Some clustering with k-means: acoustics of a prosodic contrast in a corpus of spontaneous speech

Anca Chereches̩, Cornell University

Many theories of prosody (e.g. Selkirk 1996, Calhoun 2006) predict that function words like *some* will be highly reduced in English (1a), except in a few specific environments, such as under focus (1b-c). The present study quantifies the link between focus and the acoustic realization of the quantifier *some* using a novel technique.

1. a) **[sm]** people brought us fresh cilantro.
   b) Cilantro? **[sʌm]** people love it, **[sʌm]** people hate it.
   c) I guess **SOME** people like cilantro.

A previous corpus study linked focus to prosodic prominence (Calhoun 2006), but it did not zoom in on function words and it was conservative in its focus annotation. The annotation guidelines favor explicit contrasts (1b), and recommend avoiding trickier cases of implicit contrast, such as in (1c), where a prominent *some* can trigger the scalar inference that not all people like cilantro. Such inferences are assumed by most semanticists to be very common, but judgments are actually sensitive to contextual information, including (potentially subtle) prosodic cues (de Marneffe & Tonhauser 2014). Other types of function words (prepositions, conjunctions, etc.) get focused more rarely, and often to correct some previous expression. Corrective focus tends to be very marked acoustically. Thus, *some* provides an interesting case study, where a word is predicted by prosodic theories to shift from low to high relative prominence under focus in general, but for types of focus like implicit contrast, it is not yet clear how this is reflected in its acoustics.

To address this question, I collected from online sports radio talk shows a corpus of utterances of *some* followed by *people* (resulting in 143 focused and 54 unfocused *some*s) and by *money* (4 focused, 194 unfocused *some*s), for a total of 395 tokens. I measured vowel duration, intensity, formants, F0 range, and size and alignment of F0 extrema. I manually annotated for focus. Since I felt relatively uncertain in labeling implicit contrast (e.g. scalar inferences), I did not train a supervised model like logistic regression. Instead, I ran k-means clustering on all combinations of acoustic measurements (8191 combinations) and selected for analysis the 177 models which best matched my manual labels and performed better than the highest baseline (85.3%).

The sheer number of successful unsupervised models suggests that for this dataset, focus correlates with a specific acoustic realization of *some*, which can best be characterized by increased duration, more extreme formants and an F0 peak (all measurements taken from the vowel in *some*). Many combinations of features achieve identical or near-identical cluster separation, suggesting that there is a great deal of redundancy in acoustic cues to focus. The single feature that appears in all winning models is duration of the *some* vowel. On its own, it separates the tokens with 86.7% accuracy, compared to the highest accuracy of 88.5% for duration, combined with formant and F0 values. Preliminary error analysis suggests that misclassified tokens were also difficult to label. In future work, we intend to quantify this labeling uncertainty by increasing the number of annotators.

To conclude, the present study supports prosodic analyses of function words which account for the strong influence of focus. Despite uncertainty in labeling focus when the contrast was implicit, acoustically we found remarkably good separation between focused and unfocused *some*s, especially along the duration dimension, and to a smaller extent formants and F0. Future work can strengthen these results by recruiting extra annotators and further processing the dataset to allow for more sophisticated normalization techniques.

References

- Calhoun, Sasha. 2006. "Information Structure and the Prosodic Structure of English: A Probabilistic Relationship." Ph.D., University of Edinburgh.
- de Marneffe, Marie-Catherine, and Judith Tonhauser. 2014. "Prosody Affects Scalar Implicature Generation." Poster presented at CUNY 2014.
- Selkirk, Elisabeth. 1996. "The Prosodic Structure of Function Words." Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition 187: 214.