# Learning word segmentation from acoustic data 

Kyle Gorman<br>kgorman@ling.upenn.edu<br>University of Pennsylvania

November 13, 2007

Word segmentation in infants is an important process that feeds much of phonological acquisition. How is it that children as young as 7.5 months can segment words, given the noisy input? The earliest work in Generative Grammar (Harris, 1955; Chomsky, 1955) proposes that use of transition probabilities (TPs) can be used to help children learn word boundaries, and it has been shown that infants can make use of distributional information (Saffran et al., 1996). However, TPs alone allows for poor performance. Gambell and Yang (2005) proposes an algebraic algorithm making use of lexical stress, which achieves greater accuracy, but both methods require structured information to be available to the learner.

In this talk, I'll review both the psychological and computational literature, and ask the question "What would a psychologically plausible word segmentation device look like?" I believe the answer may be considering how much can be learned from the acoustic base of the input, and what if anything can be done during the generation of the lexicon.

## References

Chomsky, N. (1955). The Logical Structure of Linguistic Theory. University of Chicago Press.

Gambell, T. and Yang, C. (2005). Mechanisms and constraints in word segmentation. Ms.
Harris, Z. (1955). From phoneme to morpheme. Language, 31(2).
Saffran, J., Aslin, R., and Newport, E. (1996). Statistical Learning by 8-Month-Old Infants. Science, 274(5294):1926-1928.

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

