Felicia Stich

Title: Detecting Form-Meaning Systematicity in the Early Lexicon: A Cross-Linguistic Corpus Study

Abstract:

Recent years have witnessed a resurgence of interest in motivated form-meaning mappings (Dingemanse et al., 2015), challenging traditional notions about the arbitrariness of the sign. One type of non-arbitrariness is systematicity – the idea that words with similar meanings tend to have similar forms. Corpus studies show that systematic mappings pervade the world's vocabularies (Dautriche et al., 2017) and are privileged in acquisition (Cassani & Limacher, 2022). It has been suggested that systematic patterns scaffold early word learning by providing formal cues towards semantic category membership when contextual information is sparse (Monaghan et al., 2014). However, the extent of systematicity in the early lexicon has never been examined directly. To investigate the systematicity of early mappings, we conducted a comparative corpus study with 19 languages from different areas and families. For each language, we calculated pairwise phonological and semantic distances between all nouns in Wordbank, a cross-linguistic database which indexes 500–700 words typically acquired at 16 to 36 months. We used Euclidean distances between phonemic feature vectors and cosine distances between fastText embeddings as distance metrics. The correlation between phonological and semantic distances was assessed by Mantel tests with 100,000 permutations. Our results showed subtle, but highly significant negative correlations for Dutch, Norwegian, Danish, Finnish, and Cantonese, such that semantically similar words tended to be phonologically distinct (Spearman's $\rho$ between –.130 and –.111, all $p < .001$). Results for the other languages failed to reach significance but were qualitatively similar, except for German, Spanish, Portuguese, Persian, and Korean. Our findings indicate that the early lexicon is biased more towards discriminability than towards categorizability, despite cross-linguistic variation. Thus, the role of systematicity in acquisition seems less clear-cut than assumed. The learning advantage reported by previous studies may be restricted to localized clusters which were not detected by our lexicon-wide analysis. Hence, our study opens up exciting avenues for research into the small-scale structure of the early lexicon.