Speech Segmentation and Word Discovery: PARSER

Taylor Cassidy[1], Danny Nassre[1]

[1] CUNY Graduate Center

Introduction

• How do infants segment continuous, unfamiliar speech input and discover words from it, when there aren't pauses between most words?

• The word boundary probability (WBP) of a two-phoneme sequence p_1p_2 in a corpus:

$$WBP(p_1p_2) = \frac{\#\ occurrences\ of\ p_1p_2\ across\ a\ word\ boundary}{\#\ occurrences\ of\ p_1p_2\ within\ a\ word\ boundary}$$

• The word boundary probability distribution (WBPD) is highly bimodal, and can be thought of as a probabilistic manifestation of phonotactic constraints.

• Drawing a word boundary in between any phonemes in a corpus with a WBP of greater than .5 yields a precision of 75.5% and a recall of 69.2%[1].

• How can the WBPD/phonotactic knowledge be accessed without any prior language-specific knowledge? PARSER might be one way.

PARSER and PARSER 2

• PARSER[2] models two cognitive processes:

  • the (initially random) chunking of unfamiliar input.

  • the retention in memory of more frequently chunked sequences (which are more likely to be actual words) in a lexicon, and their use in subsequent chunking.

  • One cycle of the algorithm:

    (1) Randomly determine size of next chunk (range:1-3).

    (2) Processing an utterance (or remainder of utterance) from left to right, add segments to chunk until chunk size is reached. A segments can be a primitive unit of input (a syllable, in the original implementation) or items in the lexicon built from these units – whichever is longer.

    (3) Add new chunk to lexicon, giving it a weight of 1, or increase the weight of the lexical item corresponding to the chunk by 1 if it already exists.

    (4) Increment the weights of any lexical items comprising the chunk by .5.

    (5) Decrement the weights of all lexical items not encountered during the chunking.

    (6) Return to (1).

  • The output of the model is a lexicon of perception shapers, which can be seen as hypothesized words, and their associated weights.

• PARSER2 is our implementation of PARSER. It processes a corpus utterance-by-utterance (decrementing the items in the lexicon after each utterance) and it uses either purely phonemic or phoneme-based "partially-syllabic" input:

  • There is evidence that children are more sensitive to syllables than to phonemes early on[3,4].

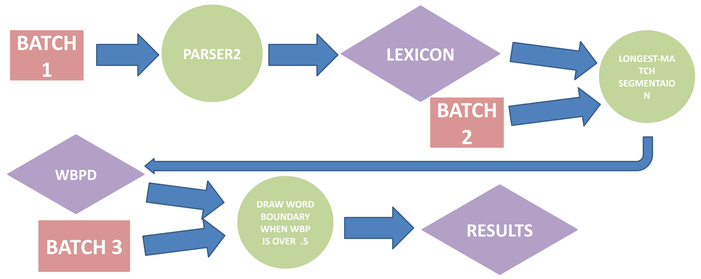
    • But knowledge of syllable *boundaries* does not follow from this. So, the input to PARSER2 is processed on a phoneme-by-phoneme basis, with the restriction that all chunks must contain at least one vowel.

• This models a language-independent process in which children first form syllables around vowels, and then refine their understanding of word boundaries as they proceed.

•Automatic extraction of highly-weighted lexical items was also introduced, to avoid repeatedly chunking them with other segments.

• Various linear forgetting rates were used.

Procedure



• Multiple instances of PARSER2 were run on a batch of 200,000 CHILDES[5] utterances with variation in the forgetting rate (.005 to .5), input type (phonemic or partially syllabic), and the automatic recognition parameter (the required weight a lexical item must achieve to be automatically extracted; none to 10).

•The lexicons produced by PARSER2 were used to segment a second batch of 100,000 utterances by finding the longest match from the beginning of each utterance or utterance remainder. WBPDs were calculated from the boundary information given by this set of segmentations.

•These WBPDs were then used to segment a third batch of 100,000 utterances, by drawing a word boundary between two phonemes with a WBP greater than .5.

Results

•Lexicon that yielded best segmentation:

  • 410 correct word types out of 774 candidates = 53% precision.

  • In addition to correct types, multi-word clusters and syllables comprised 282 additional candidates.

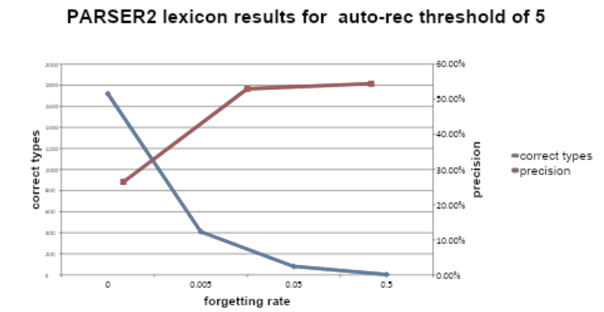
  • Correct types, multi-word clusters, and syllables accounted for 89% of the lexicon.

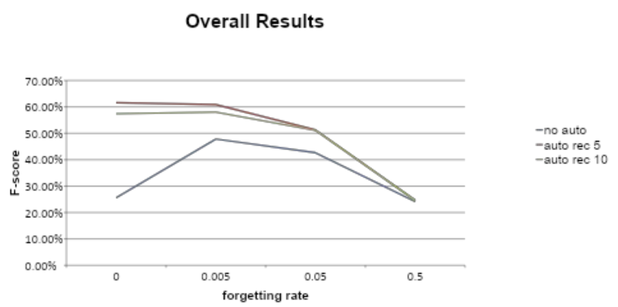
  • 82% of the word tokens in the data were accounted for by the word types in the lexicon.

•Segmentation results:

  • Auto-recognition significantly improved performance.

  • Best-performing instance with forgetting: 60% precision and 61% recall (compare to 75% precision and 69% recall with perfect word boundary information).





Discussion

• Non-word items in the lexicon (multi-word utterances and syllables) can still be useful for segmentation[7].

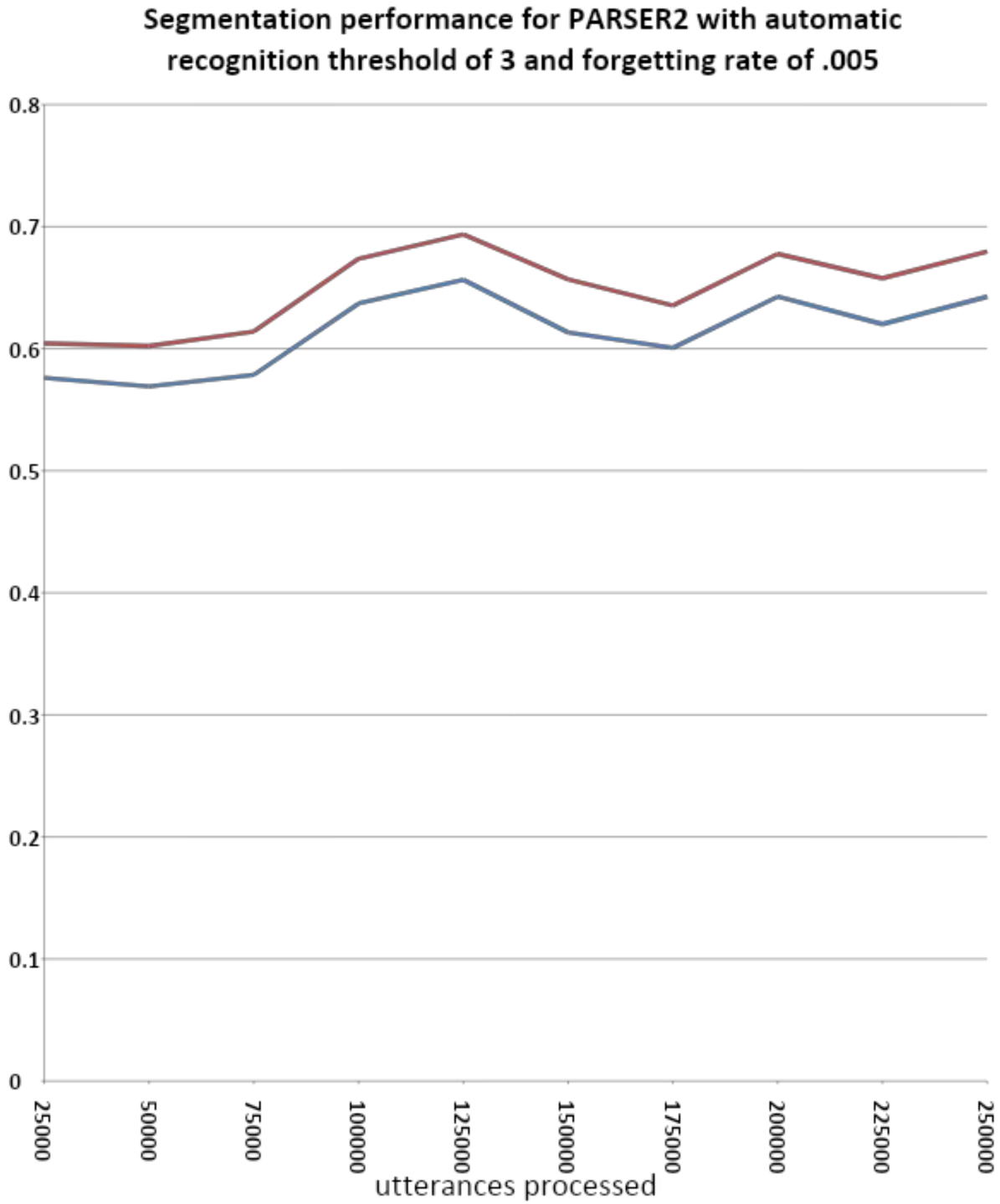
• The 200,000 utterances processed by PARSER likely represent only a few hundred hours of parent-child interaction, according to extrapolation from hour-long transcripts in the Brown Corpus in CHILDES.

• There is evidence that phonotactic awareness plays an important role in SSWD[6]. The WBPD created by PARSER yields a segmentation whose accuracy is close to what's achievable with a perfect WBPD; PARSER might be how this is achieved.

• A less-naïve method of using the PARSER lexicon to get word boundary information (*e.g.* by incorporating the items' weight) or the use of the lexicon as input to a more sophisticated word recognition model, should yield significantly improved results.

References

1.Christiansen, M., Onnis, L, & Hockema, S. (2009). The secret is in the sound: from unsegmented speech to lexical categories. *Developmental Science, 12*, 388–395.
2.Perruchet, P. and Vintner, A. (1998). PARSER: a model for word segmentation. *Journal of Memory and Language.*, *39*, 246–263.
3.Bijeljac-Babic, R., Bertoncini, J. and Mehler, J. (1993) How do four-day-old infants categorize multisyllabic utterances? *Developmental Psychology, 29*, 711–721.
4.Liberman, I., Shankweiler, F., Fisher, F. and Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.
5.MacWhinney, B., & Snow, C. (1985). The Child Language Data Exchange System. *Journal of Child Language, 12*, 271–295.
6.Graf Estes, K. (2011). Phonotactic constraints on infant word learning. *Infancy, 16*, 180-197.
7.Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, 50, 86–132.


Segmentation performance for PARSER2 with automatic recognition threshold of 3 and forgetting rate of .005
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
25000
50000
75000
100000
125000
150000
175000
200000
225000
250000
utterances processed