

Phonological Cues to Syntactic Structure in a Large-Scale Corpus

Ethan Wilcox

Harvard University, Cambridge, Massachusetts, United States

Abstract

The Prosodic Bootstrapping Theory (PBT) states that prosodic and phonetic cues assist infant language learners to segment the speech stream into words and assemble those words into phrase structures. However, many of the studies demonstrating a link between prosody and syntax were conducted on small data sets and on a narrow range of syntactic structures. This work uses a state-of-the-art parser to syntactically annotate the BU Radio News Corpus of around 16,000 diverse sentences, which are prosodically tagged and annotated. A decision tree classifier was fit, using six prosodic features and achieving 87% accuracy at differentiating words internal to major syntactic phrases vs. words that mark phrase boundaries. However, the models tested are unable to differentiate between phrasal categories based on prosodic information alone. These results provide new evidence in support of the Prosodic Bootstrapping Theory, suggesting it is possible to identify phrasal boundaries based on prosodic information alone.