

## How arbitrary is language?

**Authors:** Padraic Monaghan<sup>1\*</sup>, Richard C. Shillcock<sup>2,3</sup>, Morten H. Christiansen<sup>4,5</sup>, Simon Kirby<sup>3</sup>

<sup>1</sup>Centre for Research in Human Development and Learning, Department of Psychology,  
Lancaster University, Lancaster UK

<sup>2</sup>School of Informatics, University of Edinburgh, Edinburgh UK

<sup>3</sup>School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh  
UK.

<sup>4</sup>Department of Psychology, Cornell University, Ithaca NY, USA.

<sup>5</sup>Haskins Laboratories, New Haven, CT, USA.

\*Correspondence to: [p.monaghan@lancaster.ac.uk](mailto:p.monaghan@lancaster.ac.uk).

## Abstract

It is a long established convention that the relationship between sounds and meanings of words is essentially arbitrary – typically the sound of a word gives no hint of its meaning. However, there are numerous reported instances of systematic sound-meaning mappings in language, and this systematicity has been claimed to be important for early language development. In a large-scale corpus analysis of English, we show that sound-meaning mappings are more systematic than would be expected by chance. Furthermore, this systematicity is more pronounced for words involved in the early stages of language acquisition and reduces in later vocabulary development. We propose that the vocabulary is structured to enable systematicity in early language learning to promote language acquisition, whilst also incorporating arbitrariness for later language in order to facilitate communicative expressivity and efficiency.

One of the central “design features” of human language is that the relationship between the sound of a word and its meaning is arbitrary<sup>1,2</sup>; given the sound of an unknown word it is not possible to infer its meaning. Such a view has been the conventional perspective on vocabulary structure and language processing in the language sciences throughout much of the last century (see Perniss, Thompson, & Vigliocco<sup>3</sup> for review). Since de Saussure’s<sup>1</sup> notion of the arbitrariness of the sign, such a property has been assumed to be a language universal property, and has even assumed a definitional characteristic: According to Hockett<sup>2</sup>, for instance, a communication system will not count as a language unless it demonstrates such arbitrariness. In contrast, throughout most of human intellectual history<sup>4,5</sup>, the sound of a word was often assumed to directly express its meaning, a view recently revived in studies exploring sound-symbolism<sup>6,7,8</sup>. So, is spoken language arbitrary or systematic?

Sound-meaning mappings may be non-arbitrary in two ways<sup>9</sup>. First, through absolute iconic representation where some feature of the language directly imitates the referent, as in onomatopoeia. For example, incorporating the sound that a dog makes into the sign for the sound itself (i.e., *woof woof*) is one example of this absolute iconicity. Second, the sound-meaning mapping could be an instance of relative iconicity, where statistical regularities can be detected between similar sounds and similar meanings though these may not be restricted to imitative forms<sup>3</sup>. In this case, the iconicity is not transparent, but is generally only observable once knowledge of the sound-relationships and meaning-relationships is determined. An example of this is for certain phonoaesthemes<sup>6</sup>, such as *sl-* referring to negative or repellent properties (e.g., *slime, slow, slur, slum*). Other phonoaesthemes may indeed represent absolute iconicity (such as *sn-* referring to the nose via onomatopoeic properties of its functions), and there is debate about which phonoaesthemes are indeed absolute or relative in their iconicity. Nevertheless, in the

literature, both of these forms of iconicity have been referred to as systematicity in sound-meaning mappings, to contrast with arbitrariness. In spoken language it is not clear that absolute iconicity could occur without relative iconicity. In the case of onomatopoeia, for instance, the iconic relationship between the actual sound the animal makes and the linguistic sign carries some relationship to the nature of the beast (front vowels are more likely in words for small animals' calls than large animals' calls, compare *cheep cheep* for chicks versus *roar* for a lion). Hence, such instances of absolute iconicity are likely to be reflected in relative sound similarity measures.

Arbitrariness of form-meaning mappings introduces a profound cost for learning: as the mapping between the sign and its referent has to be formed anew for each word, knowing all the other words in the vocabulary does not assist in learning a new word. Besides the cost for processing and learning of the language, to Renaissance scholars the absence of apparent systematicity between form and meaning was seen as an offensive property of language<sup>4</sup>. Arbitrariness was interpreted in terms of the story of the Tower of Babel, in which a previously globally understood language was confounded through divine intervention. There are numerous accounts of scholars aiming to rediscover the “universal language” – the pre-Babel tongue where form and meaning were perfectly aligned. John Wilkins, a founder of the Royal Society, produced one of the most complete systems of language that related forms closely to meanings, a system exemplifying relative iconicity<sup>10</sup>. Wilkins' language, entertainingly depicted in Eco's<sup>4</sup> treatise, formed a hierarchy of categories of increasing specificity, with each category and subcategory indicated by a particular letter. For instance, in Wilkins' system, plants begin with the letter “g”, and animals with the letter “z”. Then, for the subcategories of animals, exanguious animals begin with “zα”, fish begin with “za”, birds with “ze”, and beasts with “zi”. For further

subcategories, additional letters are appended. Such a language would clearly result in much inheritance of information across words. So, on encountering a new word, the general meaning could be determined based on its form.

However, computational modelling and experimental studies of vocabulary acquisition have suggested that arbitrariness may, contrary to initial expectations, actually result in a learning advantage. In a series of connectionist computational models, that learned to map phonological forms of words onto meaning through an associative learning mechanism, Gasser<sup>11</sup> demonstrated that, as the size of the vocabulary increased, arbitrariness in the mappings between inputs and outputs of the model resulted in better learning. This result was interpreted as being due to greater flexibility in the interleaving of new items into an already learned set of mappings. For systematic sound-meaning mappings, the resources assigned to the new word are recruited from those already assigned to mapping between similar words, whereas for arbitrary mappings, the resources for learning the new word can be drawn from anywhere in the system. For an associative learning system, learning to form a mapping can be similar to discovering the principal components from the input-output pairings<sup>12</sup>. For systematic mappings, the set of mappings can be effectively described with a single component, and space on this component can become crowded. For arbitrary mappings, a separate component is required for every mapping individually, reducing the possibilities for interference between words represented by distinct components.

In a series of experimental and computational studies, Monaghan, Christiansen and Fitneva<sup>13</sup> demonstrated that for learning novel words, arbitrariness in the sound-meaning mapping was advantageous compared with a vocabulary with a systematic form-meaning mapping. However, this advantage was only prominent when an additional contextual cue was

provided for the learner within the language, either in the form of co-occurrence with a word that related to the general categorical meaning of the word, or in terms of a morphological feature that related to category. For instance, in this contextual cue condition, utterances comprised a marker word (either “weh”, which always occurred when the referent was an object or “muh” which always occurred when the referent was an action) along with a referring word (e.g., “paab”), which was heard simultaneously with viewing a picture referent. Arbitrariness or systematicity was carried in the relationships between the sounds of the referring words and the category distinction between objects and actions in the set of referents. Without the marker word (“weh” or “muh”), learning was not advantageous in the arbitrary condition. In the same study, the computational studies were connectionist models that implemented an associative learning mechanism in order to learn to map form onto meaning representations, either with or without context. Again, when context was present the arbitrary mapping was optimal for learning. Analysis of the computational model’s solution to the mapping demonstrated that arbitrariness permits maximising of the potential information in the learning situation, resulting in effective mapping to be achieved. In the systematic condition, words with similar sounds occurred in similar contexts, reducing distinctiveness in the environment for identifying the intended referent, and resulting in less effective mappings to be formed. These effects were precisely in line with Wilkin’s own errors in transcription whereby closely related words suffered mislabelling: Eco notes that *Gade* (barley) was written in place of *Gape* (tulip) in Wilkins’ Essay<sup>4</sup>.

In contrast to the view of the arbitrariness of the sign, there are a growing number of corpus analyses and behavioural studies that demonstrate some systematicity in spoken language. For some features of meaning, such as vowel quality relating to size, the sound-symbolic

properties are language-universal<sup>6,7,9</sup>; for instance, the nonwords “mil” and “mal” are typically understood to express small and large, respectively, across cultures. High and low vowel contrasts, exemplified by the i/a distinction, have also been shown to occur in small/large expressives, respectively, across most, if not all, languages<sup>14</sup>. There are also numerous language-specific properties, such as phonoaesthemes that refer to clusters of phonemes relating to specific meanings. For example, in English, words associated with the nose and its functions tend to begin with *sn-*, or words referring to light often begin with *gl-*<sup>6</sup>. Preferences for certain sound-meaning relationships, have been demonstrated to affect learning of novel adjectives<sup>15</sup>, verbs<sup>16,17</sup>, nouns<sup>18,19</sup>, and mixes thereof<sup>20</sup> though these studies generally test a forced choice between two alternatives. When the semantic distinction is not immediately available, as in a forced-choice test between two objects from different categories, then learning is less evident but still present under some learning conditions<sup>21</sup>.

Sound symbolism has been proposed to be vitally important for language acquisition because inherent properties of meaning in sound would enable children to discover that words refer to the world around them. Sound-symbolic words not only represent their meaning, but can literally incorporate the senses to which they refer within the sound, as in onomatopoeia. This mechanism could facilitate acquisition not only of particular sound-meaning mappings, but also the knowledge that there *are* mappings between sounds and meaning<sup>8</sup>. Such preferences for certain sound-meaning mappings have now been shown for young children. For instance, there are numerous studies with adults demonstrating that nonsense words such as *bouba* and *kiki* are found to reliably relate to rounded and angular objects, respectively (see Monaghan et al.<sup>21</sup> for review). However, Ozturk, Krehm and Vouloumanos<sup>22</sup> demonstrated that 4 month old children have a similar preference, indicating that substantial knowledge about language is not required in

order to form these preferences. Similarly, Walker, Bremner, Mason, Spring, Mattock, Slater and Johnson<sup>23</sup> showed that 3- to 4-month old infants were able to form cross-modal correspondences between spatial height and angularity with auditory pitch, demonstrating that cross-modal correspondence preferences can precede substantial language learning rather than being a consequence of the fact that a particular language instantiates these correspondences<sup>24</sup>.

Yet, we have seen that systematicity in sound-meaning mappings in the vocabulary comes at a cost in terms of reducing the distinctiveness of words that have similar meanings, potentially increasing confusion over intended meaning<sup>9,13</sup>. So, given this tension between the linguistic convention of arbitrariness and the growing body of studies demonstrating sound symbolism in language and its proposed importance for early language acquisition, the long-standing question remains open as to how arbitrary language actually is. Are the observed systematic clusters, such as phonoaesthemes, merely a “negligible fraction” of the lexicon<sup>25</sup>, or is systematicity a more substantial feature of spoken language? This is an important question to address because it provides insight not only into the properties of the vocabulary that support acquisition and processing, but also more generally into the manner in which mappings between representations are constructed in the brain. There is evidence that systematicity in mappings between sensory regions of the cortex may be more efficient<sup>26</sup>, consequently, there is potentially a balance to find between implementational constraints in the brain with potential advantages of arbitrariness for communicative efficiency. We return to this point in the Discussion.

To our knowledge, there are three previous published studies that have developed a measure of the properties of sound meaning mappings present in natural language. Tamariz<sup>27</sup> investigated subsamples of Spanish vocabulary, relating distances in sound space to distances in meaning space, where meanings were derived from the contextual occurrence of words<sup>28</sup>. For



carefully selected subsets of Spanish words, she demonstrated that the relationship between sound and meaning contained a small degree of systematicity, particularly in the relationship between consonants and categories of meaning. Otis and Sagi<sup>29</sup> examined the relationships between sets of letters and meaning for phonaesthemes, where meanings were derived from Infomap<sup>30</sup>, a variant of latent semantic analysis<sup>31</sup>. They focused on sets of phonaesthemes proposed in the literature<sup>32</sup>, which formed statistically significant clusters of related meanings. They found that, of 46 phonaesthemes proposed by Hutchins<sup>32</sup> as present in the English language, 27 were statistically significant clusters, including *sn-* and *gl-*. Third, a study<sup>33</sup> of a small sample of the most frequent monomorphemic words of English resulted in an estimate of sound symbolism and found results consistent with those of Tamariz<sup>27</sup>.

However, there has as yet been no comprehensive analysis of the relationships between form and meaning for a large-scale representative vocabulary. The first aim of this study was to determine the properties of the form-meaning mapping for a broad and representative set of words in English. Previous studies have focused on a single measure of sound and of meaning, and have assessed only subsamples of the vocabulary. We sampled all the monosyllabic words in English for the analyses. Monosyllabic words constitute 70.9% of all word uses in English<sup>34</sup>, and so confining analyses to just these words is a reasonable approximation to the whole vocabulary. To ensure that the limitation to monosyllabic words did not adversely affect the results, we also gathered a corpus of all monomorphemic words of all lengths (we refer to this in the following as polysyllabic). However, we assume that language processing and language acquisition are influenced by the frequency with which words occur in the linguistic environment, and so caution must be taken to ensure that the many long multisyllabic words that occur very rarely in language<sup>35</sup> do not skew the results towards a non-representative subsample of the vocabulary.

Furthermore, the current study examines the robustness of the observed sound-meaning mapping to different representations of sound and meaning, to ensure that estimates of systematicity or arbitrariness of the vocabulary are not prone to a particular interpretation of sound or meaning similarity.

The second aim of the study was to examine the contribution of individual words to the overall system of form-meaning mappings. This enables us to determine whether the relationship between form and meaning in the vocabulary is due to small clusters of words that are related or unrelated across form and meaning representations, or whether the properties of the mapping are generalizable across the whole vocabulary. Furthermore, it means that the relationship between an individual word's systematicity and its psycholinguistic properties can also be measured. In particular, we related systematicity at the word level to the age at which a word is learned. If sound symbolism is critical for language acquisition, then we would expect to see enhanced systematicity for the words that children first acquire.

## **Method**

### **Materials**

**Corpus preparation.** We took all the English monosyllabic words from the CELEX database<sup>32</sup>. We also extracted all the monomorphemic words from the CELEX database in preparation for the polysyllabic analyses. To ensure that the measure of sound-meaning systematicity in the vocabulary was not due to the particular representation of sound or meaning, we computed several measures of sound and meaning similarity.

For sound similarity, we tested the three alternative approaches following Monaghan, Christiansen, Farmer, and Fitneva<sup>36</sup>. Testing multiple sound measures is important in order to ensure that apparent relationships between sound and meaning are not due to particular types of representation of sound similarity. First, each phoneme in the word was converted to a phonological feature representation<sup>37</sup>, and then the sound similarity between each pair of words was determined to be the minimum number of phonological features changes required to convert one word to the other (phoneme feature edit distance). This measure closely corresponds to psycholinguistic measures of sound similarity<sup>38,39</sup>. The second sound similarity measure was optimal string alignment Damerau-Levenshtein distance over phonemes<sup>40</sup>, where sound similarity is the number of phoneme changes required to convert one word to the other (phoneme edit distance). The third measure was the Euclidean distance between phonological feature representations of words (phoneme feature Euclidean distance). In the results, we first report similarity based on the phoneme feature edit distance, before indicating whether the effects are robust to different sound similarity measures.

For meaning similarity, we constructed two representations of meaning. The first was based on contextual co-occurrence vectors<sup>28</sup> which were generated by counting words appearing within a +/-3 word window with each of 446 context words in the British National Corpus<sup>41</sup>. Words with similar meaning tend to have similar usage, which is in turn reflected in terms of similar co-occurrence vectors<sup>30,31</sup>. As with the sound distance measures, an additional measure of meaning was used. This was in order to ensure that relationships between sound and meaning did not depend on a particular choice of one of the representations. For instance, it could be the case that words used in similar contexts tend to have similar (or distinct) sounds because some processing constraint on production encourages (or prohibits) similar sounding words occurring

close together in utterances. Hence, the second meaning representation was based on semantic features derived from WordNet, which reflected groupings of words according to hierarchical relations and grammatical properties<sup>37</sup>. Both types of meaning representation reflect behavioural responses to semantic similarity between words as measured through free associations and semantic priming studies<sup>28</sup>, though to varying degrees<sup>42</sup>. For each type of meaning representation, meaning similarity was 1-cosine distance between the representations for each word pair, such that small distances indicate similar meanings. In the results, we first report the meaning similarity measure based on contextual co-occurrence vectors. The semantic feature representation was not available for the monomorphemic polysyllabic words because it was derived only for monosyllabic words.

There were 5138 monosyllabic words with both cooccurrence and feature-based semantic representations. However, this vocabulary set contained both simple and complex morphological forms; inflectional and derivational morphology both express systematic sound to grammatical category relations that reflect semantic aspects of words<sup>43</sup>. In order to remove the contribution of morphology to the systematicity of the vocabulary, we derived the subset of word lemmas, which omitted morphologically inflected forms (e.g., *cat* but not *cats* was included),  $n = 3203$ , and also monomorphemes (e.g., *warm* but not *warmth* was included),  $n = 2572$ , which omitted all complex morphological and compound forms, based on CELEX classifications. The polysyllabic monomorphemic set of words, with contextual co-occurrence vectors, comprised 5604 words.

One potential source of sound-meaning systematicity in the vocabulary is due to etymology; word variants with the same historical meaning may consequently have similar phonological forms<sup>44</sup>. For instance, for the phonaestheme *gl-*, *glass*, *gleam*, *glitter*, *glisten*, and *glow*, are all proposed to derive from Proto-Indo-European root *\*ghel-*, meaning “to shine,

glitter, glow, be warm.”<sup>45</sup> Less distantly, gleam, glimmer, and glimpse are proposed to derive from the Old English root *\*glim-*, meaning “to glow, shine.”<sup>46</sup> We consulted etymological entries<sup>45,46</sup> for each of the monosyllabic monomorphemic words. Words with proposed common roots in one or more of Old English, Old French, Old Norse, Greek, Latin, Proto-Germanic, or Proto-Indo-European were omitted. There were 2572 monomorphemic words with etymology entries, of which 1732 words had no listed common origins, which were assessed to determine systematicity of the vocabulary independent of proposed common origins of words.

**Psycholinguistic properties.** For each monomorphemic word, we determined the age at which words are acquired by consulting age of acquisition ratings from Kuperman, Stadthagen-Gonzalez, and Brysbaert<sup>47</sup>. In order to assess the role of age of acquisition it is important to control for a set of other psycholinguistic variables, which may be correlated with age of acquisition. We generated measures of log-frequency, orthographic similarity (neighbourhood size, based on Coltheart’s  $N^{40}$ ), and word length from CELEX<sup>34</sup>. A word’s neighbourhood is defined as the number of other words in the vocabulary that are generated by changing one letter of the target word, and is a predictor of speed and accuracy of word retrieval<sup>48</sup>. All psycholinguistic variables were available for 2787 words.

## **Procedure**

**Testing form-meaning mappings.** To test the relationship between sound-meaning mappings, for every word pair measures of sound and meaning similarity were computed, resulting in  $(5138 \times 5137)/2$  distinct pairs of distances. To determine the relationship between sound and meaning for the entire set of words, the correlation between these pairs was measured. Note that this calculation assesses the relative iconicity of words. Figure 1 illustrates the cross-correlation

between distances within the sound space and within the meaning space. A positive value indicates that distances in sound space are related to distances in meaning space, values close to zero indicate that distances in sound and meaning are not related, i.e., arbitrary. In order to determine whether the correlation between sound and meaning is significant, we applied the Mantel test<sup>49</sup>, where every words' meaning was randomly reassigned, then the correlations between sound and randomised meaning was computed, with 10,000 random reassignments of words' meanings. The position in this distribution of the correlation resulting from the sound meaning mappings in the actual language against the correlations from random reassignments, in a Monte-Carlo test, indicates the significance of the systematicity or arbitrariness of the vocabulary.

Mantel tests were conducted for each of the sound and meaning distance measures, for all words, word lemmas, monomorphemes, and monomorphemic words with no common etymology.

**Testing arbitrariness of individual words.** In order to determine the contribution of each word to the overall systematicity or arbitrariness of the language, we computed each word's individual systematicity. For each target word, it was omitted from the set of pairs of sound and meaning distances, and the correlation of the vocabulary with this word omitted was then reassessed. The size and direction of change in the new correlation against the original correlation including the target word was then recorded. Positive values indicate the omitted word contributed to systematicity of the vocabulary, negative values demonstrate that the word was arbitrary in terms of its sound-meaning mapping.

## Results

## Correlations between sound and meaning

The results for the Mantel tests of sound-meaning mappings for the phoneme feature edit distance measure for sound and the contextual co-occurrence measure for meaning are shown in Figure 2. For all words, we found that the English vocabulary was more systematic than expected by chance,  $p < .0001$ , though the amount of variance explained was very small ( $r^2 < .002$ ). For the word lemmas – that is, considering words with no derivational or inflectional morphology – the results were similar,  $p < .0001$ . Analysing the monomorphemic word set, i.e., removing all morphology from the words, again did not change the results – word roots were more systematic than expected by chance,  $p < .0001$ . Finally, for the analyses of words with no common etymology, the results again supported systematicity in sound-meaning mappings,  $p = .0002$ : only one of the randomized rearrangements of meaning distances resulted in a higher correlation than the actual word set.

We next tested the various combinations of sound distance and meaning distance measures to ensure that the results were generalizable across these different ways to determine similarity. The results for each word set are shown in Table 1. The results were similar: for the co-occurrence semantic similarity and each phonological similarity measure, there was greater than chance systematicity, explaining small amounts of variance in the vocabulary. For the semantic feature similarity measure, the results were again similar for all phonological similarity measures and for all word sets, with the exception of the words with no common etymology, where the relationship was found to be marginally significant.

In order to determine the generality of the effects to polysyllabic words, we repeated the analyses on the 5604 monomorphemic polysyllabic word set. The results supported the original monosyllabic analyses: For the phoneme feature edit measure,  $r = .009$ ,  $p = .005$ , for the

phoneme edit measure,  $r = .016$ ,  $p = .0160$ , and for the Euclidean distance measure,  $r = 0.012$ ,  $p = .0018$ .

### **Arbitrariness of individual words**

In order to assess the distribution of systematicity across the vocabulary, we measured the systematicity of individual words in the language by determining whether omitting each word increased or decreased the correlation between sound and meaning for the whole vocabulary. The landscape of systematicity and arbitrariness of individual words is shown in Figure 3, which shows the systematicity of the sound space of words. The plot projects the relative position of monomorphemic words according to their sound similarity onto a two-dimensional plane using multiple dimensional scaling, with the systematicity of each word on the vertical axis, and the landscape was then smoothed using linear interpolation. As illustrated, the vocabulary indicates both peaks of sound-symbolism as well as troughs of arbitrariness.

In order to determine the properties of this landscape, we examined whether the overall systematicity of the vocabulary is driven by small pockets of sound symbolism, or whether it is a general characteristic of the entire set of words. If the systematicity of the vocabulary is confined to, and driven by, a small set of clusters – illustrated in the peaks of Figure 3 – then the distribution of systematicity should exhibit divergence from the distribution of individual words' systematicity when words' meanings are randomly reassigned, as in the randomisation for the Mantel test in the previous section. Alternatively, if systematicity is due to the distribution across the whole vocabulary, then the distribution should not diverge from a randomised distribution. Note that any distribution of systematicity across the whole vocabulary would result in peaks and troughs, but the issue is whether these peaks and troughs differ from that expected from the general distribution.



We assessed the distribution of peaks and troughs across the space by comparing the distribution of systematicity to 1,000 distributions resulting from randomly reassigning the meaning representations of words and determining the systematicity of each word following this randomisation. If systematicity of the whole vocabulary is a consequence of a few small pockets of sound symbolism then the actual distribution of systematic words should be significantly different than the distribution resulting from randomised distributions. Figure 4 shows the probability density function distribution of the systematicity values for the set of monomorphemic words, indicating that it lies within the range of the set of randomised distributions. We conducted Wilcoxon signed rank tests comparing the distribution of systematicity of actual words against each of the randomisations. None were significantly different than chance with Bonferroni correction, minimum  $p = .2$ . Thus, the apparent peaks (and troughs) of sound symbolism in the vocabulary are anticipated from the distribution of systematicity across the whole vocabulary. Therefore the observed systematicity of the vocabulary is not a consequence only of small pockets of sound symbolism, but is rather a feature of the mappings from sound to meaning across the vocabulary as a whole.

Finally, we determined whether systematicity is differently expressed in the vocabulary across stages of language development. If sound symbolism is critical for language acquisition<sup>8,50</sup>, then we would predict greater systematicity for words that are implicated in early language acquisition than those related to later language use. We related each individual monomorphemic word's systematicity to the estimated age that it was acquired, controlling for other psycholinguistic variables<sup>47</sup> using multiple linear regression. For these other psycholinguistic variables, there was no significant effect of log-frequency,  $\beta = -.046$ ,  $t = -1.864$ ,  $p = .063$ , orthographic length,  $\beta = -.025$ ,  $t = .872$ ,  $p = .383$ , or phonological length,  $\beta = .003$ ,  $t =$

.122,  $p = .903$ , and there was a small effect of orthographic similarity,  $\beta = .054$ ,  $t = 2.081$ ,  $p = .038$ . Critically, for age of acquisition, we found that early acquired words contributed more to systematicity than late acquired words,  $\beta = -.075$ ,  $t(2781) = -3.022$ ,  $p = .003$ . Figure 5 illustrates the mean systematicity for words binned into age of acquisition years from age 2 to 13 and older (note that words are not reliably judged to be acquired before 2 years old). The significant effect in the regression analysis is due to sound symbolism being more available during early stages of language acquisition, whereas arbitrariness is dominant within the developed adult vocabulary. The effect of age of acquisition relating to systematicity of words was robust over analyses using all words, word lemmas, words with no common etymology, and applying the different measures of sound and meaning similarity.

One possible driver of the age of acquisition results is the different distribution of nouns and verbs at different stages of language acquisition – a large proportion of early acquired words are nouns. If nouns are more systematic than verbs then part of speech may be the source of the age of acquisition effect rather than systematicity being an inherent and independent property of early acquired words generally. In order to control for this we determined for the monosyllabic monomorphemes whether the word was a noun or a verb (in terms of most frequent usage in CELEX), if the word was from another category we omitted it from the analyses. This resulted in 2252 nouns and 329 verbs. Whether the word was a noun or verb was entered as an additional predictor variable into the regression analysis. This resulted in a significant effect of phoneme length,  $\beta = .057$ ,  $t = 2.261$ ,  $p = .024$ , a significant effect of orthographic similarity,  $\beta = .056$ ,  $t = 2.035$ ,  $p = .042$ , and a significant effect of age of acquisition,  $\beta = -.086$ ,  $t = -3.30$ ,  $p = .001$ . No other variables were significant. This indicates that the age of acquisition effects are robust and not due to effects of grammatical category.

The advantage of considering all words simultaneously is that they can be assessed against the same distribution of form-meaning mappings and thus can be directly compared for the arbitrariness or systematicity present in vocabulary at different stages of language acquisition. However, using this method the systematicity measure for the early acquired words is determined by comparison to the whole vocabulary. To establish whether systematicity is present in the early acquired words only for those words that children acquire first, we measured the sound-meaning mapping amongst the 300 monomorphemic monosyllabic words that children acquire up to the age of 4 years old. For co-occurrence vector semantic similarity and phoneme feature edit distance similarity (other similarity measures result in similar effects), the mapping was systematic,  $r = .045$ ,  $p = .0442$ .

## Discussion

We have shown that the sound-meaning mapping is not entirely arbitrary, but that systematicity is more pronounced in early language acquisition than in later vocabulary development. This seems to conflict with the “design feature” and Saussurian view of the arbitrariness of the sign<sup>1,2</sup>, the dominant view throughout the last century of language science, which contends that form-meaning mappings are arbitrary. Some systematicity may be anticipated from the morphological structure of the vocabulary – we know that derivational and inflectional morphology carries information about words’ usage and can indicate certain features of meaning<sup>43</sup>, such as the distinction between nouns and verbs, or the tense of the action being described, or the relationship between the length of morpheme and the quantity implied by comparatives and superlatives (e.g., *long*, *longer*, *longest*)<sup>51</sup>. However, even for the monomorphemic words, when morphology was not exerting an influence on the sound-meaning

mappings, the vocabulary is more systematic than expected by chance. Furthermore, we have demonstrated that the observed systematicity is also not due to common historical roots for words. For monomorphemic words with no shared etymological origin there is greater systematicity than expected by chance.

The analyses of the landscape of the form-meaning mappings demonstrated that systematicity in the vocabulary is not a consequence of small clusters of sound symbolism, rather, it is a general property of the whole language. Systematicity, then, is not a consequence of small exceptional clusters of form-meaning correlation, which could indicate that the structure of the vocabulary is affected or has been altered by specific isolated features of sound relating to meaning. Instead, the general property of systematicity indicates that the vocabulary is more likely to be configured by principles that apply across the whole language.

Crucially, the presence of systematicity of form-meaning mappings varies across the vocabulary. For words that feature early in language acquisition, systematicity is prominent, but for later acquired words, the form-meaning mappings reveal increasing arbitrariness. The enhanced systematicity for the early vocabulary supports views that systematicity is useful for language acquisition<sup>15,16,17,18,52</sup>. Systematicity promotes understanding of the communicative function of language early in development, as the form provides information to the learner about the meaning, potentially enabling the child to learn that words have referents. The corpus analyses we have conducted are entirely consistent with views that sound symbolism may be necessary for bootstrapping word learning. The greater systematicity for early acquired words is also consistent with studies that have demonstrated that, under certain conditions, and for small sets of words, sound symbolism facilitates identification of the actual referent associated with the

spoken word<sup>6,8</sup>, and also studies of form-meaning mappings in sign languages, where iconicity improves acquisition<sup>53</sup>.

Systematicity has been suggested to lie at the origins of language. Ramachandran and Hubbard<sup>7</sup> proposed that non-arbitrary preferences across modalities – such as between visual appearances of objects and certain sounds or shapes of the mouth (as in the example of the sounds *bouba* and *kiki* relating to rounded and angular objects, respectively) – became conventionalised in human communication. Though any one cross-modal preference may have been too weak to propagate a proto-language, multiple cross-modal correspondences could have interacted to create a system where spoken sounds communicated the intended referent (see also Cuskley and Kirby<sup>44</sup> for discussion of cross-modal processes and language evolution).

The systematicity of early language also accords with the ontogenesis of topographic maps in the neocortex<sup>26</sup>, where similar stimuli are encoded in close cortical space<sup>54</sup>. Computational models of cortical topography demonstrates that it is more efficient to encode cross-modal correspondences that exploit the topography within each modality<sup>55,56</sup>. Representations that activate regions that are close together in one sensory cortical area can be mapped onto close regions in another sensory cortex with less white matter than mappings that do not reflect areal topography. Hence, there are pressures within the neural substrate toward forging systematic mappings between modalities. It may be that this mechanism for systematicity accounts for how sound symbolism may come to be expressed in language – if encouraged to generate a novel word for a concept, a similar sounding word would respect cross-modal constraints. Similarly, the same mechanism may well explain how systematicity can initially promote learning mappings between sound and meaning, as is observed in words that occur in early language acquisition<sup>8</sup>.

Yet, systematicity comes at a cost in terms of efficiency of information transmission<sup>5</sup>, because it reduces the distinctiveness available within the sounds of words used to refer to similar sensory experiences. This apparent tension appears to be addressed within the vocabulary by reducing systematicity as the vocabulary increases – for words acquired between ages 2 and 6 the vocabulary is systematic, after this age, the vocabulary is more arbitrary, with most arbitrariness observed for words acquired at age 13 and older. For the child with a small vocabulary, ensuring distinctiveness amongst a smaller set of words is less critical because the distribution of the set of words entails that distinctions in meaning are likely to be greater. In the contextual co-occurrence vectors, this difference is evident. For words acquired age up to 3 years of age, mean cosine distance between meaning vectors is .224 (SD = .099), whereas the distance between vectors for words acquired from age 3 upwards is .116 (SD = .071), which is significantly different,  $t = 44.996$ ,  $p < .001$ . This has the consequence that systematicity in form-meaning mappings can be tolerated because fine discriminations between the meanings of words do not have to be discerned from only on the phonological form of the word. This result is not a trivial consequence of comparing a smaller and a larger vocabulary, because it could have been the case that earlier acquired words densely occupied a smaller region of the possible meaning space<sup>57</sup>, in which case meaning distinctiveness would not differentiate first acquired words compared to the entire vocabulary.

The increased arbitrariness for later acquired words assists the mature language user in determining nuanced distinctions in meaning, as arbitrariness maximises the information available in the communicative discourse<sup>11,13</sup>, especially important when distinctions between meanings, in terms of contextual information, are less available. This arbitrariness of the later acquired words is also important in establishing that the results are not just due to increasing

levels of noise in the semantic representations for later, more complex, potentially lower frequency words. If the later acquired words effects were merely due to increasing noise then the systematicity of the words would decline to chance level, whereas in fact the systematicity polarises to below chance level, thus indicating that these representations are carrying important information complementary to the early acquired words.

The results are also consistent with a number of other observations about the relationship between meaning and communicative distinctiveness. When nuanced distinctions are not so critical, as is the case for certain circumscribed sets of words in the vocabulary, such as expressives<sup>9</sup> (where identifying the difference between, for instance, *gigantic* and *ginormous* is not absolutely essential for communicative effectiveness), then systematicity appears to be more tolerated in the language. Consequently, expressives seem to be one of the very few language universal properties where systematicity is observed<sup>14,58</sup>. Relatedly, and in addition to the systematicity observed in monomorphemic words, morphology provides an additional source of systematicity in form-meaning mappings. Thus, ending in *-ed*, such as for *mapped* or *learned*, is a strong indicator that the word is a verb and that it refers to an event that is past, whereas ending in *-er* is a strong indicator of a noun (as in *mapper*, *learner*). This systematicity is likely to be advantageous because it provides information about the general category of the word, rather than at the level of the individual word<sup>43,59</sup>. For mapping from form onto such category levels, systematicity in the spoken word is beneficial<sup>21,60</sup>, but for the more specific task of individuating words' meanings, arbitrariness is advantageous, at least for larger vocabularies<sup>13</sup>. For both categorisation and individuation, division of labour within the structure of the word may be beneficial<sup>13,27,61</sup>.

The greater systematicity for early acquired words is consistent with computational models that demonstrate that pressures from vocabulary size prohibit systematicity. Gasser reported that the arbitrariness advantage for mappings between form and meaning was only observed in his computational model when the vocabulary exceeded a certain size, where the precise size was dependent on the distinctiveness available in the signal<sup>11</sup>. Thus, whilst the vocabulary is small, as in early stages of acquisition, there is no pressure against systematicity in the mappings. Only when the vocabulary is larger is arbitrariness required for efficient learning.

In spoken language, the issue of distinctiveness is closely related to arbitrariness, because the dimensions available to create variation in the signal are limited to sequences of sounds, expressed in segmental and prosodic phonology<sup>62</sup>. Hence, it is not possible to ensure that words with similar meanings have distinct sounds without simultaneously introducing divergence between form and meaning. In spoken language, there is thus a conflation between absolute and relative iconicity. However, in sign languages, distinctiveness can be distinguished from arbitrariness due to several properties. First, in sign languages the number of dimensions available to form distinctiveness may potentially be much greater<sup>63</sup>, and the aspect of the sign that can relate to meaning for each word can vary accordingly. British Sign Language, for instance, expresses signs (at least) in terms of initial hand shapes and positions, hand shape changes, hand movements, as well as facial expression. Second, the sign can relate to various visual properties of the referent, using any of the phonological features of the sign (hand shape/position, hand shape changes/movements). In contrast, in spoken language, iconic relationships can only occur between the sound of the word and sound properties of the referent – a much more restricted set. Third, the dimensions in the sign can be expressed simultaneously, meaning that the dimensions can add to distinctiveness because they are processed in parallel. In



contrast, in spoken language, the sequential nature of speech production and processing requires that distinctiveness be available early in the word, again meaning that systematicity in spoken language would result in a greater reduction in distinctiveness than in sign language. Thus, absolute iconicity between form and meaning can be accommodated in sign language without compromising distinctiveness, and potentially also without also introducing relative iconicity in the mappings, because independent aspects of the signal can be varied to maximise distinctiveness but also to permit iconic relationships between sign and meaning.

Consider, for instance, the sign for cat and the sign for dog in British Sign Language and in spoken English. The sign for cat has initial hand configuration as open with fingers apart and slightly bent, with starting position of the hands at each side of the face, and then short movement of the hands outwards. The sign for dog has initial hand configured as index and middle fingers of each hand extended and pointing downwards. Starting hand position is in front of the body, and then short movements up and down. *Dog* and *cat* have some similarities in terms of meaning – they could occur in similar contexts in discussions about household pets – yet the signs are distinct in each of the expressed dimensions, with different salient features of each animal iconically represented in the sign – for the cat it is the appearance of the whiskers, for the dog it is its behaviour reminiscent of begging. In contrast, in spoken English, the distinction is expressed in terms of different consonants and vowels, none of which are transparently iconically related to the animal. Reflecting properties of the referent in spoken forms of these words would necessitate reducing the distinctiveness in the sounds of the words – sound similarity can only be accomplished by changing the same signal dimensions as is used to ensure distinctiveness. Increasing the dimensions by which signs can be distinguished means that arbitrariness would not be required until a substantially larger vocabulary is required. Such

general principles are consistent with observations that speakers maintain a steady rate of information when communicating, where the interplay between the word's context and the sound of the word itself remains stable<sup>64,65</sup>.

### **Conclusion**

Over 2300 years ago, Plato reported the dialogue between Hermogenes and Cratylus over whether the nature of a word resides in its form, or whether the word is arbitrarily related to its meaning<sup>5</sup>. This debate can now be resolved with a classic dialectic synthesis: they are both right, but for different regions of the vocabulary. The structure of the vocabulary serves both to promote language acquisition through sound symbolism<sup>6,7,8</sup> as well as to facilitate efficient communication in later language through arbitrariness maximising the information present in the speech signal<sup>9,13</sup>.

### **Acknowledgements**

We thank Dale Barr, Simon Garrod, Jim Hurford, Bob Ladd, Pamela Perniss, Mark Seidenberg, and Gabriella Vigliocco for helpful comments on an earlier draft of this paper.

## References

1. F. de Saussure, *Course in General Linguistics*. (McGraw-Hill, New York, 1916).
2. C. F. Hockett, The origins of speech. *Sci. Am.* **203**, 89-96 (1960).
3. P. Perniss, R. L. Thompson, G. Vigliocco, Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology* **1** (2010).
4. U. Eco, *The search for the perfect language*. (Blackwell, London, 1995).
5. Plato, *The Collected Dialogues*, E. Hamilton, H. Cairns, Eds. (Princeton Univ. Press, Princeton, 1961), pp. 421-476.
6. B. K. Bergen, The psychological reality of phonaesthemes. *Language* **80**, 290-311 (2004).
7. V. S. Ramachandran, E.M. Hubbard, Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies* **8**, 3–34 (2001).
8. F. Spector, D. Maurer, Synesthesia: A new approach to understanding the development of perception. *Dev. Psychol.* **45**, 175-189 (2009).
9. M. Gasser, N. Sethuraman, S. Hockema, in *Empirical and experimental methods in cognitive/functional research*, J. Newman, S. Rice, Eds (CSLI Publications, Stanford, 2010), pp. 163-180.
10. J. Wilkins, *An essay towards a real character and a philosophical language* (Sa. Gellibrand, London, 1668).
11. M. Gasser, The origins of arbitrariness of language. In *Proceedings of the 26<sup>th</sup> Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum (2004), pp.434-439.
12. G. W. Cottrell, P. Munro, Principal components analysis of images via back propagation.

Proc. SPIE 1001, Visual Communications and Image Processing '88: Third in a Series, 1070 (1988).

13. P. Monaghan, M. H. Christiansen, S. A. Fitneva,, The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *J Exp Psychol Gen.* **140**, 325-347 (2011).
14. L. Hinton, J. Nichols, J. Ohala, (Eds.), *Sound symbolism* (Cambridge University Press, Cambridge, 1994).
15. K. Kantartzis, S. Kita, M. Imai, Japanese sound-symbolism facilitates word learning in English-speaking children. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum (2009), pp. 591– 595.
16. M. Imai, S. Kita, M. Nagumo, H. Okada, Sound symbolism facilitates early verb learning. *Cognition* **109**, 54-65 (2008).
17. M. Imai, S. Kita (placeholder for this issue).
18. V. Kovic, K. Plunkett, G. Westermann, The shape of words in the brain. *Cognition* **114**, 19-28 (2010).
19. A. Nielsen, D. Rendall, The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition* **4**, 115-125 (2012).
20. L. C. Nygaard, A. E. Cook, L. L. Namy, Sound to meaning correspondences facilitate word learning. *Cognition* **112**, 181–186 (2009).
21. P. Monaghan, K. Mattock, P. Walker, The role of sound symbolism in word learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 1152-1164 (2012).
22. O. Ozturk, M. Krehm, A. Vouloumanos. Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *J. Exp. Child Psychol.* **114**, 173-186

(2013).

23. P. Walker, J. G. Bremner, U. Mason, J. Spring, K. Mattock, A. Slater, S. Johnson, Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychol. Sci.* **21**, 21-25 (2010).
24. M. Peña, J. Mehler, M. Nespors, The role of audiovisual processing in early conceptual development. *Psychol. Sci.* **22**, 1419-1421 (2011).
25. J. Ohala, The frequency code underlies the sound-symbolic use of voice pitch. In *Sound Symbolism*, L. Hinton, J. Nichols, & J. J. Ohala, Eds, (Cambridge University Press, Cambridge, 1994), pp.325–347.
26. J. H. Kaas. Topographic maps are fundamental to sensory processing. *Brain Res Bull.* **44**, 107–112 (1997). NEW 26.
27. M. Tamariz, Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *Mental Lexicon* **3**, 259-278 (2005).
28. F. Huettig, P. Quinlan, S. McDonald, G.T.M. Altmann, Word co-occurrence statistics predict language-mediated eye movements in the visual world. *Acta Psychol.* **121**, 65-80 (2006).
29. K. Otis, E. Sagi, in *Proceedings of the 30<sup>th</sup> Annual Meeting of the Cognitive Science Society*, B. C. Love, K. McRae, V.M. Sloutsky, Eds (Cognitive Science Society, Austin TX, 2008).
30. J. Schütze, *Ambiguity Resolution in Language Learning*. (CSLI Publications, Stanford CA, 1997).
31. T. K. Landauer, S. T. Dumais, A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychol. Rev.* **104**, 211-240 (1997).
32. S. S. Hutchins, The psychological reality, variability, and compositionality of English

- phonesthemes. *Dissertation Abstracts International*, 59(08), 4500B. (University Microfilms No. AAT 9901857) (1997).
33. R. C. Shillcock, S. Kirby, S. McDonald, C. Brew. Filled pauses and their status in the mental lexicon. *Proceedings of the 2001 conference of disfluency in spontaneous speech*, (Edinburgh: International Speech Communication Association), 53–56 (2001).
34. R. H. Baayen, R. Pipenbrock, L. Gulikers *The CELEX Lexical Database*, (Univ. Pennsylvania, Philadelphia, 1995).
35. G Zipf, *Human behavior and the principle of least effort*. (Addison-Wesley, New York, 1949).
36. P. Monaghan, M.H. Christiansen, T.A. Farmer, S.A. Fitneva, Measures of phonological typicality: Robust coherence and psychological validity. *Mental Lexicon* **5**, 281-299 (2010).
37. M.W. Harm, M.S. Seidenberg, Computing the meaning of words in reading: Cooperative division of labor between visual and phonological processes. *Psychol. Rev.* **111**, 662–720 (2004).
38. G. A. Miller, P. E. Nicely, An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* **27**, 338–352 (1955).
39. T. M. Bailey, U. Hahn, Phoneme similarity and confusability. *J Mem Lang.* **52**, 347-370 (2005).
40. T. Yarkoni, D. Balota, M. Yap, Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychon. Bull. Rev.* **15**, 971–979 (2008).
41. Burnard, Lou, ed. *British National Corpus: Users Reference Guide British National Corpus Version 1.0*. (Oxford Univ., Oxford, 1995).

42. G. Vigliocco, D. P. Vinson, W. Lewis, M. F. Garrett. Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cog. Psych.* **48**, 422-488 (2004).
43. M. S. Seidenberg, L. M. Gonnerman, Explaining derivational morphology as the convergence of codes. *Trends Cogn Sci.* **4**, 353–361 (2000).
44. Cuskley, C. and Kirby, S. (Forthcoming). Synaesthesia, cross-modality and language evolution. *Oxford Handbook of Synaesthesia*, J. Simner, E. M. Hubbard (Eds).
45. E. Klein, *A Comprehensive Etymological Dictionary of the English Language: Dealing with the Origin of Words and their Sense Development thus Illustrating the History of Civilization and Culture*. (Elsevier, Amsterdam, 1966).
46. OED Online (Oxford University Press, Oxford, accessed July 05, 2013).
47. V. Kuperman, H. Stadthagen-Gonzalez, M. Brysbaert, Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods.* **44**, 978-990 (2012).
48. A. Pollatsek, R. Perea, K. S. Binder, The effects of “neighborhood size” in reading and lexical decision. *J. Exp. Psychol. Hum. Percept. Perf.* **24**, 1142-1158 (1999).
49. N. Mantel, The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
50. D. Maurer, T. Pathman, C. J. Mondloch, The shape of boubas: Sound–shape correspondences in toddlers and adults. *Dev. Sci.* **9**, 316–322 (2006).
51. M. Haspelmath, Frequency vs. iconicity in explaining grammatical asymmetries. *Cogn. Linguist.* **19**, 1-33 (2008).

52. L. C. Nygaard, A. E. Cook, L. L. Namy, Sound to meaning correspondences facilitate word learning. *Cognition* **112**, 181-186 (2009).
53. R. L. Thompson, D. P. Vinson, B. Woll, G. Vigliocco, The road to language learning is iconic: Evidence from British Sign Language. *Psychol. Sci.* **23**, 1443-1448 (2012).
54. M. E. Diamond, R. S. Petersen, J. A. Harris. Learning through maps: Functional significance of topographic organization in primary sensory cortex. *J Neurobiol.* **41**, 64–68 (1999).
55. V. Walsh. A theory of magnitude: Common cortical metrics of time, space and quality. *Trends Cogn. Sci.* **7**, 483–488 (2003).
56. C. Spence. Crossmodal correspondences: A tutorial review. *Atten. Percept. Psychophys.* **73**, 971–995 (2011).
57. M. Steyvers, J. B. Tenenbaum, The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cogn. Sci.* **29**, 41–78 (2005).
58. N. Evans, S. Levinson, The myth of language universals: Language diversity and its importance for cognitive science. *Brain Behav. Sci.* **32**, 429-492 (2009).
59. L. Onnis, M. H. Christiansen, Lexical categories at the edge of the word. *Cogn. Sci.*, **32**, 184-221 (2008).
60. T. A. Farmer, M. H. Christiansen, P. Monaghan, Phonological typicality influences on-line sentence comprehension. *Proc. Nat. Acad. Sci.* **103**, 12203–12208 (2006).
61. M. C. St. Clair, P. Monaghan, M. Ramscar, Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cogn. Sci.*, **33**, 1319-1329 (2009).



62. L. C. Nygaard, D. S. Herold, L. L. Namy, The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cogn. Sci.* **33**, 127-146 (2009).
63. W. Sandler, D. Lillo-Martin, *Sign language and linguistic universals* (Cambridge University Press, Cambridge, 2006).
64. T. Jaeger, Redundancy and reduction: Speakers manage syntactic information density. *Cogn. Psychol.* **61**, 23–62 (2010).
65. S. T. Piantadosi, H. Tily, E. Gibson, The communicative function of ambiguity in language. *Cognition* **122**, 280-291 (2012).

## Table

**Table 1.** Correlations between different implementations of measures of sound similarity and meaning similarity for each word set.

Word Set	Meaning Similarity Measure	Sound Similarity Measure	r	Mantel test, <i>p</i> =
Word Forms	Co-occurrence	Phoneme Feature Edit	.035	.0001
		Phoneme Edit	.035	.0001
		Euclidean Distance	.028	.0001
	Feature	Phoneme Feature Edit	.033	.0001
		Phoneme Edit	.016	.0001
		Euclidean Distance	.014	.0001
Word Lemmas	Co-occurrence	Phoneme Feature Edit	.031	.0001
		Phoneme Edit	.032	.0001
		Euclidean Distance	.028	.0001
	Feature	Phoneme Feature Edit	.015	.0001
		Phoneme Edit	.016	.0001
		Euclidean Distance	.012	.0033
Monomorphemes	Co-occurrence	Phoneme Feature Edit	.034	.0001
		Phoneme Edit	.036	.0001
		Euclidean Distance	.031	.0001
	Feature	Phoneme Feature Edit	.011	.0001
		Phoneme Edit	.011	.0001
		Euclidean Distance	.008	.0001
Etymology	Co-occurrence	Phoneme Feature Edit	.035	.0002
		Phoneme Edit	.044	.0001
		Euclidean Distance	.030	.0010
	Feature	Phoneme Feature Edit	.008	.0654

Phoneme Edit	.009	.0548
Euclidean Distance	.008	.0716

---

## Figure Legends

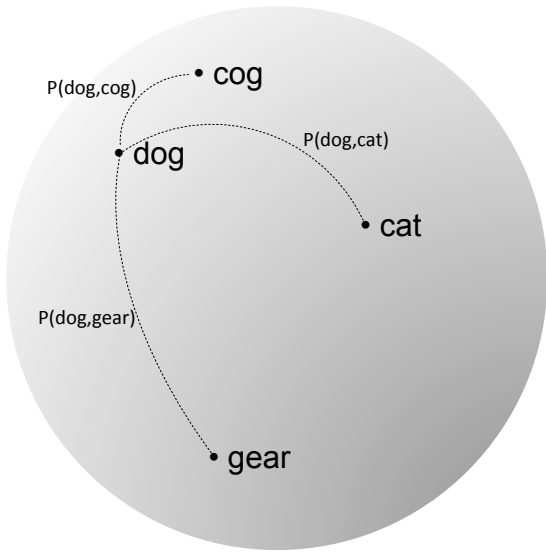
**Figure 1.** Example of correlating distances in the sound and meaning spaces. Points indicate sound and meaning representations of words,  $P(x,y)$  and  $S(x,y)$  indicate distances in sound and meaning spaces, respectively, between words  $x$  and  $y$ . Only 4 of the 5138 words are indicated, and distances only from the word “dog” are shown. Correlations are performed by pairing  $P(\text{dog,cat})$  with  $S(\text{dog,cat})$ ,  $P(\text{dog, gear})$  with  $S(\text{dog, gear})$ , etc.

**Figure 2.** Correlation between sound and meaning ( $\diamond$ ), against distribution of 10,000 randomised sound-meaning mappings, for all words, lemmas, monomorphemes, and words with no proposed common etymology. Distribution shows mean (-), first and third quartile (box) and range (bars) of randomised mappings. Positive values for correlation indicate a systematic relationship between sound and meaning.

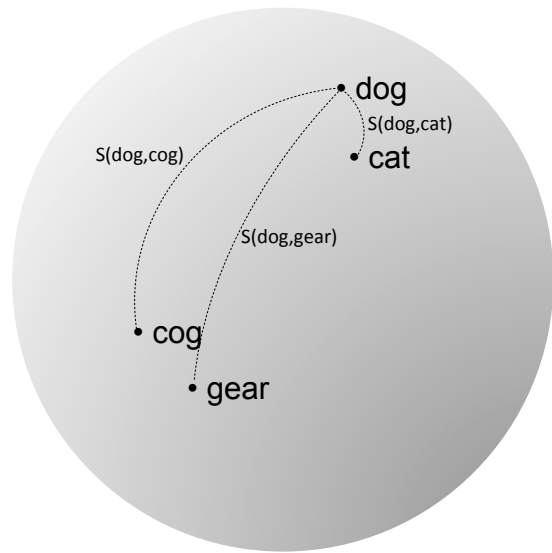
**Figure 3.** Systematicity and arbitrariness of the sound space for monomorphemic words. Positive peaks indicate “sound symbolic” regions, negative peaks indicate decorrelated, arbitrary sound-meaning regions of the vocabulary.

**Figure 4.** Probability density distribution of words’ systematicity (negative: word is arbitrary, positive: word is systematic) for actual vocabulary (solid line) and 1000 randomisations (grey shading) indicating that the landscape of peaks and troughs are consistent with general distribution of systematicity over the whole vocabulary rather than driven by peaks of systematicity.

**Figure 5.** Systematicity by age of acquisition. Mean and standard error for systematicity of words binned by each year of age of acquisition, indicating that words acquired early tend to be more systematic.



Phonological space



Semantic space

