Statistical word segmentation in spontaneous child-directed speech of Korean

Anonymous ACL submission

Abstract

The present study demonstrates advantages of child-directed speech (CDS) over adult-directed speech (ADS) in statistical word segmentation of spontaneous Korean. We derived phonetic input from phonemic corpus by applying a set of phonological rules. For modeling the statistical word segmentation based on transitional probability (TP), we used two syllable-based algorithms (i.e., Absolute and Relative) in two directions (i.e., Forward TP and Backward TP). Results show that (i) segmentation accuracy is greater with phonetic input than phonemic, (ii) The model performs better when trained on CDS than ADS, and (iii) segmentation accuracy improves with child age.

1 Introduction

A prerequisite for infants to build a lexicon for word learning is the ability to segment words out of the speech stream. (Jusczyk and Aslin, 1995). One of the postulated mechanisms for attaining such an ability is statistical segmentation based on transitional probabilities (TP; Brent and Cartwright, 1996; Harris, 1955; Saffran et al., 1996; Aslin et al., 1998). Behavioral studies suggest that infants segments words more easily in CDS (child-directed speech) than ADS (adult-directed speech) (Thiessen et al., 2005). However, CDS advantages in word segmentation has not yet been verified in statistical modeling and is debated (Cristia et al., 2019). For instance, Fourtassi et al. (2013) show that segmentation performance is better with CDS than ADS whereas Cristia et al. (2019) report mixed or minimal advantages of CDS depending on various algorithms applied.

Another important issue with CDS as the input to the processing system of language-learning children is the adaptive nature of CDS over the course of child's language development (Snow, 1972). For example, repetitions in CDS decrease, utterance length increases and vocabulary types increase with child age (Henning et al., 2005; Soderstrom, 2007). Due to the changing nature of the input, it could be that the segmentation accuracy on CDS based on statistical algorithm might also change with child age.

This research investigates the question of CDS advantages over ADS in statistical segmentation of words with Korean, an agglutinative verb-final language with a phonemic syllabary. Despite much research on word segmentation, very few studies have investigated word segmentation in non-European languages with behavioral or statistical methods. One of the unique aspects of our data is that we converted the phoneme-based transcription to phonetic input by applying a comprehensive phonological rules. We first compare the performances of various algorithms based on transitional probability (TP) to seek the most optimal algorithm for segmentation in Korean. We then investigate the question of CDS advantages in word segmentation based on spontaneous corpora of CDS and ADS. Finally, in consideration of the fine-tuning hypotheis of CDS (Snow, 1972), we examine any developmental changes in the model performance of segmentation.

2 Methods

2.1 Corpora

Our modeling was based on two corpora of spontaneous speech. For the CDS data, we used the Ko corpus (Ko et al., 2020) containing 35 mothers freely interacting with their own children for about 40 minutes. The same corpus also contains ADS in which the mother talks to their family members and experimenters for about 10 minutes. To complement the relatively small data size of the ADS, we used additional data from the Call Friend Korean corpus (Ko et al., 2003), which contains casual telephone conversations between friends.

Dataset	Child-directed speech		Adult-directed speech	
	Phoneme	Phonetic Input	Phoneme	Phonetic Input
Types				
Words	8,987	9,970	12,452	13,895
Syllables	1,093	1,260	1,051	1,296
Tokens				
Words	65,940	62,188	68,516	64,434
Syllables	149,269	149,222	147,188	146,928
Types/tokens ratio				
Words	7.337	6.237	5.502	4.637
Syllables	136.568	118.430	140.045	113.37

Table 1: Descriptive statistics of the data

2.2 Rule-based phoneme-to-phonetic input

The process of grapheme-to-phoneme usually refers to the conversion of written word forms to phonetic transcription. This process is essential for processing languages where the spelling system is phonetic but contains irregularities (e.g. English) or the writing system is complex (e.g. Chinese or Japanese). The Korean writing system is a phoneme-based syllabary, with the alphabet system invented in 1446 by King Sejong based on linguistic principles. In this sense, Korean does not need a separate grapheme-to-phoneme process beyond a simple transliteration of Korean to Roman letters.

What sets apart our process from previous approaches is an application of phonological rules to turn the orthographic/phonemic corpus to phonetic transcription reflecting phonological alternations. In English, for example, the /n/ in 'green book' is pronounced as [m] due to the *assimilation* rule applying across word boundaries. Most previous research, however, simply adopts the dictionary pronunciation of each word without reflecting such phonotactic alternations occurring across word boundaries. An accurate reflection of the phonotactic patterns, however, is important for modeling since infants use them as one of the important cues for word-segmentation (Mattys and Jusczyk, 2001). For example, when they encounter a sequence of phones not allowed within words (e.g. [f][t]), they posit a word boundary between the two phones.

In this study, we applied a comprehensive set of phonological rules to approximate the actual input to infants' processing system. At the segmental level, for example, an application of the *assimilation* rule changes */n/* into *[l]* as in 잘 노네/jal none/ -> [jallone] '(Someone) plays well.'. Prior to the application of the segmental rules, however, we also applied *Accentual Phrase (AP) formation* at the prosodic level, whereby a monosyllabic function word or a high frequency adverb cliticizes to its host to form an AP (Jun, 1996) as in 잘 노네 /jal none/ -> 잘노네 [jalnone]. There is extensive evidence showing that the AP is the basic unit of phonological processing in Korean phonology (Cho and Flemming, 2015; Kim, 2000; Jun, 1998). We thus assumed that the statistical processing for word-segmentation might operate over the prosodically defined unit of AP for infants. Table 1 presents the detailed statistics of data for phonemic and phonetic input [1].

2.3 Syllable-based word segmentation model

For model training, we devised syllable-based Transitional Probability (TP) models by employing two algorithms (i.e., absolute and relative) and two measures (i.e., Forward TP and Backward TP). Forward TP (FTP) for a syllable sequence AB, for example, measures the frequency of the occurrence of the sequence AB divided by the frequency of A. Backward TP (BTP) for AB measures the frequency of occurrence of the syllable sequence AB divided by the frequency of B. The Relative algorithm assigns a boundary when a dip in TP is found. For instance, given the syllable sequence ABCD, there will be a boundary between B and C if the TP of BC is lower than the TP of AB and CD. The Absolute algorithm assigns a boundary when the TP is lower than the mean TP's at word boundaries in the entire corpus. We then employed the *k*-fold cross-validation technique to obtain a normalized result from each model (Stone, 1974). We set the value of *k* as 10 and repeated the cross-validation 10 times,

[1] Our data and entire code are available at: Github URL

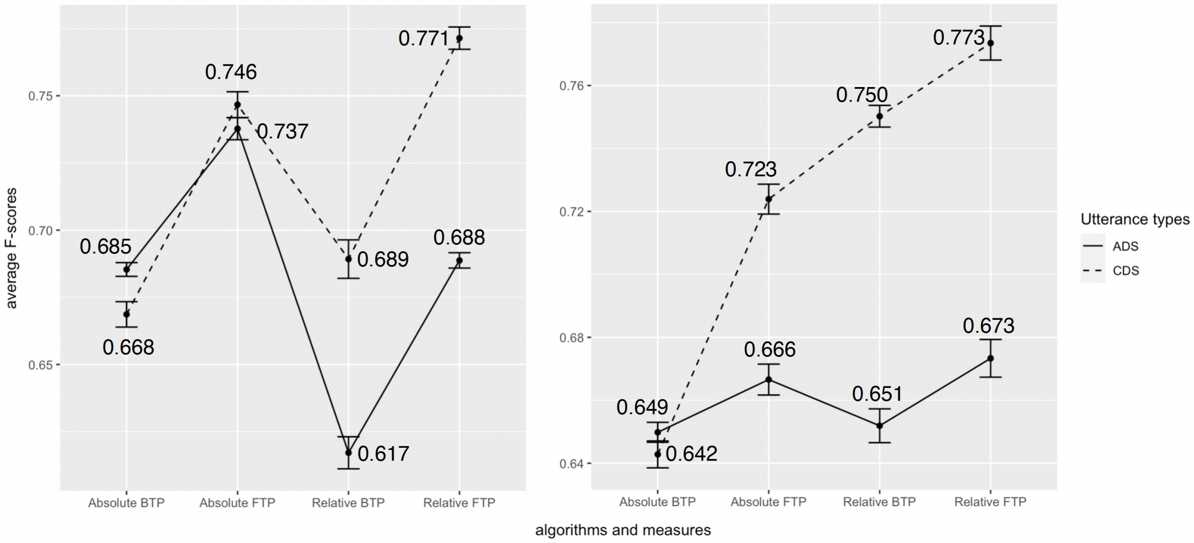

Figure 1: Means of F-scores by different algorithms and measures for phoneme (left) and phonetic input (Right)

with each sub-sample used exactly once as the test set for the model training. Model performance was measured by comparing the word boundaries in the original input sentence with the word boundaries generated via each model.

3 Results and Discussion

Our main interest is in testing the CDS advantages in segmentation suggested in behavioral studies based on computational modeling. A related question is investigating any developmental changes in the segmentation accuracy in model performance. Before reporting our findings on these question, however, we will compare the model performance in the original orthographic/phonemic corpus and the phonetic corpus we derived.

3.1 Does phonetic input make segmentation easier than the phonemic stream?

Overall, models performed better when trained with phonetic input (Mean F of CDS=0.722; Mean F of ADS=0.660) than phonemic corpus (Mean F of CDS=0.719; Mean F of ADS=0.682) (Figure 1). The highest mean F-score was found with the model trained on the phonetic CDS input (0.773; Relative FTP).

As shown in Table 1, the number of syllables and word types and tokens differs in the phonemic and phonetic corpus due to the phonological rules applied. Specifically, applying the *AP formation* has an effect of reducing the number of word tokens by prosodically cliticizing certain monosyllabic words to a nearby host. Meanwhile, the application of segmental phonological rules could both increase (e.g. *word-initial devoicing* changes the consonant such as /b,d,g,j/ into [p,t,k,c^h]) or decrease (e.g. *neutralization* changes the syllable-final /s,j,t^h,c^h/ into [t]) the number of word types.

The individual effect of each of these rule applications on model performance is currently being investigated. The AP formation has an effect of increasing the word length, which could have had a negative effect on segmentation accuracy. However, given the improved model performance with the phonetic corpus, it might be actually easier to segment the speech into AP's than orthographic words. It could also be that the application of various segmental rules had a positive effect over and above the negative effect of AP formation because of its effect on phonotactics and type/token numbers. We plan to report on the results on these investigations in the revision.

3.2 Is CDS easier to segment than ADS in Korean?

We investigated if distributional cues are enhanced in some way in CDS compared to ADS, leading to an easier segmentation. Figure 1 shows the model performance of the word segmentation by different algorithms and measures for the phonemic and the phonetic transcription. Results show that the average F-score was the highest in CDS (0.773; Relative FTP in phonetic) and the lowest in ADS (0.617; Relative BTP in phonemic); other models yielded accuracy scores ranging from 0.642 to 0.771. As

Age group	Syllable type/token	Word type/token	Utterance length (mean (sd))	Syllables per word (mean (sd))
Age0	846 / 49,064	4,417 / 19,696	8.749 (7.761)	2.333 (1.222)
Age1	869 / 46,348	4,321 / 19,776	8.386 (7.833)	2.201 (1.102)
Age2	842 / 53,983	5,111 / 22,716	9.463 (8.566)	2.234 (1.041)

Table 2: Statistics of the data in CDS by different age groups

shown in Figure 1, the segmentation performance was higher in CDS than ADS in all models except for the Absolute BTP model. Overall, therefore, our results indicate that the model performs better when trained with CDS than ADS.

Our results, therefore, provide support to the notion in behavioral research (e.g., Thiessen et al., 2005) that CDS is easier to segment, and corroborate the findings in earlier modeling research (Fourtassi et al., 2013) that CDS provides distributional cues that makes it easier for infants to segment words than ADS. Features of CDS likely to conspire to yield a higher model performance are short utterances, a high proportion of isolated words, and frequent repetitions (e.g., Bernstein Ratner and Rooney, 2001; Soderstrom, 2007).

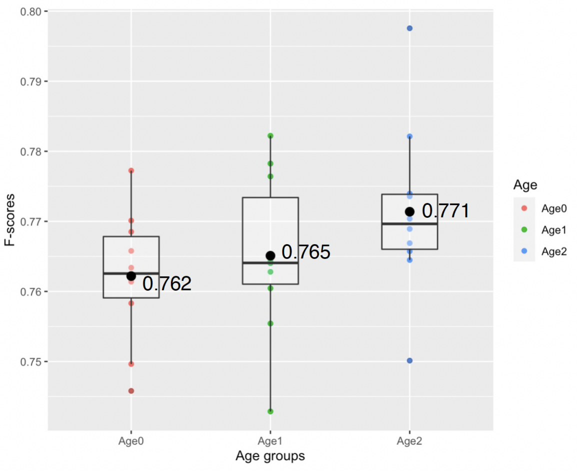

Figure 2: Changes of FTPs F-scores by ages (relative)

3.3 Does segmentation become easier with child age?

Our CDS corpus is a cross-sectional data set containing 11 or 12 mother-child dyads in each of the Age 0 (M = 0;08, preverbal), Age 1 (M = 1;02, early-speech), and Age 2 (M = 2;03, multi-word) group. We compared the model performance across these developmental stages to inspect any age effect in segmentation. We tested the hypothesis that segmentation might become easier with child age by measuring the model performance with the relative FTP.

As shown in Figure 2, we found the model performance improved with child age: the model performance was better in the Age2 (0.771) than the Age1 (0.765) or the Age0 (0.762) group. As shown in Table 2, word types and the mean length of utterance increase with child age. On the other hand, repetition decreases (Henning et al., 2005; Soderstrom, 2007). Contrary to the common notion, frequent repetitions, a characteristic of CDS, seems to lead to poorer segmentation. And statistical regularities seem to become more informative with the increased word types and tokens with child age.

4 Conclusion

The main findings of the current study can be summarized as follows. First, the model for segmentation performed better with a phonetic over a phonemic corpus. Second, CDS seems to have distributional cues yielding advantages over ADS in segmentation. Third, the model performance on CDS segmentation improved with child age. These findings, however, would need to be further verified by comparing the performance against the gold standard corpus for the phonetic input.

Our TP model is one of the first attempts to model statistical word segmentation with Korean, which turned out to be quite different from the models reported in European languages (e.g., Saksida et al., 2016). The difference could be due to typological differences in the language but also to methodological differences such as the data and the derivation of phonetic input based on phonological rules. While these issues need to be further clarified, our results are meaningful in that it provides demonstrations of CDS segmentation advantages based on a data set approximating ecological validity in its spontaneous nature and the phonetic derivation. Further, our finding of the age effect is one of the first addressing the developmental changes in the distributional cues for statistical word-segmentation in CDS.

References

Richard N. Aslin, Jenny R. Saffran, and Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4):321–324.

Nan Bernstein Ratner and Becky Rooney. 2001. How accessible is the lexicon in motherese? *Language Acquisition and Language Disorders*, 23:71—-78.

Michael R. Brent and Timothy A. Cartwright. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61(1):93–125. Compositional Language Acquisition.

Hyesun Cho and Edward Flemming. 2015. Compression and truncation: The case of seoul korean accentual phrase.

Alejandrina Cristia, Emmanuel Dupoux, Nan Bernstein Ratner, and Melanie Soderstrom. 2019. Segmentability Differences Between Child-Directed and Adult-Directed Speech: A Systematic Test With an Ecologically Valid Corpus. *Open Mind*, 3:13–22.

Abdellah Fourtassi, Benjamin Borschinger, Mark Johnson, and Emmanuel Dupoux. 2013. Whyisenglishsoeasytosegment. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 1–10.

Zellig S. Harris. 1955. From phoneme to morpheme. *Language*, 31(2):190–222.

Anne Henning, Tricia Striano, and Elena V.M. Lieven. 2005. Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development*, 28(4):519–536.

Sun-Ah Jun. 1996. *The phonetics and phonology of Korean prosody: Intonational phonology and prosodic structure*. Taylor & Francis.

Sun-Ah Jun. 1998. The accentual phrase in the korean prosodic hierarchy. *Phonology*, 15(2):189–226.

P. W. Jusczyk and R. N. Aslin. 1995. Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29(1):1—-23.

Soo-Jung Kim. 2000. *Accentual effects on phonological rules in Korean*. The University of North Carolina at Chapel Hill.

Eon-Suk Ko, Na-Rae Han, Stephanie Strassel, and Nii Martey. 2003. Korean telephone conversations transcripts. Accessed May. 16, 2003.

Eon-Suk. Ko, Jinyoung Jo, Kyung-Woon On, and Byoung-Tak Zhang. 2020. Introducing the ko corpus of korean mother-child interaction. *Frontiers in Psychology*.

Sven L Mattys and Peter W Jusczyk. 2001. Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78(2):91–121.

Jenny R. Saffran, Richard N. Aslin, and Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928.

A. Saksida, S. Iannuzzi, C. Bogliotti, Y. Chaix, Bricout L. Démonet, JF., C. Billard, MA. Nguyen-Morel, MF. Le Heuzey, I. Soares-Boucaud, F. George, JC. Ziegler, and F. Ramus. 2016. Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Dev Psychol*, 52(10):1503–1516.

Catherine E. Snow. 1972. Mothers' speech to children learning language. *Child Development*, 43(2):549—-565.

Melanie Soderstrom. 2007. Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27(4):501–532.

Mervyn Stone. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)*, pages 111–147.

Erik D. Thiessen, Emily A. Hill, and Jenny R. Saffran. 2005. Infant-directed speech facilitates word segmentation. *Infancy*, 7(1):53–71.