Syllable-based Bayesian inference: A (more) plausible model of word segmentation

Because knowledge of words plays a crucial role in acquisition and children seem to accomplish word segmentation very early (~7.5 months (Jusczyk et al., 1999; Echols et al., 1997; Jusczyk et al., 1993a)), many models have been proposed for how children learn to identify words in their native language. Most models of word segmentation have assumed that the learner has already discovered their language's phonemic inventory at the age that they begin tackling this problem (e.g., Goldwater, Griffiths, Johnson 2009 (GGJ 2009); Brent & Siskind, 1999; Pearl, Goldwater & Steyvers, 2011 (PGS 2011)). However, experimental evidence suggests that infants do not yet possess a full grasp of their language's phonemic inventory this early on (Jusczyk et al., 1993b; Jusczyk et al., 1994). Instead, syllables seem to be stronger representational units for infants (Jusczyk & Derrah, 1987; Eimas, 1999). We modify existing, highly successful, phoneme-based statistical models of word segmentation that use Bayesian inference (GGJ 2009; PGS 2011) to operate over syllables and so create a more psychologically faithful model of word segmentation. In doing so, we demonstrate the robustness of a Bayesian approach, and additionally find support for the utility of processing constraints on language acquisition.

We test our syllable-based models using child-directed speech (Pearl-Brent corpus, CHILDES: MacWhinney, 2000). We compare the results of our *ideal* learners, which have no processing constraints, to *constrained* learners that segment utterances as they are encountered and sometimes perform non-optimal statistical inference. Additionally, we compare modeled learners that assume words are produced independent of all other words (a *unigram* assumption (GGJ, 2009)) with modeled learner that assume a word depends on the word that occurred directly before it (a *bigram* assumption (GGJ, 2009)).

We show that syllable-based models perform comparably with previous phoneme-based models of word segmentation (as shown in Table 1), suggesting that Bayesian inference-based strategies are still plausible for infants. Interestingly, while we do find that an ideal Bayesian learner with perfect, unlimited memory resources can succeed at this task, crucially we find results similar to PGS (2011) that constrained learners who learn incrementally and with limited memory, as in the case of actual children, *outperform* the ideal learner. Although PGS (2011) found this to be the case only for certain types of learners, here we find that the constrained learner outperforms the ideal learner *in all cases*. This is in line with the "Less is More" hypothesis (Newport, 1990), where limited memory resources help, rather than hinder, aspects of language acquisition. In the broader picture, this study highlights the benefit of grounding computational models with empirical research: not only can we find what strategies are likely to be used by children, but we may also discover potential explanations for existing, sometimes puzzling, observations about child language acquisition, as with the "Less is More" hypothesis.

Unigram Models						
	TP	TR	TF	BP	BR	BF
Ideal	54.01	32.64	40.69	99	43.96	60.88
DPM	76.23	57.14	65.31	96.48	62.53	75.88
DPS	76.86	60.93	67.97	95.19	67.46	78.96
Bigram Models						
Ideal	73.95	60.17	66.35	96.14	70.96	81.65
DPM	81.73	71.72	76.4	95.07	78.71	86.12
DPS	83.8	75.3	79.33	95.16	81.59	87.86
Comparison Models						
TransProb	53.03	37.57	43.98	90.00	53.14	66.82

Table 1. Model results are compared on two sets of measures, Word Tokens (T) and Boundaries (B). Precision (P) = % of guesses made that were correct. Recall (R) = % of tokens/boundaries correctly identified. F-score (F) = average of P & R. Ideal = optimally segments, sees all data at once. DPM = optimally segments, but sees data incrementally. DPS = sub-optimally segments and sees data incrementally. A syllabic transitional probability learner (from Saffran, Aslin, & Newport (1996)) is provided as a baseline.

References

Brent, M.R. & Siskind, J.M. 2001. The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* 31-44.

Echols, C.H., Crowhurst, M.J. & Childers, J.B. 1997. The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, 36, 202-225.

Eimas, P.D. 1999. Segmental and syllabic representations in the perception of speech by young infants. *Journal of the Acoustical Society of America*, 105(3), 1901-1911.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. 2009. Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20,* 579–585.

Goldwater, S., Griffiths, T. & Johnson, M. 2009. A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition 112*(1), 21-54.

Jusczyk, P.W. & Derrah, C. 1987. Representation of speech sounds by young infants. *Developmental Psychology*, 23(5), 648-654.

Jusczyk, P.W., Cutler, A. & Redanz, N.J. 1993a. Infants' preference for the predominant stress patterns of English words. *Child Development*, 64(3), 675-687.

Jusczyk, P.W., Friederici, A.D., Wessels, J.M.I., Svenkerud, V.Y. & Jusczyk, A.M. 1993b. Infants's sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 402-420.

Jusczyk, P.W., Luce, P.A. & Charles-Luce, J. 1994. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630-645.

Jusczyk, P.W., Houston, D.M. & Newsome, M. 1999. The beginnings of word segmentation in English learning infants. *Cognitive Psychology*, 39, 159-207.

MacWhinney, B. 2000. *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Newport, E. 1990. Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

Pearl, L., Goldwater, S., & Steyvers, M. 2011. Online Learning Mechanisms for Bayesian Models of Word Segmentation, *Research on Language and Computation*, special issue on computational models of language acquisition. DOI 10.1007/s11168-011-9074-5.

Xu, F. & Tenenbaum, J.B. 2007. Word learning as Bayesian inference. *Psychological Review, 114*(2), 245-272.