants compared to total number of letters
**: total number of vowels (counting long vowels and diphthongs double) divided by total number of letters

The phonological features of the 10% most positive words (with a valence score of 7.05 or higher) and the 10% most negative words (with a valence score of 2.75 or lower) were compared. Alternative ways to analyze the dataset, for instance by taking the entire original word list or a tertiary split of the data file, resulted in overall comparable results.

Results and Discussion

For each of the three languages separately, the cue selection was conducted using bivariate correlations on all three data sets to check which cues were linearly related to valence score. If the cues did not significantly correlate with valence in more than one data set, they were excluded from further analyses for that particular language, for any effect they might have on valence would not be significant. Subsequently, a mixed-effects model was conducted on all data sets, because of the ability of the model to make measurements on clusters of related statistical units, in this case for example linguistic categories that overlap to a certain degree. After both analyses were conducted, their results were used to filter out cues that proved to have no effect on the valence score for the language in question.

A mixed effects model using the phonological cues as fixed factors and the valence scores as dependent variables, showed significant relations between approximately 10 cues per language and their valence scores. As Tables 2a, 2b and 2c demonstrate, five cues (all at the word level) correlate with valence scores for all three languages. Valence scores negatively correlated with consonant score, syllables and length in phonemes. Frequency and word complexity positively correlated with valence score. However, the few cues that are marked not showing a significant effect in these tables were in fact significant in both the original dataset and the tertiary split dataset and were therefore still considered in the subsequent analyses.

These findings support the positive language bias and the hypothesis stated earlier by Augustine et al. (2011) with more frequent words more likely to have a positive connotation (Augustine et al., 2011). Somewhat surprisingly, the findings only correspond with the complexity bias (not to be confused with word complexity) for German, where negative words are significantly longer.

Table 2a. Phonological cues predicting valence (English)

Cue	*F*-score	*Corr. direction*
Word:		
Frequency	34.93**	+
Consonants	8.35**	-
Word complexity	1.23	+
Voiced consonants	12.74**	-
Glottals	4.07*	-
Dentals	21.62**	-
Height near close	8.64**	-
Rounded vowels	5.29*	-
Onset:		
Fricatives	7.90**	+
Approximants	9.30**	+
Alveolars	2.22	-

Table 2b. Phonological cues predicting valence (Dutch)

Cue	*F*-score	*Corr. direction*
Word:		
Frequency	39.94**	+
Syllables	22.47**	-
Word complexity	2.72	+
Voiced consonants	15.05**	-
Approximants	15.52**	-
Velars	11.16**	-
Palatals	1.77	+
Height near close	4.49*	-
Onset:		
Approximants	0.59	-
First Consonants:		
Coronals	5.43*	-
Nasals	10.76**	-

Table 2c. Phonological cues predicting valence (German)

Cue	*F*-score	*Corr. direction*
Word:		
Frequency	25.00**	+
Consonants	12,74**	-
Length in phonemes	7.95**	-
Word complexity	9.00**	+
Vowel density	2.69	-
Approximants	17.56**	-
Velars	9.18**	-
Rounded vowels	8.88**	-
Onset:		
Alveolars	0.29	-
First consonants:		
Voiced consonants	6,25*	+
Nasals	2.07	-
Approximants	1.12	+

Even though these findings indicate a significant difference between high and low valence of words on the basis of individual sound features, Study 1 does not indicate whether phonological cues predict valence across language for a combination of a small set of phonological cues.

Study 2

In the second study we investigated which phonological cues would predict valence across the three languages. The principle parsimony was adhered to, preferring an easy (fewer phonological cues) explanation to a complicated explanation (many phonological cues).

In order to find the combination of the least cues with an accuracy as high as possible, all cues that were deemed to be important for the language(s) concerned were used in a stepwise discriminant analysis and a binary logistic regression, using a leave-one-out cross-validation.

When a cue was not used in the discriminant analysis model and/or the accuracy of the cues for both tests was not affected when removing this cue, it was removed from the analyses permanently. For the logistic regression accuracy, the original grouped cases were used, while the cross-validated grouped cases-percentage was used for the discriminant analysis.

Additionally, a Pearson bivariate correlation test was conducted in case two cues would be significantly correlated, and therefore would account for the same effect on the valence score. For example, it might be suspected that length in phonemes and length in syllables would be correlated, as they both concern the word length. If this was the case, either the cue that correlated with more than one other cue or the cue that had the least effect on the accuracy score was discarded.

When the selection of cues was finished, the accuracy of the valence prediction based on an enter discriminant analysis and a binary logistic regression was computed. The resulting accuracy is a percentage, where 50% indicates a chance-level performance and 100% translates to a perfect performance in predicting a word's valence category (either positive or negative).

To avoid that these findings could be attested to the choice of statistical classifiers, four common classification algorithms in Weka (Waikato Environment for Knowledge Analysis) 3.7.10 were used: NaiveBayes, LibSVM, MultilayerPerceptron and Logitboost. A 10-fold cross-validation was used in all classification tasks to prevent overfitting.

Results and Discussion

Figure 1 represents a Venn diagram in which an overview of all selected cues and their obtained average accuracy scores can be found.

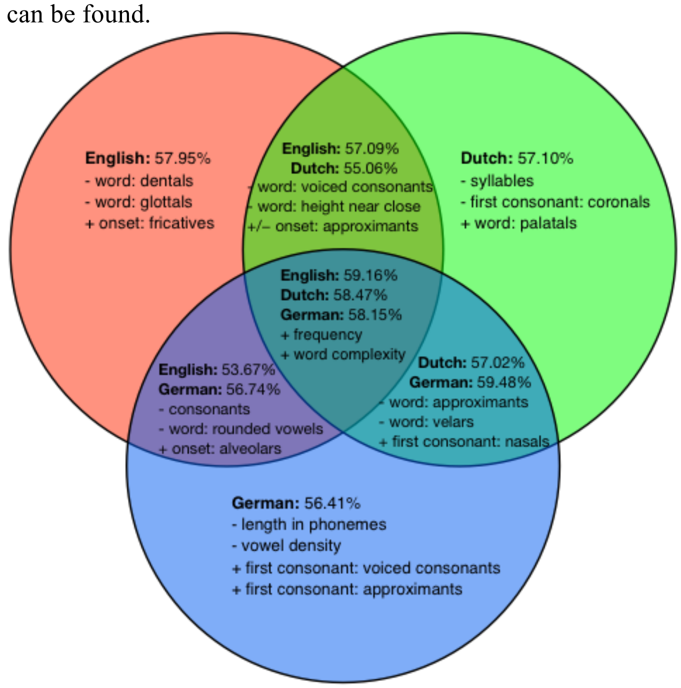

Figure 1. Venn diagram of cues for distinguishing between high and low valence words in English, Dutch and German, with their obtained average accuracy. A plus preceding a cue indicates that it is positively significant and a minus preceding a cue indicates negatively significant correlation.

The percentage of correctly classified words using all cues relevant for a language combined was even higher: 60.3% for Dutch, 60.6% for English and even 61.2% for German.

Some may find the significant 8-percent-above-chance level unimpressive opposed to percentages obtained in previous studies. However, those mostly focused only on selected cues for languages separately, as in the study by Monaghan et al. (2007) when they discerned open/closed class words and nouns/verbs (Monaghan et al., 2007), whereas the current studies used real words, adhered to parsimony and emphasized cues relevant for more than one language. The obtained accuracy percentage is therefore considerable all the more.

Cues concerning the whole word, the onset of words and the first consonant of words were all part of the selection as was expected (Gasser, 2004; Monaghan et al., 2011). Although the first consonant proved not to be useful in predicting valence for English words regardless of language combination, it was conversely quite important in the prediction for German words (voiced consonants, approximants and nasals were all selected). Furthermore, all selected cues correlated in the same direction for each language (apart from the approximants in the onset). This further suggests that the cues can be used similarly to predict valence in various languages.

As Figure 1 shows, the cues that proved to be the most useful for classifying words on their valence in all three languages were frequency and word complexity. There was no significant interaction and therefore no co-linearity between frequency and word complexity (as checked with a bivariate correlation on all three data sets). The detailed results of the accuracy analyses for these two cues are reported in Table 3. For clarification, an example of a positive word with high frequency, complexity and valence score in ANEW is 'relationship', while 'comatose' is a negative word with low frequency and complexity.

For all languages, the accuracy was above chance level, as it was 58.15% or higher. Frequency and word complexity both positively correlated with valence, again confirming the positivity bias.

General Discussion

In two studies, using a large database of valence ratings and phonological features, we have demonstrated that approximately 58% of the valence scores can be predicted using only two linguistic features: frequency and word complexity.

These two studies present accumulating evidence for the positivity bias, as frequency indeed correlated with valence score positively – even being selected as a cue that could predict valence above chance level for all three languages under investigation here. However, contrary to the expectation posed by the complexity bias that positive words would be shorter than negative words, this was only the case for German. The reason for this is not quite known, but is at least of interest to our ongoing investigations.

A possible explanation might lie in what could also be considered a weakness of this study: the fact that only Germanic languages were used. In that sense, the current data still results in a skewed database, as Ahlner and Zlatev (2010) described, and results that are not representative for human language in general (Ahlner & Zlatev, 2010). However, the similarities in results across languages become all the more interesting when considering systematicity in a conventional way (meaning that phonological representation and meaning are associated with each other out of convention instead of a reflection of cognitive nature), as this could possibly mean an advantage in language learning. Fully determining the type of systematicity remains a challenge for further research.

Additionally, as Google Translate does not deliver perfect translations (e.g. takes context cues into account), it would be interesting to use a professionally translated database in future work. Affixes might be addressed in following studies as well, as approximately 4-5% of the negative words used in the analyses contained an affix. Finally, a future study using solely phonological cues would be interesting to further consider to what extent sound relates to meaning.

Despite of this, the results presented in this paper demonstrate that the relationship between the sound of a word and its meaning is not simply arbitrary, as some studies claim (Glenberg & Kaschak, 2002). The accumulating evidence that the sound-meaning relation is not arbitrary (Lewis, et al., 2014; Monaghan, et al, 2005; 2011; 2014) not only places a different perspective on a century-old claim on the nature of language, but also sheds light on the nature of cognition. A common view in the cognitive sciences is the notion that cognition is fundamentally embodied because no meaning can be derived from the (sound of the) word itself (Glenberg & Kaschak, 2002). If evidence indicates that at least some meaning can be derived from the sound features of a word, it would suggest that both language statistics and perceptual simulation can explain language processing, a claim we have advocated elsewhere (Louwerse, 2011).

Table 3. Accuracy of valence prediction for cues 'frequency' and 'word complexity' in English, Dutch and German

	Discriminant analysis	Logistic regression	Avg. Statistical	Naive-Bayes	LibSVM	Multi-layer Perceptron	LogitBoost	Avg. machine learner	*Avg. Overall*
English	58.1%	61.2%	59.65%	54.89%	60.28%	57.66%	61.82%	58.66%	*59.16%*
Dutch	56.9%	59.3%	58.1%	54.69%	63.01%	56.37%	61.25%	58.83%	*58.47%*
German	56.8%	60.0%	58.4%	53.86%	60.28%	57.77%	59.68%	57.90%	*58.15%*

The findings in this paper again suggests that De Saussure's (1916) claim '*le signe est arbitraire'* might be elegant, but is also too strong, leaving no room for non-arbitrary elements. We have argued that, at least for the sound of valence, non-arbitrary elements make De Saussure's (1916) claim approximately 60% false.

References

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*(1/4), 248-398.

Aiken, M., & Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal*, *16*(2).

Augustine, A. A., Mehl, M. R., & Larsen, R. J. (2011). A positivity bias in written and spoken English and its moderation by personality and gender. *Social Psychological and Personality Science*, *2*, 508–515.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Pennsylvania: Linguistic Data Consortium.

Balk, E. M., Chung, M., Hadar, N., Patel, K., Yu, W. W., Trikalinos, T. A., & Kong Win Chang, L. (2012). *Accuracy of data extraction of non-English language trials with Google Translate*.

Benor, S. B., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language*, 233-278.

Burling, R. (1999). Motivation, conventionalization, and arbitrariness in the origin of language. *The origins of language: What nonhuman primates can tell us*, 307-350.

Cooper, W. E., & Ross, J. R. (1975). World order. In R. E. Grossman, L. J. San, & T. J. Vance (Eds.), *Papers from the parasession on functionalism* (pp. 63-111).

Bradley, M., & Lang, P. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (pp. 1–49).

Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton: Princeton University Press.

De Saussure, F. (1916). *Cours de linguistique générale* (C. Bally & A. Sechehaye, eds.), trans. W. Baskin, Course in General Linguistics, Glasgow: Fontana/Collins, 1977.

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

Gasser, M. (2004). The origins of arbitrariness in language. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 4–7).

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558-565.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, (203), 5–12.

Hutchinson, S., & Louwerse, M. M. (2013). Statistical linguistic context and embodiment predict metaphor processing but participant gender determines how much. *Cognitive Linguistics*, 24, 667–687.

Hutchinson, S., & Louwerse, M. M. (2014). Language statistics explains spatial-numerical association of response codes. *Psychonomic Bulletin & Review, 21*, 470-478.

Lewis, M., Sugarman, E., & Frank, M. C. (2014). The structure of the lexicon reflects principles of communication. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp.845-850). Austin, TX: Cognitive Science Society.

Louwerse, M. M. (2008). Embodied representations are encoded in language. *Psychonomic Bulletin and Review, 15*, 838-844.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3*, 273–302.

Louwerse, M.M., Raisig, S., Hutchinson, S. & Tillman, R. (under review). Time after time in words: Chronology through language statistics.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*, 143-182.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*, 259–305.

Monaghan, P., Christiansen, M. H., & Fitneva, S. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology. General*, *140*(3), 325–47.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology. Leaning, Memory and Cognition*, *38*(5), 1–50.

Monaghan, P., Shillcock, R., Christiansen, M., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Nielsen, F. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs.

Ohala, J. J. (1997). Sound Symbolism. In *4th Seoul International Conference on Linguistics* (pp. 98–103).

Parsons, T. (1990). *Events in the Semantics of English*. Cambridge, Ma: MIT Press.

Ramachandran, V., & Hubbard, E. (2001). Synaesthesia -- A Window Into Perception, Thought and Language. *Journal of Consciousness Studies*, *8*(12), 3–34.

Rozin, P., Berman, L., & Royzman, E. (2010). Biases in use of positive and negative words across twenty natural languages. *Cognition and Emotion*, *24*(3), 536–548.

Tanz, C. (1971). Sound symbolism in words relating to proximity and distance. *Language and Speech*, 266–276.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–207.