

Examining Iconicity Information in Semantic and Phonetic Word Embeddings

Extended Abstract

Jamie Wright

Supervised by Thomas Brochhagen

Universitat Pompeu Fabra

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-ø-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. 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 vector representations and distributional semantics models (DSMs). 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 that iconicity is found in words with certain phonetic and semantic qualities. This aligns with the findings of previous studies which also indicate that DSMs contain iconicity information (Dingemanse & Thompson, 2020; Thompson et al., 2020; De Varda & Strapparava, 2022). 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. These findings contribute to our understanding of the distribution of iconicity in language, and indicate that computational tools such as DSMs and phonetic vector representations are able to detect iconicity in natural language to a significant degree, despite not having direct access to the words' referents, a fundamental component in an iconic sign.

References

- Barnes, K. (2023). Subjectivity, perception and convention in ideophones and iconicity. *SKASE Journal of Theoretical Linguistics*, 20(1).
- 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.
- Ferrara, L., & Hodge, G. (2018). Language as description, indication, and depiction. *Frontiers in Psychology*, 9, 716.
- 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.
- 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.
- Saussure, F. M. (1972). *Course in general linguistics*. Open Court.
- 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.

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.

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