The Semantic and Phonetic Distribution of Iconicity in English: A Computational Study

Jamie Wright

Abstract

Iconicity is a perceived similarity between aspects of the form of a word or sign and its meaning. This study makes use of iconicity ratings for English, derived from native speaker judgements, and embeddings based on distributional semantic and articulatory phonetic information to model the relationships between iconicity, semantics and phonetics for a vocabulary of over 14,000 English words. Linear regression models are employed to connect these elements, investigating the extent to which iconicity information is captured in English phonetic and semantic word embeddings. The models successfully predict iconicity ratings to a certain degree, suggesting the presence of iconicity information in these embeddings. Moreover, the models' predictions align with patterns found in previous research on the phonetic and semantic dimensions of iconicity. This research shines light on the relationship between iconicity, form and meaning, and contributes to our knowledge of the cognitive and linguistic mechanisms underlying the phenomenon of iconicity.

1. Introduction

The conventional notion in linguistics of the 'arbitrariness of the sign', that there is no connection between the form of a word and its meaning, was perhaps most famously stated in Saussure's (1972: 67-68) 'first principle', where he argues that there is no inherent connection "between the idea 'sister' and the French sequence of sounds s-ö-r which acts as its signal", and that this "is the organizing principle for the whole of linguistics". Nonetheless, there is evidence that non-arbitrariness is present in language, meaning that in some cases form does create, affect or modulate meaning, with the most common forms of non-arbitrariness being systematicity and iconicity (Dingemanse et al., 2015). This study focuses on iconicity, which refers to a relationship of similarity between the form and meaning of a given word and which can be found to varying degrees in natural language. One of the most prominent examples of iconicity in spoken language is onomatopoeia, where the sound of the word imitates the real-world sound it represents; for example, the form of *roar* reflects the deep cry it represents. However, iconicity in spoken language is not limited to this class of words, with ideophones being another example of iconicity. These are "depictive words that stand in an iconic relation to their real world referents" (Barnes, 2023: 89), making use of other aspects of the form of the word to depict aspects of meaning, as in the Japanese *kibikibi* ('energetic') (Dingemanse, 2018: 605-606). Indeed, rather than being a fringe remnant of non-arbitrariness in fundamentally arbitrary languages, it has increasingly been argued that iconicity is an essential component of both spoken and signed languages (Perniss, Thompson & Vigliocco, 2010; Ferrara & Hodge, 2018; Winter et al., 2023), and in particular that it plays an important role in language acquisition (Imai & Kita, 2014; Perry et al., 2018) and language evolution (Perniss & Vigliocco, 2014; Vinson et al., 2021).

Research on iconicity in spoken languages has shown that certain phonemes have iconic qualities in certain semantic contexts (see discussion below). As such, this study intends to further the research on iconicity by making use of natural language processing tools and native speaker iconicity ratings (from Winter et al., 2023) to discern the extent to which information about iconicity is contained in phonetic and semantic word embeddings. To this end, this study employs linear regression models using phonetic and semantic word embeddings as the explanatory variables and native speaker iconicity ratings as the response variable. We expect there to be a degree of iconicity information in both the semantic and phonetic embeddings, given the aforementioned relationship between certain phonemes, semantic contexts and iconicity. This study found that these models do indeed have a certain degree of success in predicting iconicity ratings, and that the trends in their predictions are consistent with other research on iconicity.

2. Background

2.1 The Distribution of Iconicity in Language

The existence of iconicity in language is predicated on the idea that the form of a linguistic sign can be linked in some way to its meaning, and as such the question of how form is linked to meaning is fundamental to the study of iconicity. The phonetic form of a word cannot wholly determine its meaning, the fact that foreign languages are uninterpretable to us is just one indicator of this, and thus iconicity must function in a different way. Ahlner & Zlatev (2010) make use of Peircian semiotics to explain the functioning of iconic signs. In this framework, a *representamen* (or sign) represents an *object* to an *interpretant* (that is, the person interpreting the sign). The idea which links the sign to the object is called the *ground* (see Figure 1). Peirce identified three 'ideal types' of signs, which are differentiated by their type of *ground* (Peirce, 1992). In an iconic sign, the nature of the ground is that of similarity: the sign and the object share some similar qualities. Indexical signs are based on relations in time and space, but are not relevant to the current discussion, while symbolic signs are based on a ground of convention, and thus are typical of the archetypal Saussurean 'arbitrary' sign. Ahlner and Zlatev (2010) point out that, in reality, signs often consist of facets of these three 'ideal types', rather than being a pure example of any of them. These three types of signs (arbitrary, indexical and iconic) correspond to what Ferrara and Hodge (2018) argue are the three methods of signalling that make up our use of language: describing, indicating and depicting. These three methods are often used in tandem, in what are termed 'multi-modal composite utterances', which are typical of both spoken and signed languages (Enfield, 2009). Iconicity has also been argued to play an especially important role in language acquisition and evolution, with the Sound Symbolism Bootstrapping Hypothesis making the claim that cross-modal iconic mappings aid communication in preverbal infants and other research showing that iconicity is more prevalent in the vocabulary of children than of adults (Imai & Kita, 2014; Perry et al., 2018).

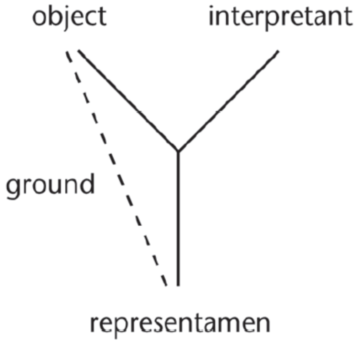

Figure 1: Illustration of Peircian semiotics (from Ahlner & Zlatev, 2010). The representamen (sign) is connected to the real word object by a ground. In the case of iconic signs, this ground is one of similarity.

Given this increased interest in the role of iconicity in language, a growing body of research addresses the extent and distribution of iconicity in language. One of the most well-known and researched cases of iconicity is the Bouba-Kiki effect, which demonstrates the linking of nonwords containing certain phonemes with round and spiky objects across languages and cultures (Ramachandran & Hubbard, 2001; Ćwiek et al., 2022). Research on ideophones, which make use of iconic depictions of their referents (Barnes, 2023), has shown that they form a major lexical class in a number of languages, such as Basque, Gbeya, Japanese, Korean, Turkish and Zulu (Dingemanse, 2018). With respect to iconicity in English, the language of study in this paper, Sidhu et al. (2021) showed that the phonemes associated with roundness/spikiness in the Bouba-Kiki effect were more common in English words referring to round/spiky objects, while Winter et al. (2017) used native speaker ratings to show that words with sensory meanings are more iconic (for example, *hissing* was judged to be highly iconic, as onomatopoeia often are, whereas abstract terms like *permission* were judged not iconic). Utilising computational tools and native speaker iconicity ratings (from Winter et al., 2023), this study extends previous research on iconicity in English to analyse the presence of iconicity information within phonetic and semantic word embeddings.

2.2 Distributional Semantics Models

In this study, a distributional semantics model (DSM) will be used to represent the semantic content of the words in the vocabulary. These models generate numerical vector representations for words, which are derived from an abstraction of occurrences of those words in context (Westera & Boleda, 2019). That is to say, DSMs contain distributional information about the words they represent, where words used in similar contexts, or with a similar distribution, have similar vector values. Their use is based on the Distributional Hypothesis, which asserts that similarity in meaning corresponds to similarity in linguistic distribution, an idea first popularised in the 1950s but which received renewed attention with the advent of computational linguistics tools (Firth, 1957; Boleda, 2020). DSMs generally represent words as vectors in a multi-dimensional space, within which geometric distance models semantic similarity, although they can vary in terms of the units being represented (words, morphemes, n-grams, sentences or documents), the training corpus used and the abstraction mechanism implemented (Boleda, 2020). The dimensions, or features, which make up these vector representations are abstract in nature: they do not make reference to any semantic information, although some approaches have managed to extract interpretable information from these features (Günther, Rinaldi & Marelli, 2019). This lack of interpretability is in contrast to the phonetic embeddings used in this paper, which consist of dimensions which refer to specific articulatory features (Mortensen et al., 2016).

A point of debate in the discussion around DSMs is whether they constitute a psychologically plausible model of human language. The Distributional Hypothesis is a cognitive hypothesis of meaning, which makes "cognitive assumptions about how these meanings are acquired – through repeated experience" (Günther, Rinaldi & Marelli, 2019: 7), and DSMs have been shown to replicate human behaviour on a number of semantic benchmarks (Baroni, Dinu, & Kruszewski, 2014). While this provides good support for DSMs as models of human language, the fact that they only make use of linguistic information (a word's semantic representation is derived from its context: other words) has been highlighted as theoretically problematic. A number of theories of cognition emphasise the role of sensorimotor experience in building our semantic representations, arguing that such representations are *grounded* in our senses (Barsalou, 2008). On this view, DSMs are ungrounded in nature, and are affected by the *symbol grounding problem*, in that they relate linguistic symbols to a distribution of other linguistic symbols, without directly accessing the sensorimotor information through which we interact with the world (Wingfield & Connell, 2022; Baroni, 2016). There is evidence that non-linguistic information can be implicit in DSMs; for example, Caliskan, Bryson and Narayanan (2017) proved that DSMs reflect human social and cultural biases drawn implicitly from text. Nonetheless, Bruni et al. (2012) showed that DSMs failed to relate concrete objects (e.g. *wood*) with their most obvious characteristic colour (*brown*), as the textual information upon which DSMs are built does not necessarily contain visual information, exemplifying the limitations of a language model which lacks sensorimotor input. These limitations seem to indicate that DSMs should not be taken as a model of human semantic knowledge in its entirety, but rather as "an essential component of semantics that is grounded in a complementary sensorimotor component" (Wingfield & Connell, 2022: 1222). The role of non-linguistic, sensorimotor information seems particularly pertinent when studying iconicity, as iconic signs rely on perceived similarity between the form of the sign and what is represented. If DSMs are able to differentiate words high in iconicity from those low in iconicity, it would imply that iconic words have a differentiable distribution: that they are used in certain contexts with certain meanings. It may also imply that the sensorimotor information which is required to make an iconic connection is somehow implicit in DSMs.

Previous research has shown that DSMs likely contain information relevant to different native speaker lexical ratings including iconicity (Thompson et al., 2020), systematicity (Monaghan et al., 2014) and valence (Passaro, Bondielli & Lenci, 2017). Thompson et al. (2020) make use of DSMs and regression models to predict iconicity ratings for words in English and American Sign Language (ASL), with a method similar to that used in this study, and find that these models more accurately predict iconicity ratings for English than ASL. This method was used in Dingemanse and Thompson's (2020) study on the relationship between funniness and iconicity to get 'imputed' iconicity ratings for English words not covered by the original dataset, and the validity of this method was confirmed by the similarity in correlation with funniness ratings between imputed and actual iconicity ratings. The findings of these papers indicate that DSMs do contain information relevant to iconicity, which in turn implies that highly iconic words have similar linguistic distribution and therefore are semantically related. This study builds on these findings by examining the performance of DSMs in iconicity models in more detail, comparing them with models based on articulatory phonetic embeddings, and combining them with the phonetic model.

2.3 Phonetic Vector Representations

While DSMs contain distributional information about how a word is used in context, the phonetic embeddings used in this paper break up words into segments and give information about the articulatory features of each, such as whether a vowel is sonorant or a consonant is voiced (Mortensen et al., 2016). These articulatory feature vectors allow for phonological similarity between segments to be calculated using vector feature edit distance (Mortensen et al., 2016). An important difference between the DSMs and phonetic embeddings used is that the features contained in the phonetic embeddings are directly interpretable. Phonetic vector sequences were used in De Varda and Strapparava's (2022) large-scale deep learning-based analysis of nonarbitrariness in language, in conjunction with DSMs and visual embeddings, to show that systematic form-meaning mappings can be transferred across languages. Monaghan et al. (2014) made use of the aforementioned phonetic vector similarity measure along with semantic vector similarity to show that systematicity is more present in English than would be expected by chance. While these studies focus on systematicity (form-meaning mappings which are not necessarily depictive in nature), this paper uses the same tools to extend the research to iconicity, through the use of native speaker iconicity ratings (from Winter et al., 2023). The primary objective of this paper therefore is to ascertain the degree to which information pertaining to iconicity is encapsulated within both phonetic and semantic word embeddings through the use of linear regression models. By leveraging these resources, the study intends to deepen our understanding of the relationship between iconicity, phonetics, and semantics.

3. Methods

3.1 Iconicity Ratings

This study makes use of the iconicity rating dataset from Winter et al. (2023). The dataset contains the results of an experiment where 1400 American English-speaking participants were given a definition of iconicity along with examples of iconic words, including *screech,* where the word sounds like the sound it represents, and *twirl* and *ooze*, which are examples of cross-modal iconicity, where there is a similarity between the the way the words sound and the movements they represent. The participants rated English words on a Likert scale from 1, not iconic at all, to 7, signifying a very iconic word. The dataset used in this study contains a total of 14,776 words along with their mean iconicity ratings. For the purposes of modelling, the iconicity ratings were rescaled to 0 to 1. See Table 1 for an overview of the words with the highest and lowest iconicity ratings.

Word	**Iconicity rating**
oomph	0.99
swish	0.99
wiggle	0.98
creak	0.97
clunk	0.97
…	
gnome	0.07
are	0.06
if	0.05
partial	0.05
how	0.05

Table 1: Words with highest and lowest ratings in the iconicity dataset from Winter et al. (2023), rescaled to 0 to 1.

3.2 Phonetic Embeddings

As a measure of the phonetic content of the words in the dataset, the Epitran-Panphon pipeline was used to obtain phonetic vectors for each token. Epitran (Mortensen et al., 2018) is used to convert tokens into phonemic representations using the International Phonetic Alphabet (IPA), while Panphon (Mortensen et al., 2016) converts the IPA representations into articulatory feature vectors per segment. The articulatory features are represented by vectors with 24 dimensions with values of -1, 0 and 1, with -1 and 1 referring to - and + for each feature, as in [±consonantal], while 0s occur in cases where that particular feature cannot apply. Since both -1 and 0 refer to an absence of the given feature, they were both treated as 0 for the purposes of modelling. The tokens in the dataset vary in their number of segments: for example, *shh* (/ʃ/) has just one segment while *retina* (/ɹɛtənə/) has six. Since a linear regression model requires the same number of variables for each datapoint, the mean of the segmental features was taken as a general representation of the articulatory information for each token. However, this does mean that sequential information is lost. Another approach was trialled, where the length of each vector sequence was normalised through the addition of empty vectors, but this failed to produce useful results. In order to narrow down the features to be used for linear regression, a correlation matrix was used as an indicator of how the mean phonetic information is distributed among the words in the dataset. As Figure 2 shows, a number of features had no values, specifically the spread/constricted glottis, velaric, long, high tone and high register features (10, 11, 20, 22, 23 and 24 respectively); this is very likely because these features are not present or are unspecified in English. These features were then eliminated for the purposes of the linear regression, since they are not informative. Figure 2 also shows there are not any strong correlations between the remaining variables, and since our primary goal is to see whether such feature-based representations are predictive of iconicity, we included all informative dimensions. For more information about the phonetic information captured by the Panphon features, see Appendix I.

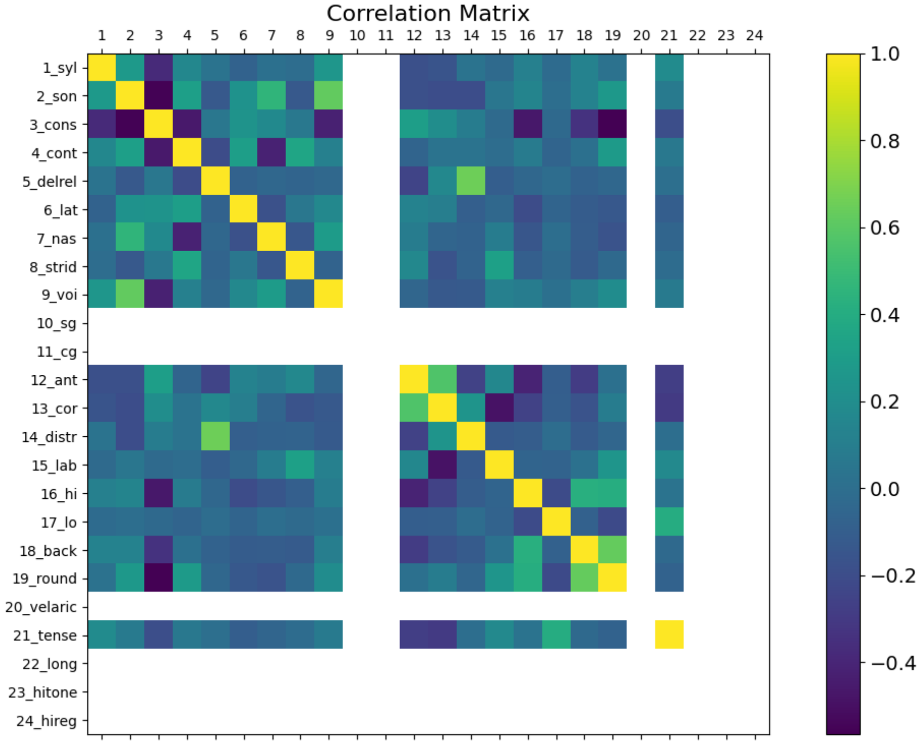

Figure 2: Panphon Pearson’s correlation matrix for mean phonetic representations of words in the dataset.

3.3 Semantic Embeddings

The DSM used for this study is the Fasttext English pre-trained word embeddings (Mikolov et al. 2017). These embeddings consist of 300 dimensions, and the number of dimensions needed to be reduced in order to obtain optimal results. Linear regression models with very high numbers of features compared to the number of samples in the dataset can run into issues such as overfitting and a lack of interpretability, although in this case the semantic features are not directly interpretable in the first place, along with increased computational costs. Principal Component Analysis (PCA), as implemented in the Scikit-learn Python package (Pedregosa et al., 2011), was used in order to reduce the number of dimensions. In order to ascertain the most informative number of dimensions for the linear model, preliminary linear models were run with 1-300 dimensions, each using a randomised 70/30 train/test data split to avoid overfitting, and their root mean square errors (RMSE) were taken as a measure of their performance. The results of this step are shown in Figure 6. 185 dimensions were used for the final model as this model had the lowest RMSE. Additionally, in order to assess the relationship between the semantic model's predictions and concreteness/abstractness, the concreteness ratings from Brysbaert et al. (2014) were added for each word. These ratings were also based on native speaker intuitions, where participants were asked to rate words on a scale of 1 (abstract) to 5 (concrete), and these ratings were also rescaled to 0 to 1 for the purposes of analysis.

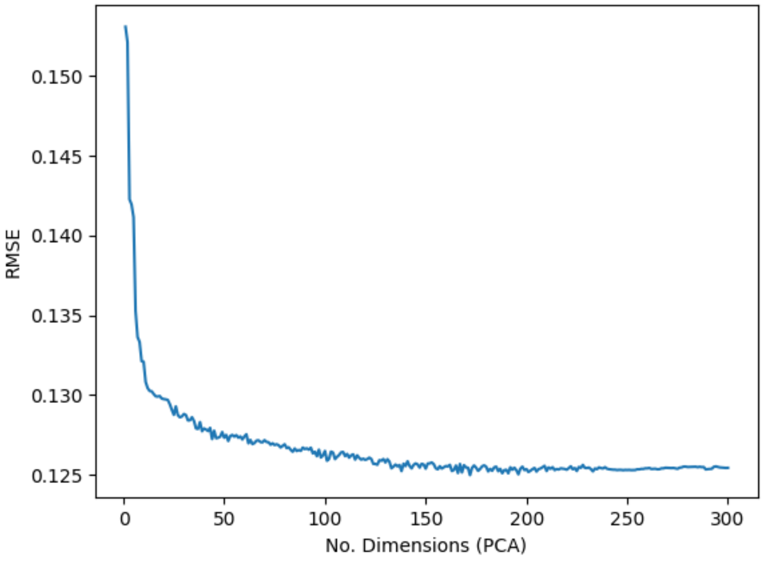

Figure 3: Semantic vector models - number of dimensions (x-axis) vs. RMSE (y-axis).

Taking stock, the iconicity dataset, phonetic vectors and semantic vectors were used to create three linear regression models: a **phonetic model** using the 18 phonetic features as independent variables, a **semantic model** using the 35 dimensions from PCA as independent variables, and a **combined model** which concatenated both the phonetic and semantic features for the independent variables. In all three cases the iconicity score was the dependent variable. The linear regression models were fitted with the Scikit-learn Python package (Pedregosa et al., 2011). R^2, mean absolute error (MAE), mean square error (MSE) and root mean square error (RMSE) were taken as performance metrics for each model. The Kendall's

τ score was taken as the rank correlation between the word list ordered by each model prediction measured against the same list ordered by iconicity rating, and also as the rank correlation between the semantic model's predictions and the aforementioned concreteness ratings.

The Python code for preprocessing and analysis can be found at: https://github.com/jotadwright/iconicity_embeddings

4. Results

The performance metrics for each model can be found in Table 2. It should be noted that higher numbers of variables artificially inflate the R^2 score, which may be the case for the semantic and combined models (which have 185 and 203 variables respectively). The Kendall's **τ** score is given for the rank correlation with iconicity ratings for all model predictions, while the Kendall's **τ** score for correlation with concreteness is just given for the semantic model. The R^2 score and error measures show that the semantic model is better at predicting the iconicity score of a given word, while the Kendall's **τ** scores for correlation with iconicity show that the phonetic and combined models are better at ranking the vocabulary from most to least iconic. This is also reflected in Figure 4, which compares the distribution of iconicity scores with the predictions of the different models. All three models tend to make predictions with a more limited range of scores than the original iconicity scores, while the semantic model's predictions have a wider distribution than the very limited range of predicted scores from the phonetic and combined models. The phonetic and combined models perform very similarly on all measures, suggesting that the method of concatenating phonetic and semantic data was not productive.

Model	**R^2**	**MAE**	**MSE**	**RMSE**	**Kendall's τ** (correlation with iconicity)	**Kendall's τ** (correlation with concreteness)
Phonetic	10.38	0.11	0.02	0.14	0.21	-
Semantic	34.11	0.1	0.02	0.12	0.12	0.10
Combined	11.55	0.11	0.02	0.14	0.22	-

Table 2: Performance metrics for the three models

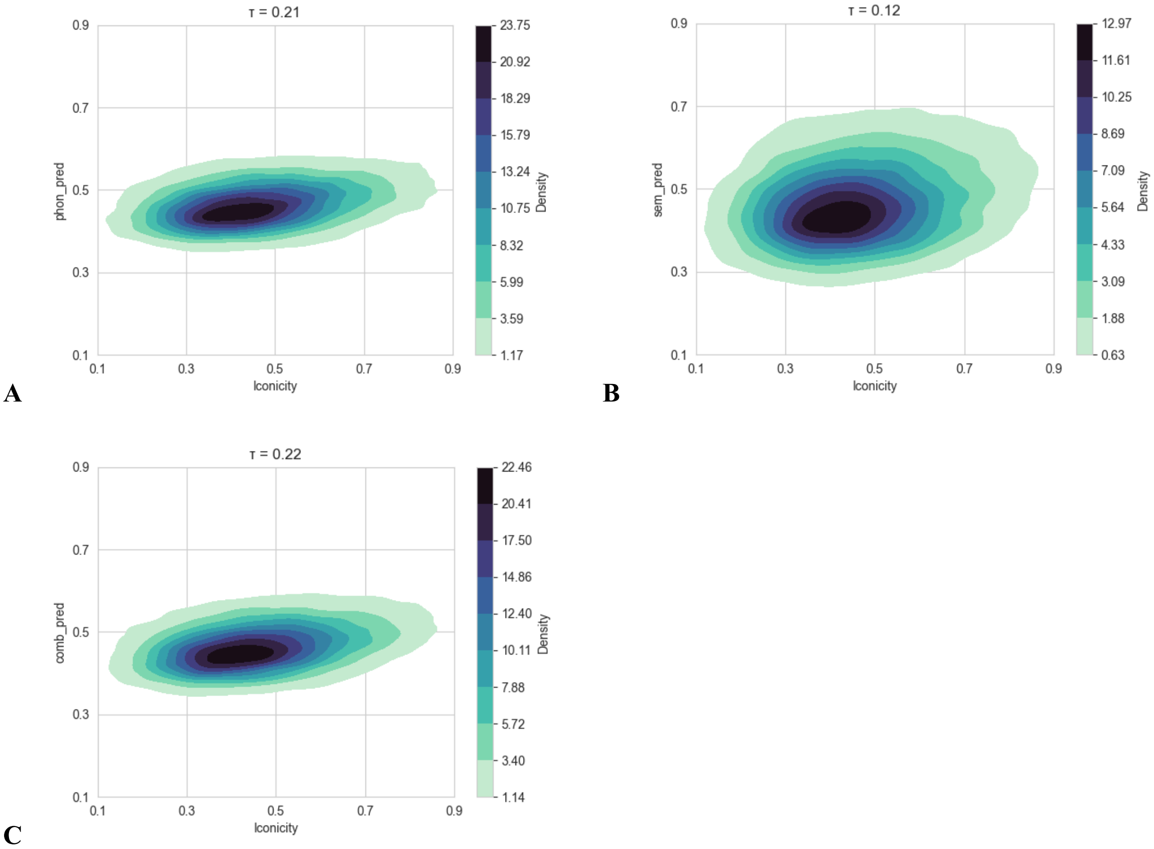

Figure 4: 2D kernel density plots showing the distribution of iconicity ratings (x-axes) versus the distribution of the predictions (y-axes) of the phonetic model (A), semantic model (B) and combined model (C).

The feature coefficients for the phonetic model (see Appendix II) indicate the correlation between each feature and iconicity ratings in the phonetic model, bearing in mind that the mean of all segments was taken for each word. These coefficients indicate that [±consonantal], [±sonorant], [±distributed], [±high] and [±tense] features are positively correlated with iconicity, while [±nasal], [±strident] and [±syllabic] segments are negatively correlated. For more information about the Panphon features, see the Appendix I. Since the semantic dimensions are not directly interpretable, this has been omitted for the semantic and combined models.

Tables 3, 4 and 5 show which words each model predicted as the 10 most and 10 least iconic, which serve to illustrate the trends in words which each model selects as iconic or not iconic, although this obviously represents a small sample of the 14,000 word vocabulary. These tables contain the iconicity ratings (from Winter et al., 2023) and the model predictions for comparison, and Table 4 also contains the concreteness ratings from Brysbaert et al. (2014). Examining these results we can see examples of where the model has correctly predicted a word's iconicity rating (e.g. *hooch*, Table 3; *area*, Table 5), examples where the model errs in giving a high prediction for iconicity (e.g. *clue*, Table 3; *booth*, Table 5) or by giving a low prediction (e.g. *dwell* and *unfulfilled*, Table 4). These tables also demonstrate what is shown in Figure 4: that the phonetic and combined models fail to make predictions in the upper and lower extremes of the iconicity scale.

Word	IPA	Iconicity	Phonetic model prediction
shh	ʃ	0.9	0.67
hooch	hut͡ʃ	0.62	0.65
booth	buθ	0.35	0.64
who	hu	0.43	0.64
hoop	hup	0.61	0.64
cool	kul	0.72	0.63
clue	klu	0.21	0.63
glue	glu	0.6	0.63
ghoul	gul	0.67	0.63
puke	puk	0.74	0.63
heir	ɛɹ	0.28	0.31
err	ɛɹ	0.85	0.31
air	ɛɹ	0.38	0.31
retina	ɹɛtənə	0.39	0.31
Aurora	ɹ̩ɔɹə	0.4	0.29
area	ɛɹiə	0.23	0.29
enema	ɛnəmə	0.14	0.28
aura	ɔɹə	0.39	0.27
era	ɛɹə	0.3	0.25
eh	ɛ	0.77	0.04

Table 3: Phonetic model predictions. Darker hues indicate values closer to 1 while lighter hues indicate those closer to 0.

Word	IPA	Iconicity	Semantic model prediction	Concreteness
pooch	put͡ʃ	0.52	0.84	0.83
mouse	maws	0.6	0.83	0.96
huge	hjud͡ʒ	0.58	0.82	0.64
clipped	klɪpt	0.6	0.8	0.6
igloo	ɪglu	0.53	0.78	0.93
coop	kup	0.5	0.78	0.87
sooth	suθ	0.73	0.78	0.28
cheesecake	t͡ʃizkejk	0.67	0.78	0.99
camp	kæmp	0.45	0.77	0.84

blues	bluz	0.68	0.77	0.33
abandoned	əbændənd	0.53	0.23	0.38
evangelicalism	ɛvænd͡ʒɛlɪkəlɪzəm	0.39	0.23	0.13
fantasize	fæntəsajz	0.5	0.23	0.29
sufficiency	səfɪʃənsi	0.42	0.23	0.2
unfulfilled	ʌnfʊlfɪld	0.62	0.23	0.16
repair	ɹɪpɛɹ	0.45	0.23	0.58
fern	fɹ̩n	0.37	0.22	1
dwell	dwɛl	0.74	0.22	0.41
erotica	ɹ̩ɑtɪkə	0.58	0.22	0.57
pain	pejn	0.58	0.2	0.63

Table 4: Semantic model predictions. Darker hues indicate values closer to 1 while lighter hues indicate those closer to 0.

Word	IPA	Iconicity	Combined model prediction
shh	ʃ	0.9	0.69
booth	buθ	0.35	0.66
hoop	hup	0.61	0.66
hooch	hut͡ʃ	0.62	0.65
puke	puk	0.74	0.65
who	hu	0.43	0.65
coup	ku	0.4	0.65
sooth	suθ	0.73	0.65
gloom	glum	0.87	0.64
clank	klæŋk	0.89	0.64
nausea	nɔziə	0.53	0.31
etcetera	ɛtsɛtɹ̩ə	0.63	0.31
retina	ɹɛtənə	0.39	0.3
Aurora	ɹ̩ɔɹə	0.4	0.3
air	ɛɹ	0.38	0.3
enema	ɛnəmə	0.14	0.29
aura	ɔɹə	0.39	0.29
area	ɛɹiə	0.23	0.27
era	ɛɹə	0.3	0.24
eh	ɛ	0.77	0.12

Table 5: Combined model predictions. Darker hues indicate values closer to 1 while lighter hues indicate those closer to 0.

5. Discussion

The performance metrics shown in Table 2 indicate that the semantic model performs best overall, with lower error measures and a higher R^2. The phonetic model performs worse on these metrics, but it does have a better Kendall's **τ** score, which suggests that the model's ranking of words from most to least iconic is more similar to that of the human iconicity ratings. The combined model performs very similarly to the phonetic model, making a very slight improvement in the metrics, suggesting that the combined model prioritises the phonetic variables over the semantic ones.

The phonetic model coefficients suggest that high proportions of consonantal, sonorant and distributed segments are positively correlated with iconicity, while nasal, strident and syllabic segments are negatively correlated. However, examination of the top and bottom predictions for the phonetic model perhaps give a clearer picture of the kind of phonemic structures the phonetic model prefers: we can see the repetition of the phonemes /u/, /k/ and /h/, and a trend of generally short words. Examining the words that the phonetic model deems least iconic we can see the repetition of the phonemes /ɛ/, /ə/ and /ɹ/. The top predicted word, /ʃ/, is a pure consonant, in line with the high weight the model gives the consonantal variable. The fact that /u/, a rounded back vowel, occurs so frequently in the top predictions is interesting given that the rounded and back features were not given particularly high importance by the model. This perhaps suggests that interaction between the different phonetic variables is rather complex, and the averaging of the scores over a word likely also contributes to this. Nonetheless, the model does seem to consistently select certain segments, which results in some relatively accurate approximations of high-iconicity words (*shh*, *cool*, *hooch*) with some phonetically similar errors (*clue, booth*). There seems to be some convergence with Sidhu et al.'s (2021: 1395) examination of roundness and spikiness in English nouns, where the phoneme /u/ was found to be iconically related to roundness, while /ʃ/, /t͡ʃ/ and /k/ were all linked to object spikiness. All of these phonemes occur among the top ten words as predicted by the phonetic model, /u/ and /k/ repeatedly. /k/ and /u/ were also found in a number of strong worldwide sound-meaning associations in Blasi et al. (2016). Overall, the metrics for the phonetic model show that variation in the averaged phonetic vectors can explain variation in iconicity rating to some extent, although this relationship is perhaps not as strong as for the semantic model, and analysis of the predictions shows some concurrence, in terms of phonemes proven to have iconic properties, with other research.

The combined model performs very similarly to the phonetic model as the performance metrics show. The combined model's top predictions also reveal great similarity with the phonetic model, as we can see a repetition of the phonemes /u/, /k/ and /h/ among the top predictions and of the phonemes /ɛ/, /ə/ and /ɹ/ among the bottom predictions. The two models are not identical, there are differences in the predicted iconicity they give for many words, but the overall picture indicates that the combined model is very strongly influenced by the same variables which are included in the phonetic model.

The semantic model, however, clearly gives very different results. The coefficients for the semantic model are derived from pre-trained word embeddings, reduced to a smaller number of dimensions via PCA, and it is not possible to directly interpret them, as the dimensions themselves do not refer to semantic properties but instead are used conjunctively to locate words in a semantic space (Boleda, 2020). The top predictions for the semantic model do not generally seem to be semantically related. Nevertheless, comparing the top and bottom predictions does suggest that this model selects concrete words over abstract words, with the top words containing concrete terms such as *mouse, clipped* and *huge*, while abstract concepts such as *abandoned, fantasize* and *evangelicalism* are found with low iconicity predictions. A notable exception to this trend is *fern*, with a low iconicity prediction and high concreteness rating. This trend is reflected in the Kendall's rank correlation between the semantic model's iconicity predictions and concreteness ratings for the same words. The score reflects a positive, although relatively weak, correlation between the semantic model's predictions and concreteness ratings, meaning that the model generally selects words with higher concreteness ratings as being more iconic. Given that the semantic model learns from the distributional information in semantic word embeddings, it makes sense that the model would pick up on the difference in use of abstract and concrete words. This is in line with Lupyan & Winter's argument that "iconicity limits abstraction and abstraction limits iconicity" (2018: 6), as it is generally harder to perceive similarity between a sound and an abstract concept than, say, a shape. However, it should be noted that there are some contradictory findings with respect to concreteness and iconicity, with concreteness ratings found to negatively correlate with iconicity ratings in some studies (Winter et al., 2023). Surprisingly, the most obvious element that almost all the words in the top semantic predictions share is the phoneme /u/, suggesting that the model has somehow indirectly selected words with this phoneme. Word length also seems to be a factor, with the top ten semantic model predictions having a mean length in number of segments of 3.8, while the bottom ten has a mean of 8. Overall, the semantic model has the strongest performance metrics, although the rank correlation shows that it is not as strong at ordering the words from highest to lowest iconicity, and analysis of the model's predictions shows that it is likely tapping into the aforementioned concreteness/abstractness phenomenon.

DSMs, the semantic embeddings used in this study, derive semantic representations for words based on their linguistic contexts. In this system, a word's meaning is determined by the words which it tends to be used with, and words used in similar contexts will have similar meanings. As discussed in section 2.2, this is a model of language which only makes use of linguistic information, and does not directly include the sensorimotor information with which we interact with the world around us. In iconic words, the ground which links the representamen with the object represented is one of similarity, and as such the nature of the object itself is an essential component of an iconic representation. The findings of this study, and others in which DSMs have been shown to contain information about iconicity (Dingemanse & Thompson, 2020; Thompson et al., 2020; De Varda & Strapparava, 2022), suggest that DSMs contain information about this relationship of similarity between signifier and signified, one which is grounded in sensorimotor information, despite not having direct access to sensorimotor input.

The purpose of this study was to identify whether there is information about iconicity in phonetic and semantic word embeddings through the use of linear regression models and native speaker iconicity ratings. The results of these models show that these embeddings do have some explanatory value with respect to iconicity ratings. However, these findings have some limitations. Firstly, iconicity ratings such as those used in this study (from Winter et al., 2023) are based on native speaker intuitions. This is in part a strength, as iconicity is subjective by nature, yet there is a degree of consistency in this phenomenon, in the way in which users of a language identify iconicity across the lexicon, and taking average ratings can capture this (Winter & Perlman, 2021). However, these ratings do not at all explain why any given word is iconic or not, and do not give information about specific form-meaning relationships. What's more, the issue of the extent to which participants confuse the concept of iconicity with other ideas, such as semantic transparency, has been raised with respect to the results of such iconicity rating experiments. Dingemanse and Thompson (2020) pointed out that compound words such as 'dishwasher' and 'skateboard' received high iconicity ratings in Perry et al. (2017), yet are not actually iconic, and they attributed this to the phrase *a word sounding like what it means* causing a confusion between semantic transparency and iconicity. As such, there is a possibility of a bias towards such words having inflated iconicity ratings in datasets like the one used for this study. Similar criticisms have been made over the reliability of the concreteness ratings used in the analysis of the semantic model's predictions, such as that the mean ratings in the middle sometimes reflect where there is disagreement between participants (Pollock, 2018). Another limitation is in the nature of semantic word embeddings, which locate words in vector spaces where the semantic distance between words can be measured. However, the dimensions that make up semantic word embeddings do not refer to any specific semantic properties, and therefore it is not possible to extrapolate, for example, what semantic content caused the semantic model to rate *pooch* as more iconic than *independence,* from the word embeddings themselves. Nonetheless, the embeddings and ratings do have explanatory value for the purposes of this research, and the convergence of the results presented here with preexisting iconicity research supports this.

Further research on this subject could build upon these findings with multi-modal word embeddings, which are enhanced with perceptual input and could be exploited to explore the cross-modal nature of iconicity in language (Verő & Copestake, 2021). The incorporation of sensorimotor input via visual embeddings (as used in De Varda & Strapparava, 2022) could give insight into the importance of sensorimotor information in iconicity, and the extent to which such information is latent in DSMs. This study focuses on iconicity ratings in English, and similar work in other languages would be valuable and create opportunities for cross-linguistic analysis. Panphon has functionality for a number of different languages, semantic word embeddings for a large variety of languages are also widely available, and iconicity ratings are being developed for other languages too (for example, Hinojosa et al. (2021) for Spanish).

6. Conclusion

This study explored the phenomenon of iconicity in language and its relationship to phonetics and semantics. The concept of iconicity challenges the conventional Saussurean notion of the arbitrariness of linguistic signs, suggesting that there can be a connection between the form of a word and its meaning. The study examined the different types of signs in Peircian semiotics, highlighting the role of similarity as the ground for iconic signs and the cross-modal nature of many iconic signs. The Bouba-Kiki effect, which demonstrates a cross-modal sound-to-shape mapping, and the presence of ideophones in various languages are examples of iconicity in action. By utilising natural language processing tools and native speaker iconicity ratings, this study aims to contribute to current research on iconicity. Linear regression models with phonetic vector representations and DSMs as explanatory variables and iconicity ratings as the response variable were used with the aim of determining the extent to which information about iconicity is encoded in these embeddings. The findings of this study indicate that these models were able to predict iconicity ratings with a certain degree of success. Moreover, the trends observed in the predictions align with previous research on iconicity, particularly with certain phonemes repeatedly found in iconic words, and in the relationship between iconicity and concreteness. These findings also have implications for understanding the degree to which sensorimotor information is latent in DSMs, given that iconicity is grounded in sensorimotor information. This research demonstrates the potential for incorporating natural language processing techniques and word embeddings in exploring the role of iconicity in language. Future studies could build upon these findings by utilising multi-modal word embeddings to further enrich the words' semantic representations, or by using the approach taken in this paper with other languages.

Bibliography

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

Baroni, M. (2016). Grounding distributional semantics in the visual world. *Language and Linguistics Compass*, 10(1), 3-13.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers*) (pp. 238-247).

Barnes, K. (2023). Subjectivity, perception and convention in ideophones and iconicity. *SKASE Journal of Theoretical Linguistics*, *20*(1).

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, 59, 617-645.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818-10823.

Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, 6, 213-234.

Bruni, E., Boleda, G., Baroni, M., & Tran, N. K. (2012). Distributional semantics in technicolor. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 136-145).

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, 46, 904-911.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., ... & Winter, B. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200390.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, *19*(10), 603-615.

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: a journal of general linguistics*, 3(1).

Dingemanse, M., & Thompson, B. (2020). Playful iconicity: Structural markedness underlies the relation between funniness and iconicity. Language and Cognition, 12(1), 203-224.

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge University Press.

Ferrara, L., & Hodge, G. (2018). Language as description, indication, and depiction. *Frontiers in Psychology*, 9, 716.

Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*, 10-32.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Scienc*e, 14(6), 1006-1033.

Hinojosa, J. A., Haro, J., Magallares, S., Duñabeitia, J. A., & Ferré, P. (2021). Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behavior Research Methods*, 53, 1262-1275.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the Royal Society B: Biological sciences*, 369(1651), 20130298.

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170137.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.0940*5.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130299.

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., & Levin, L. (2016). Panphon: A resource for mapping IPA segments to articulatory feature vectors. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 3475-3484).

Mortensen, D. R., Dalmia, S., & Littell, P. (2018). Epitran: Precision G2P for many languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Passaro, L. C., Bondielli, A., & Lenci, A. (2017). Learning affect with distributional semantic models. *IJCoL. Italian Journal of Computational Linguistics*, 3(3-2), 23-36.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.

Peirce, C. S. (1992). On a new list of categories (1867). *The writings of Charles S. Peirce: a chronological edition*, 2, 49-58.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in psychology*, 1, 227.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130300.

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, 21(3), e12572.

Pollock, L. (2018). Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study. *Behavior Research Methods*, 50(3), 1198-1216.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. Journal of consciousness studies, 8(12), 3-34.

Saussure, F. M. (1972). *Course in general linguistics*. Open Court.

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, 28, 1390-1398.

Thompson, B., Perlman, M., Lupyan, G., Sehyr, Z. S., & Emmorey, K. (2020). A data-driven approach to the semantics of iconicity in American Sign Language and English. *Language and Cognition*, 12(1), 182-202.

de Varda, A. G., & Strapparava, C. (2022). A Cross‐Modal and Cross‐lingual Study of Iconicity in Language: Insights From Deep Learning. *Cognitive Science*, 46(6), e13147.

Verő, A. L., & Copestake, A. (2021). Efficient Multi-Modal Embeddings from Structured Data. *arXiv preprint arXiv:2110.02577*.

Vinson, D., Jones, M., Sidhu, D. M., Lau-Zhu, A., Santiago, J., & Vigliocco, G. (2021). Iconicity emerges and is maintained in spoken language. *Journal of experimental psychology. General*, 150(11), 2293–2308.

Westera, M., & Boleda, G. (2019). Don't blame distributional semantics if it can't do entailment. *arXiv preprint arXiv*:1905.07356.

Wingfield, C., & Connell, L. (2022). Understanding the role of linguistic distributional knowledge in cognition. *Language, Cognition and Neuroscience*, 37(10), 1220-1270.

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies*, 18(3), 443-464.

Winter, B., & Perlman, M. (2021). Iconicity ratings really do measure iconicity, and they open a new window onto the nature of language. *Linguistics Vanguard*, 7(1).

Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., & Perlman, M. (2023). Iconicity ratings for 14,000+ English words. *Behavior Research Methods*, 1-16.

Appendix I

[±syllabic]	Is the segment the nucleus of a syllable?
[±sonorant]	Is the segment produced with a relatively unobstructed vocal tract?
[±consonantal]	Is the segment consonantal (not a vowel or glide, or laryngeal consonant)?
[±continuant]	Is the segment produced with continuous oral airflow?
[±delayed release]	Is the segment an affricate?
[±lateral]	Is the segment produced with a lateral constriction?
[±nasal]	Is the segment produced with nasal airflow?
[±strident]	Is the segment produced with noisy friction?
[±voice]	Are the vocal folds vibrating during the production of the segment?
[±anterior]	Is a constriction made in the front of the vocal tract?
[±coronal]	Is the tip or blade of the tongue used to make a constriction?
[±distributed]	Is a coronal constriction distributed laterally?
[±labial]	Does the segment involve constrictions with or of the lips?
[±high]	Is the segment produced with the tongue body raised?
[±low]	Is the segment produced with the tongue body lowered?
[±back]	Is the segment produced with the tongue body in a posterior position?
[±round]	Is the segment produced with the lips rounded?
[±tense]	Is the segment produced with an advanced tongue root?

Panphon features used, descriptions taken from Mortensen et al. (2016)

Appendix II

Feature	**Coefficient**
[±syllabic]	-0.187
[±sonorant]	0.152
[±consonantal]	0.433
[±continuant]	0.057
[±delayed release]	0.045
[±lateral]	-0.083
[±nasal]	-0.204
[±strident]	-0.13
[±voice]	-0.002
[±anterior]	0.054
[±coronal]	-0.044
[±distributed]	0.146
[±labial]	0.077
[±high]	0.162
[±low]	0.059
[±back]	0.027
[±round]	0.048
[±tense]	0.156

Phonetic model feature coefficients.