Integrating Linguistic and Referential Cues Into Probabilistic Models of Word Learning

Michael C. Frank (mcfrank@mit.edu)
Vikash Mansinghka (vkm@mit.edu)
Edward Gibson (egibson@mit.edu)
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
43 Vassar St., Cambridge, MA 02139 USA

Keywords: Language acquisition; statistical learning; Bayesian models; word learning.

How do infants and young children learn the words of their native language? Although some words are modeled by caregivers as single word utterances, most words that children learn occur only in fluent, un-segmented speech. Proposals for how infants learn to segment fluent speech have tended to split between accounts focusing on the use of cues such as the conditional probability between particular syllables (e.g., Swingley, 2005; Aslin, Saffran, & Newport, 1998) and accounts focusing on prosodic cues such as word-level stress (e.g., Cutler & Butterfield, 1992).

We examine a family of probabilistic computational models derived from Brent (1999). Rather than attempting to use one particular information source (either conditional probability or stress) to find word boundaries, these models use probabilistic inference methods to derive a lexicon which might have generated a particular corpus. We present experimental work which shows that adult participants learning artificial, un-segmented languages can (1) learn a prosodic cue for segmentation and generalize it to novel vocabulary in that language, and (2) learn associations between novel objects and words. These results suggest that word-based models of segmentation allow for a parsimonious integration of a variety of information sources for word learning and word segmentation.

Experiment 1: Learning Prosodic Cues

We investigated whether adult participants with pre-existing knowledge of their native language would be able to learn a novel prosodic cue to segment utterances in an artificial language. Participants heard randomly generated, unsegmented utterances created via the concatenation of six words with different lengths but a uniform prosodic shape (a 30Hz dip in pitch on either the initial or final syllable of every word, location varied between subjects). At test, participants heard utterances in a novel vocabulary with the same stress pattern as their initial vocabulary. After each utterance they were asked to choose which one of two parts of the utterance sounded most like a word in the language: a word or a segment of the same length which crossed a word boundary (e.g., "this-is-an-ele-phant" would have the test items "an-ele" and "ele-phant"). We found that, while initial stress was easier, participants in both conditions were able to segment the novel utterances above chance ($t(299) = 5.00$, $p < .001$) using the stress cue they heard.

Experiment 2: Learning Word Meanings

We asked whether adults were able to use distributional information from two modalities simultaneously in the service of learning the form and meaning of words in a simple artificial language. Participants in our experiment heard randomly generated sentences of un-segmented, synthesized speech created via concatenation of words of various lengths. Each sentence contained both a random number of filler words and exactly one meaning word which corresponded to a simultaneously presented picture of a novel object, all arranged in a random order. A control condition was identical save that the association between the meaning words and the novel objects was not fixed, so no meaning word was associated with any particular object. Participants were tested both on the forms of the words they heard as well as on the correspondence between particular meaning words and objects. Participants in the experimental condition learned both word meanings and word forms at a level significantly greater than those in the control condition ($t(568) = 3.68$, $p < .001$, and $t(1141) = 4.70$, $p < .001$, respectively), suggesting that they were able to use distributional information from two different modalities to associate words in un-segmented speech with their referents.

Acknowledgments

Thanks to Josh Tenenbaum and the members of Tedlab and Cocosci for useful discussions of this material.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month old infants. *Psychological Science, 9*, 321-325.

Brent, M. R. (1999). An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning, 34*, 71-105.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31*, 218–236.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*, 86-132.

Saffran, J.R., Newport, E.L., & Aslin, R.N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.