Comparing the efficacy of bigrams and frames in cuing lexical categories for human learners

Toben Mintz
University of Southern California

Abstract: Distributional patterns in linear word sequences are informative about lexical categories in natural languages. However, it is unclear what patterns human learners use to categorize words. The current study exposed adults to sentences in two artificial languages. The Frames group heard sentences in which target words were embedded in frequent frames (FFs)—frequently co-occurring words that immediately surrounded/framed a target word (e.g., you_X_it); the FFs categorized the target words into two distinct categories. The Bigrams-Only group heard sentences with the same vocabulary, but no FFs. However, the bigram information in the two languages was equally informative regarding the categories of target words. In a 2AFC test, Frames subjects but not the Bigrams-Only subjects chose novel sentences that contained target words as "correct" if they adhered to the abstract distributional structure, compared to "ungrammatical" sentences that were identical on all surface dimensions. Thus, learners categorized using FFs but not bigrams alone.