Transitional Probability and Word Segmentation

Yingying Xie

Correspondence: Yingying Xie, Department of English, The Chinese University of Hong Kong, New Territories, Hong Kong. Tel: 852-5388-7901. E-mail: xyyamanda@yahoo.com

Abstract

This article aims at reviewing the literature in the studies of the relationship between transitional probability and word segmentation in an attempt to emphasize statistical learning as the experience-dependent factor in language acquisition. Transitional probability, the crucial cue of the statistical relationship between syllables, is characterized by its two computation directions: the forward transitional probability and backward transitional probability. Results from the empirical research on artificial languages and natural languages are also discussed to prove the effectiveness and defectiveness of transitional probability in word segmentation.

Keywords: transitional probability, statistical learning, word segmentation

1. Introduction

Before producing their first comprehensible words, infants have already acquired some linguistic knowledge of their native language by relying on innately biased statistical learning mechanism. Word segmentation, a fundamental task of identifying words in fluent speech, is a good example to prove infants' possession of experience-dependent learning mechanism. The task of extracting words in spoken language is challenging, given the inconsistent pauses between word boundaries in continuous speech. However, infants are capable of completing the task by detecting and exploiting some cues. Statistical regularities in syllabic sequences are one type of those cues for infants to discriminate potential words in fluent speech after a short period of exposure to the language (Saffran, Aslin, & Newport, 1996). Statistical cues are used by infants earlier than other cues (Thiessen & Saffran, 2003). Based on artificial language (Saffran, Aslin, & Newport, 1996) and natural-language paradigms (Pelucchi, Hay, & Saffran, 2009), studies have revealed that infants can discover the statistical relationship between adjacent syllables, known as transitional probability (TP) which is measured by the frequency of the combination of syllables in the corpus of infant speech.

2. Computation Direction of Transitional Probability

Transitional probabilities are more apparent and greater within a word than those across word boundaries (Saffran, Aslin, & Newport, 1996). TP is typically calculated based on the following equation: $TP = P(Y|X) =$ Frequency (XY) / Frequency (X). For instance, in English word like *pretty*, the syllable *pre* can be followed by a set of syllables like *ty*, *tend*, and *vent*. The probability of the syllable *pre* preceding *ty* is roughly 80% in the infant's language environment. However, in the English phrase *pretty baby*, the final syllable *ty* can appear before any syllable of another English word, resulting in the extremely low probability of *ty* and *ba* (roughly 0.03% in speech to infants). Due to the striking difference in sequential probabilities of the syllables, a conclusion can be drawn that *pretty* is more likely to be a word than *tyba* (Saffran, 2003).

The example of the real word like *pretty* and the nonword *like typa* illustrates that infants can discover word boundaries with the help of forward transitional probabilities (FTP), for the first syllable predicts the second syllable in the disyllabic sequences. However, when it comes to situations like grammatical gender in Spanish words or grammatical category of 'noun' in English, backward transitional probabilities (BTP) render more useful information than FTP, for the backward transitional probabilities are higher than the corresponding forward probabilities under those particular circumstances. The combination of FTP and BTP generate more reliable cues for infants to detect word boundaries. The directionality of computation in language has no longer been restricted to FTP since the discovery of BTP, offering more convincing proof to support the practical function of transitional probability (Pelucchi, Hay & Saffran, 2009).

3. Application of Transitional Probability in Natural Languages

To investigate whether transitional probabilities still work well in word segmentation in natural authentic

language, Pelucchi, Hay, and Saffran (2009) abandoned the simplified artificial language and utilized Italian as the testing instrument. Italian, which is typologically different from English, is novel and complex to the English learning 8-month infants in the study. Nevertheless, infants could distinguish familiar words from novel words after familiarization with Italian speech in experiment one. Experiment two dispels the doubt that infants depended on individual syllables to recognize words by demonstrating infants' rapid response to the familiarity of syllable sequences. Experiment three further displays that infants succeeded in differentiating HTP-(high transitional probabilities) and LTP-(low transitional probabilities) words in Italian passages. The three experiments produce robust evidence to support infants' sensitivity to transitional probability cues in natural language stimuli and the importance of statistical learning in real-world language acquisition. Research on cross-linguistic word segmentation was also conducted by P. Jusczyk's lab, but the results of research showed that those English learning infants failed to segment words in Chinese even after several hours of exposure to Chinese input (Newman, Tsay, & Jusczyk, 2003). The results imply that not all linguistic input is subject to the influence of statistical learning.

4. Factors Influencing Application of Transitional Probabilities in Word Segmentation

The prosodic characteristics of language are considered as a factor influencing infants' application of transitional probability to distinguish word boundaries. Using artificial infant-directed (ID) speech and artificial adult-directed (AD) speech which share the same statistical information except the acoustic feature of pitch range, researchers found that infants succeeded in segmenting words by relying on transitional probabilities between syllables only after exposure to ID speech but not after the familiarization with AD (Thiessen, Hill & Saffran, 2005). Although the pitch characteristics of ID speech do not provide information of word boundaries, they capture the children's attention and win their preference for ID speech over AD speech. The results that ID speech facilitates infants' word segmentation indicate that ID speech contributes more information about the patterns of sound in words to infants than the AD speech.

Attention is also necessary for the language users to ensure the accuracy of word extraction. With the purpose to "test the extent to which attention is necessary for speech segmentation by statistical learning", three experiments characterized by different testing instruments were conducted by Toro and his colleague to compare the participants (undergraduate students) who were requested to passively listen to the speech stream and others who were required to listen and perform a concurrent task involving extra listening and visual activity (Toro, Sinnett, & Soto-Faraco, 2005). It comes as no surprise that the performance of the passive listening groups outmatched the group with high attention load in word-segmentation. The diversion of attention across modalities (for example, auditory and vision) and within modalities (auditory) both impair the word-segmentation based on statistical regularities. The results of the study are not directly related with word segmentation by infants, but they suggest that successful word extraction depends on some degree of attention even if the listeners can accomplish the task without explicit instruction.

Participants in many studies are healthy infants with no history of hearing or visual problems. Instead, Evans, Saffran & Robe-Torres (2009) investigate the function of transitional probability by comparing normal children (5 to 12 years old) with children (6 to 14 years old) who have specific linguistic impairment (difficulty in acquiring and using language resulting from hearing loss, intellectual, emotional and neurological impairment).Their results support the hypothesis that typically growing children with normal language have no problem in detecting word boundaries by tracking transitional probabilities in speech. However, it is found that this computation mechanism does not fully extend to children with specific linguistic impairment. Although they could track transitional probabilities in speech after double exposure time, they failed to distinguish newly learned target words from highly similar-sounding foils during the testing phase of the task. Their unsuccessful attempt is caused by their inability to recall the detailed phonological form of the target words and the nature of the synthesized linguistic stimuli which lack the variety of cues cooperating with TPs in natural language. For children with SLI, implicit learning mechanism like TPs is relatively fragile and ineffective.

5. Limitations of Transitional Probability in Word Segmentation

The limits of statistical learning for word segmentation emerge when the infants are faced with more complex linguistic input. Johnson and Tyler (2010) have tested whether transitional probabilities facilitate infants to distinguish word boundaries in an artificial language which contains more variations than natural language and more complexity than the artificial languages applied in previous studies. Participants of the study are two groups of 5.5 and 8 months old Dutch-learning infants, marking the study the first evidence to show that non-English learning infants under 6 months of age can resort to transitional probabilities to segment four words of the same length (all CVCV) in the designed artificial language. However, when the artificial language

consisted of four words of mixed word length (two CVCV, two CVCVCV), infants of both age groups failed to segment the language. The results of the study suggest that infants' application of transitional probabilities to segment words is not as effective as that suggested by previous studies based on simplified input. Constraints of word length are directly related with the final outcome of word segmentation.

Lew-Williams and Saffran (2012) further emphasizes the influence of previous knowledge of word length on infants' performance in segmenting words. Corresponding with the results of Johnson and Tyler's (2010), infants only succeed in segmenting words from fluent speech by obtaining information from transitional probability when words are either uniformly disyllabic or trisyllabic in both the pre-exposure phase and test phase. The occurrence of trisylliabic words in the pre-exposure phase undermines infants' judgment in distinguishing disyllabic words in the test phase, and vice versa. The findings of the study point out the inadequacy of statistical learning mechanisms in diversified language input. They also indicate that prior experience and knowledge of language shape statistical learning. Pre-exposure to shorter or longer words leads to infants' learning preference (Lew-Williams & Saffran, 2012).

The interaction between prior knowledge and statistical learning becomes more obvious as infants grow older. Older infants may disregard transitional probabilities when their native language provides them with other learned probabilistic cues to word boundaries (Finn & Hudson Kam, 2008). For example, transitional probabilities are still chosen by 6-month-olds for word segmentation while 9-month-olds favor the strong-weak stress pattern that is characteristic of English bisyllabic words (Thiessen & Saffran, 2003). When dealing with the speech stream which contains high TP between syllables and a weak-strong stress pattern in contrast to the strong-weak pattern of English, the younger infants relied on high transitional probabilities to figure out the syllable pairs. The older infants, on the other hand, segmented out items with lower syllable TPs by referring to English word-internal stress pattern. In another study, infants by 8 months of age showed more consideration to co-articulation information than TPs (Johnson & Jusczyk, 2001). There is general agreement that infants make full use of any or all cues available in their language input to solve linguistic problems like discovery of word boundaries (Saffran, Newport & Aslin, 1996).

There are times when TPs fail to work even after children has acquired fluency in their L1, as it happens in the case of children's errors caused by the mistaken computation. Oversegmentation occurs when a child, for instance, replies, ''I am/heyv/'' in response to the parent's expression of ''Behave!''(Peters, as cited in Saffran, Newport & Aslin, 1996). Previous knowledge of the syllable "be" and its grammatical function mislead the child to consider /heyv/ as an individual word without comprehending the meaning of the word. The opposite side of this type of errors is the errors of undersegmentation when a child regards a phrase like "ham-n-egg" as a single word (Saffran, Newport & Aslin, 1996). The high frequency of the collation confuses the child to associate the syllables across the words as an instance of high TPs.

6. Conclusion

Transitional probabilities alone cannot generalize word segmentation in all linguistic input because of exceptions and variability in different languages. However, the facilitative role of TPs should not be disregarded, as the experience with sound patterns in the early stage build the foundation for the lexical acquisition leading to meaningful communication. Moreover, transitional probabilities are often integrated with other cues like stress and co-articulation by infants to solve the problem of discovering word boundaries. Future research should explore more in cross-linguistic factors and the factors related with participants like adult second foreign language learners to shed new light on the topic of word segmentation.

References

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical Learning in Children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *52*, 321–335. http://dx.doi.org/10.1044/1092-4388(2009/07-0189)

Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*, *44*, 548–567. http://dx.doi.org/10.1006/jmla.2000.2755

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*, 477–499. http://dx.doi.org/10.1016/j.cognition.2008.04.002

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345. http://dx.doi.org/10.1111/j.1467-7687.2009.00886.x

Le-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*, *122*, 241-246. http://dx.doi.org/10.1016/j.cognition.2011.10.007

Newman, R., Tsay, J., & Jusczyk, P. W. (2003). Segmentation of Chinese words after exposure to Chinese. In D. Houston, A. Seidl, G. Hollich, E. Johnson, & A. Jusczyk (Eds.), *Jusczyk Lab Final Report*. Retrieved from http://hincapie.psych.purdue.edu/Jusczyk

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*, 244-247. http://dx.doi.org/10.1016/j.cognition.2009.07.011

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, *80*(3), 675-685. http://dx.doi.org/10.1111/j.1467-8624.2009.01290.x

Saffran, J. R. (2003). Statistical Language Learning: Mechanisms and Constraints. *Current Directions in Psychological Science*, *12*, 110-114. http://dx.doi.org/10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning of 8-Month-Old Infants. *Science*, *274*, 1926-1928. http://dx.doi.org/10.1126/science.274.5294.1926

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. http://dx.doi.org/10.1006/jmla.1996.0032

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*, B25-B34. http://dx.doi.org/10.1016/j.cognition.2005.01.006

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71. http://dx.doi.org/10.1207/s15327078in0701_5

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*, 706–716. http://dx.doi.org/10.1037/0012-1649.39.4.706