

Evidence of sound-meaning cues in a typologically and genealogically diverse dataset of two hundred languages

Pablo Contreras

Abstract

Recent work has shown that, contrary to traditional assumptions (de Saussure, 1916), the sound of a lexical item can carry cues about its meaning. In one prominent case, “iconic” words are those whose sound somehow resembles the sensorimotor attributes of its referent. Iconicity has been found to be pervasive across a variety of languages’ lexicon, and it facilitates the acquisition of the specific, high-iconicity words (see Dingemanse et al., 2015, for a review).

We focus on a different, less studied type of sound-meaning pattern: sound systematic cues. These are patterns of similarity not between words and referents but between words with roughly similar meanings. In contrast to embodied iconic cues, systematic cues can inform the generalization that learners perform of their already acquired knowledge to new instances. For example, Monaghan and colleagues (2007) found systematic phonological patterns showing that verbs sound more similar to other verbs than to nouns and vice versa. These patterns have further been shown to facilitate acquisition in children (Shi et al., 1998) and language processing in adults (Farmer et al., 2006).

One weakness of previous work on sound systematic patterns is the limited set of languages in which it has been tested. Indeed, they have been tested only in a small number of languages from the industrialized world (e.g., Fourtassi et al., 2020; Monaghan et al., 2007), which limits the generalizability of the claims regarding their role in language acquisition and processing. In this study, we address this problem by evaluating the sound systematic cues present in a typologically and genealogically diverse resource: the Intercontinental Dictionary Series (IDS) (Key and Comrie, 2015). Our study analyzes word lists for 200 languages belonging to 59 different language families, most of them understudied, from a broad range of geographical diversity (Figure 1).

The words in the IDS are linked to broad domains of meaning. In our study, we compared the two largest categories: words for actions and word for things. We found that, overall, within-category similarities in sound are larger than those between the categories, confirming that there are sound-systematic patterns in the usable word lists of the IDS. Moreover, these patterns obtain in most of the languages in the resource, although the effect size varies widely between them (Figure 2).

Given that previous work attributes these patterns a role in facilitating language acquisition through providing cues for generalization, we tested whether the patterns in the IDS could be used in a learning scenario. For these, we tested the performance of different algorithms trained only on the phoneme sequences of the words: a nearest phonological neighbor procedure and a small recurrent neural network. We found that both procedures achieve better-than-chance performance, even when the latter is trained on a small portion of the data.

In summary, our study shows that phonological links between words of the same category are not confined to languages from rich or industrialized regions of the world. Instead, they can

be found in a wide array of understudied languages. Moreover, our findings confirm that these phonological patterns can be used to generalize already acquired knowledge of a language's vocabulary, further illuminating the rich environmental statistical structure that scaffolds early language acquisition.

De Saussure, F. (1916). *Cours de Linguistique Générale*. Paycot.

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

Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children's semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, 44, e12847.

Key, M. R., & Comrie, B. (2015). *The Intercontinental Dictionary Series*.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive psychology*, 55(4), 259-305.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, 25, 169–201.