Words as anchors: Known words facilitate statistical learning

Toni Cunillera[a,b], Estela Càmara[c,a], Matti Laine[b], Antoni Rodríguez-Fornells[d,a]

[a]Department of Basic Psychology, Faculty of Psychology, University of Barcelona, 08035, Barcelona, Spain

[b]Department of Psychology, Åbo Akademi University, FIN-20500 Åbo, Finland

[c]Department of Neuropsychology, Otto-von Guericke University, 39106, Magdeburg, Germany

[d]Institució Catalana de Recerca i Estudis Avançats (ICREA)

Running head: Words as anchors

Word count (abstract included, ref excluded): 3 999

Correspondence to:

Toni Cunillera

Dept. Psicologia Bàsica
Facultat de Psicologia
Universitat de Barcelona
Passeig de la Vall d'Hebron 171
Barcelona 08035
Tel: +34-93 31258155
fax: +34-93 4021363
e-mail: tcunillera@ub.edu

Abstract

Can even a handful of newly learned words help to find further word candidates in a novel spoken language? The present study shows that the statistical segmentation of words from speech stream by adults is facilitated by the presence of known words in the stream. This facilitatory effect is immediate as the known words were acquired only minutes before the onset of the speech stream. Our results demonstrate an interplay between top-down lexical segmentation and bottom-up statistical learning, in line with infant research suggesting that integration of multiple cues facilitates early language learning. The ability to simultaneously benefit from both types of word segmentation cues appear to be present through adulthood, and it can thus play a role in second language learning.

Introduction

To learn a new language, listeners must first attain a basic vocabulary. This begins with identification of word candidates in the new language through segmentation of the speech stream. This is not a trivial task as speech represents a continuous signal with no clear pauses indicating word boundaries within a sentence. The acoustic signal, however, contains some reliable cues that can help to segment words (e.g., Jusczyk, 1999; Kuhl, 2004).

Among other cues, the distributional properties of speech can help in segmenting the speech stream into words. Words can be detected by computing the transitional probabilities of syllables, a process coined as statistical learning (Saffran, Newport, & Aslin, 1996b; Saffran, Aslin, & Newport, 1996a). High transitional probabilities where the presence of the syllable *X* strongly predicts that the next syllable is *Y*, are most likely within words. In contrast, low transitional probabilities signaling a weak contingency between *X* and *Y* suggest word boundary (Saffran et al., 1996b).

Intuitively, one would expect that when the very first words are learned, further word segmentation could be facilitated by the learned words. This is because they indicate the offset and onset of adjacent words. However, in statistical learning research (Aslin, Saffran, & Newport, 1998; Saffran et al., 1996b; Saffran et al., 1996a), one has not considered how already segmented words aid to isolate the remaining words from the language stream.

The idea that familiar words can play an important role in speech segmentation is not a new one (see e.g., Peters, 1983; Pinker, 1994) but it was directly tested in infants only recently. Bortfeld et al. (2005) exposed 6 month-old infants to a series of short utterances in which a familiar word (infants' own name or mother's name) or an unfamiliar one was followed by a new word (an object name unknown for the child).

Their results proved that infants segmented new words from fluent speech only when they were followed by a familiar name. This study demonstrated that the first words infants recognize become useful segmentation cues, probably acting as anchors that indicate the offset/onset of adjacent word candidates.

While empirical evidence for an interplay between statistical learning and lexical knowledge in speech segmentation is lacking, there is at least one computational model directly relevant to this issue, namely the INCDROP (INCremental Distributional Regularity OPtimization; Brent & Cartwright, 1996; Brent, 1997; Dahan & Brent, 1999). This model emphasizes top-down lexically driven speech segmentation but also takes into account statistical learning (based on word frequency and distributional regularities) in language acquisition. In asserting that even at the earliest stages of language acquisition, experience with words of a language is the main determinant of segmentation, it differs from other proposals (e.g., Cole, Jakimik, & Cooper, 1980; Marslen-Wilson & Welsh, 1978; Mcclelland & Elman, 1986; Norris, 1994). The model predicts that when known words are first recognized in an utterance, the subsequent contiguous string of syllables is immediately inferred as a new word. Familiar words would thus be used as anchors, indicating that the remaining part of the utterance is a potential new wordlike unit.

In the present study, we examined whether known words can act as anchor words that aid adults to segment unknown words in a new language. Participants were exposed to a continuous speech stream of an artificial language that could be parsed into wordlike units only through statistical learning, i.e., by computing transitional probabilities between syllables. Prior to the presentation of the speech stream, participants learned two novel words which did or did not belong to the subsequent speech stream. We hypothesized that these recently acquired words, when recognized in the language stream, will improve speech segmentation.

Experiment 1

Method

Participants

Fifty-six (mean age 20.8 ± 2.23 SD) undergraduate psychology students at the University of Barcelona received extra course credits for their participation in the experiment. Participants were randomly assigned to one of the two experimental conditions (anchor word condition or non-anchor word condition, see below)

Stimuli

The artificial language stream. Forty-eight different consonant-vowel syllables were combined to create two language streams which followed the same structure as those created by Saffran et al. (1996). We decided to use two different language streams to control for possible arbitrary listening preferences. For each stream, eight trisyllabic nonsense words were concatenated to form a nonstop speech stream by using the text-to-speech synthesizer MBROLA with a Spanish male diphone database at 16 kHz (Dutoit, Pagel, Pierret, Bataille, & van der Vreken, 1996). Importantly, words were combined in a way that each word in the stream was followed by each of the other words the same number of times.

The use of this artificial language learning methodology enables us to rule out such potential segmentation cues as word-stress or coarticulation. Thus, all phonemes had the same duration (116 ms) and pitch (200 Hz; equal pitch rise and fall, with pitch maximum at 50% of the phoneme) in the language streams. The only reliable cue for word boundaries was the statistical structure of the language. In all streams the transitional probability of the syllables forming a word was 1.0, while for syllables spanning word boundaries it was 0.14. Each word was repeated 28 times along the stream with the constraint that the same word never occurred twice in a row. The duration of each word was 696 ms, yielding a total stream duration of 2 min 35 s and 904 ms. A written excerpt from the speech stream is as follows: "*demuri/senige/somapo/kotusa/tokuda/piruta/furake/bagoli/senige/tokuda/demuri…* ". Here the three-syllable wordlike units are separated by slashes. In the Anchor word condition, the two words that were taught prior to stream exposure were included in the speech stream.

In addition, eight non-words were created for each language by recombining the syllables of the 8 words composing the stream. Non-words were sequences of three syllables that never formed a string in the language stream (transitional probability = 0). Finally, for the two languages 112 part-words were created by recombining the syllables of the 8 words from each language. Fifty-six part-words were made by concatenating the last two syllables of a word and the first one of another (part-words 2-3-1), and the other fifty-six were made by concatenating the last syllable of a word and the first two syllables of another (part-words 3-1-2).

Word learning phase. The participants were taught two words of the new language by showing pictures together with an auditorily presented narration in Spanish. The presentation lasted for approximately 3 minutes. The synopsis was as follows: "A space traveler stops by an unknown planet looking for water and food. After he lands he decides to move to a nearby city. There he meets a local inhabitant who speaks an unknown strange language. The alien provides the traveler with water and apples and at the same time teaches him the words in his language that refer to water and apple". Each time the two new "alien" words were presented, the female narrator's voice was replaced by the synthesized speech used in the subsequent artificial language stream.

The two novel words were repeated three times during the presentation. Each word was associated either to water or apple. We decided to provide the novel word with an associated meaning to simulate a more natural learning process: the first words learned in a foreign language are usually concrete, familiar and frequent objects.

Procedure

Twenty-eight participants were randomly assigned to the anchor word condition and the other twenty-eight to the non-anchor word condition. In the word learning phase, the participants were instructed to pay attention to the slide show and to learn the new words that would be presented. Immediately after the slide show, they heard the new words separately and were to write down the corresponding meaning (Spanish translation equivalent, i.e., *agua* and *manzana*). Each participant saw the same slide show but with different words so that the word presentation became counterbalanced across participants. The segmentation task began not until the participant had identified the meaning of the new words. They were allowed to write the response up to three times, and when an erroneous response was recorded, the slide show was played. In the anchor word condition, the participants learned two of the eight words composing the novel language. In contrast, in the non-anchor word condition the participants learned two trisyllabic sequences that were not presented in the language stream.

Immediately after successful completion of the word learning phase, the participants were requested to listen carefully to the language stream and to discover the words of the novel language. They were informed that a final test would be presented at the end of the language stream. Importantly, they were not informed about the presence of the two recently learned words in the language stream. For each condition, the participants were randomly assigned to one of the two language streams (Language A or B). The two language streams were counterbalanced across participants in correspondence with the preceding slide show.

Immediately after the language stream, a standard auditory two-alternative-forced-choice (2AFC) test was presented. Test items comprised the eight words of each stream (for the anchor word condition the two previously learned words were included in the set of eight words) and eight part-words randomly selected from the pool of 112 part-words of the same stream (four part-words corresponding to the syllable structure 2-3-1 and four to the syllable structure 3-1-2; see the Stimuli section). Words and part-words were combined so that each word was paired with four different part-words but each of the eight part-words appeared equally often. This procedure rendered a total of 32 pairs that were presented in random order. After hearing each pair of test items, the participants were asked to decide by pressing a button whether the first or the second item of the pair was a word of the new language. Presentation of the items of a pair was separated by a 400 ms pause.

Results and Discussion

For the anchor word condition, only data from participants who successfully selected the two anchor words in the 2AFC test were included in the statistical analysis. A criterion of at least 3 out of 4 correct segmentations for each of the two anchor words was employed. Data from 15 participants were thus removed from the analysis and substituted by data from new participants to retain perfect counterbalancing.

We began comparing the segmentation rates between the participants who listened to the different language streams (Language A and B) for the two experimental conditions. The results revealed no significant stream differences in either the anchor word condition ($t(26) < 1$) or the non-anchor word condition ($t(26) = 1.3$, $p > 0.19$). Consequently, in all subsequent analyses the data was collapsed across the two languages.

For the anchor word condition, the mean percentage of correctly segmented words was 76.8 ± 11.3%, being different from chance [(50%), $t(27) = 12.5$, $p < 0.001$]. For the non-anchor word condition, the mean percentage of correctly segmented words was 63.6 ± 9.8%. This percentage was also different from chance [(50%), $t(27) = 7.3$, $p < 0.001$]. Segmentation performance was significantly better for the anchor word condition than for the non-anchor word condition ($t(54) = 4.66$, $p < 0.001$, effect size: $d = 1.24$).

The analyses reported above compared the two conditions in which streams of eight words were presented. However, in the anchor word condition, two of the words were already segmented because of the word learning phase. In order to provide a percentage of segmentation of the remaining six words, a new analysis was carried out where in the anchor word condition only the truly novel six words were taken into consideration (24 test pair items). The mean percentage of correctly segmented words was 71.5 ± 14.8% (percentage different from chance level (50%), $t(27) = 7.7$, $p < 0.001$). When this percentage was compared to the non-anchor word condition, a significant difference was observed ($t(54) = 2.33$, $p < 0.03$, $d = 0.62$).

These results corroborate the hypothesis that the presence of anchor words facilitated the segmentation of the language stream. However, one might also argue that the difference in segmentation performance between the two conditions is due to interference in the non-anchor word condition rather than facilitation in the real anchor word condition. In order to clarify this, we conducted a new experiment.

Experiment 2

A possible explanation for the significantly lower segmentation rate for the non-anchor word condition in Experiment 1 would be that learned words caused participants to use a detrimental mis-segmentation strategy, as the syllables that composed these words were also present in the language stream. In order to rule out this alternative, we ran an experiment where the learned words were composed of syllables that were not present in the subsequent speech segmentation task.

Method

Participants

Twenty-eight (mean age 20.1 ± 1.42 SD) undergraduate psychology students at the University of Barcelona participated for extra course credits. None of them took part in Experiment 1. Participants were randomly assigned to one of the two language streams (Language A or B).

Stimuli and Procedure

The language streams, words, part-words, the slide show and the whole procedure were the same as in Experiment 1, with the exception of different words being taught in the slide show. For the present experiment, words from Language A were used in the slide show for the Language B, and vice versa. Thus, in contrast with Experiment 1, the learned words consisted of syllables that were not present in the subsequent language stream.

Results and Discussion

No significant differences in segmentation performance were encountered between languages ($t(26) < 1$) and therefore the data were collapsed across the two languages for all subsequent analyses. The mean percentage of correctly segmented words was 63.8 ± 13.9%, being significantly different from chance [(50%), $t(27) = 5.3$, $p < 0.001$]. A comparison of the segmentation results of Experiment 2 and 1 for the non-anchor word condition showed no differences ($t(54) = 0.7$, $p > 0.9$, $d = .02$). This indicates that it was irrelevant for the speech segmentation performance whether or not "non-anchors" consisted of syllables that were present in the language stream.

Moreover, when comparing the non-anchor word condition in this experiment with the anchor word condition in Experiment 1, we observed better segmentation for the anchor word condition when considering either the 8 test words ($t(54) = 3.83$, $p < 0.001$, $d = 1.02$) or the 6 test words ($t(54) = 1.98$, $p = 0.05$, $d = 0.53$).

Experiment 3

Finally, we ran an experiment to compare the segmentation rate in the anchor word condition with a non-anchor condition of a language composed of only 6 words. An intrinsic property of the anchor word condition Experiment 1 was that although the streams consisted of 8 words, only six of them were required to be segmented out, as the participants already knew two words from the learning phase. Consequently, it could be argued that the significantly lower segmentation performance observed for the non-anchor word conditions in Experiments 1 and 2 were due to participants facing a more demanding task (segmenting 8 words) in comparison with the anchor word condition (segmenting 6 words and simply recognizing the other two words).

In order to equate the number of words that needed to be segmented, in the present experiment we reduced the words composing the non-anchor word condition from eight to six. If task difficulty was responsible for the differences reported in the previous experiments, we should observe a better segmentation rate in this new non-anchor word condition than in the previous non-anchor word conditions.

Method

Participants

Twenty-eight (mean age 20.8 ± 2.29 SD) undergraduate psychology students at the University of Barcelona who did not take part in the previous experiments were recruited for the present experiment and received extra course credits for their participation. They were randomly assigned to one of the two language streams (Language A or B).

Stimuli

Two new languages were created by recombining six of the eight words from the previously used languages. Consequently, the stream duration was reduced to 1min 56 s and 928 ms. The structure of the languages was the same as in the previous experiments (see the Stimuli section of Experiment 1). In the two streams the transitional probability of the syllables forming a word was 1.0, while for syllables spanning word boundaries it was 0.2. The number of part-words was reduced to 64 in this experiment. In addition, 6 new non-words for each language were created by recombining the syllables of the 6 words composing the language, yielding 6 syllable sequences with transitional probability equal to zero in the language stream. These non-words were used in the first phase of the experiment as the two to-be-learned words. The slide show and the overall setup were the same as in Experiments 1 and 2.

Procedure

The procedure was the same as in Experiment 1 and 2. The same 2AFC speech segmentation test was administered to the participants as in Experiment 1 and 2 but the number of item pairs was 36 for the present experiment. The six words composing the stream were exhaustively combined with 6 part-words (three part-words corresponding to the syllable structure 2-3-1 and three to the syllable structure 3-1-2) rendering 36 pair items.

Results and Discussion

No differences were observed between the languages ($t(26) < 1$) and thus the data across the two languages were collapsed. The mean percentage of correctly segmented words was 63.1 ± 13.8% (see Figure 2), being different from chance level [(50%), $t(27) = 5.0$, $p < 0.001$]. When comparing the segmentation performance between Experiments 2 and 3 no differences were found ($t(54) < .3$). This indicates that the results from Experiment 1 were not due to a difference in segmentation load. When comparing the non-anchor word condition between experiments 1 and 3, no significant differences were observed either ($t(54) < .2$).

We then compared the present results with those of Experiment 1 and again observed a larger rate of segmented words for the anchor word condition (Exp. 3 vs. Exp. 1 anchor word condition: $t(54) = 4.07$, $p < 0.001$, $d = 1.09$). When the analysis was restricted to 6-test words in the anchor word condition in Experiment 1, the difference still remained significant ($t(54) = 2.19$, $p < 0.04$, $d = 0.58$).

General Discussion

We explored how recently learned words affect statistical learning in a speech segmentation task. The results from Experiment 1 demonstrated that speech segmentation performance was increased when recently learned words were embedded in the language stream. Experiments 2 and 3 showed that the observed advantage was not due to interference caused by miscuing in the control condition or due to the different number of words to be segmented.

The present findings suggest that the very first learned words help to isolate and discover novel words of a new language. Thus the first learned words appear to aid the underlying statistical learning process when segmenting new words. Our results indicate that lexically driven segmentation, as proposed by the INCDROP model (Dahan & Brent, 1999), can work in concert with computation of transitional probabilities of syllables. The present results thus reflect the interplay between a top-down process (lexical segmentation) and a bottom-up process (computation of transitional probabilities). Bortfeld and colleagues (2005) suggested a similar process to explain how 6-month-old infants succeeded in segmenting out new words from utterances after recognizing familiar words in them. However, an important difference with our study is that the familiar names used by Bortfeld et al. (“mommy/mama” or the infants’ name) were probably well consolidated in their infants’ memory, as they heard these words every day. Our participants were able to use learned words although their experience with these words was minimal, demonstrating that lexical items can contribute to speech segmentation immediately after their acquisition.

The present results show that adult listeners can combine statistical learning with other segmentation cues available in speech. Infant research has suggested that integration of multimodal cues facilitates language learning (Bahrick & Lickliter, 2000; Hollich et al., 2000; Hollich, Newman, & Jusczyk, 2005). A recent speech segmentation study also found a positive effect of combining intrasensory statistical regularities in speech and music (Schön et al., in press). Therefore, it is plausible that the coalition of multiple cues, as far as they do not collide (see e.g., Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003), can facilitate speech segmentation. The cue-specific weights in a multi-cue context during second language acquisition are not yet clear (but see Christiansen, Allen, & Seidenberg, 1998).

Another critical issue concerns the use of top-down lexical segmentation and bottom-up computation of transitional probabilities at different ages. Our data suggests that both of these mechanisms remain active after childhood (see Braine et al., 1990; Gillette, Gleitman, Gleitman, & Lederer, 1999). In line with this, statistical learning has been demonstrated in both infants and adults when learning an artificial mini-language (Saffran et al., 1996b; Saffran et al., 1996a). Likewise, it appears that infants benefit from isolated and familiar words at the initial stages of language comprehension (Bortfeld et al., 2005; Mandel et al., 1995) and at the beginning of their vocabulary expansion (Brent & Siskind, 2001), and such an effect is present also in adults with their initial contact with a new language (Dahan & Brent, 1999).

Further evidence for similarities of adults' and infants' language learning systems comes from a word learning experiment where adults were exposed to infant-directed speech (Golinkoff & Alioto, 1995). English-speaking adults were exposed to short sentences spoken in Chinese while watching pictures corresponding to target object names embedded in the sentences. One group heard sentences pronounced in infant-directed speech, whereas the other group heard sentences pronounced in adult-directed speech. Only those exposed to infant-directed speech could segment the target words.

However, it is impossible to say as to which cue or cues contributed most to speech segmentation, as infant-directed speech has many characteristic features (slower speech rate, extended frequency range, higher fundamental frequency, repeated pitch contours, marked intensity shifts, longer pauses, simplified vocabulary, vowel lengthening; Hoff-Ginsberg & Shatz, 1982). Interestingly, some properties of infant-directed speech are observable in "foreigner talk", i.e., in native speakers interacting with non-natives (Snow, Vaneeden, & Muysken, 1981).

It thus seems plausible that when infants and adults are exposed to a new language, they both would rely on the same top-down and bottom-up strategies to isolate new words. In fact, Bortfeld et al. (Bortfeld et al., 2005) argued that there is no reason to believe that infants cannot use top-down lexical strategies for segmenting speech. While both strategies appear to be in use throughout the life span, further studies are needed to clarify the relative weight of these strategies in children vs. adults.

In summary, we show that very recently acquired words facilitate word segmentation in a new language when the learned words appear in the speech stream. This indicates a possible interplay between lexical top-down processing and bottom-up segmentation based on transitional probabilities of syllables. More generally, our results highlight the employment of multiple cues in vocabulary acquisition.

Acknowledgements

This project was supported by the Academy of Finland NEURO research program grant to ML, the Spanish Government grant to ARF (SEJ2005-06067/PSIC) and the Generalitat de Catalunya pre-doctoral stage grant to TC. We thank Irene Nogué for helping in data collection.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321-324.

Bahrick, L. E. & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36,* 190-201.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and Me - Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16,* 298-304.

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L. et al. (1990). Exploring language-acquisition in children with a miniature artificial language - Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language, 29,* 591-610.

Brent, M. R. (1997). Toward a unified model of lexical acquisition and lexical access. *Journal of Psycholinguistic Research, 26,* 363-375.

Brent, M. R. & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93-125.

Brent, M. R. & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* B33-B44.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13,* 221-268.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1980). Segmenting speech into words. *Journal of the Acoustical Society of America, 67,* 1323-1332.

Dahan, D. & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology-General, 128,* 165-185.

Dutoit, T., Pagel, N., Pierret, F., Bataille, O., & van der Vreken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In (pp. 1393-1396). **Philadelphia**.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135-176.

Golinkoff, R. M. & Alioto, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition. *Journal of Child Language, 22,* 703-726.

Hoff-Ginsberg, E. & Shatz, M. (1982). Linguistic input and the childs acquisition of language. *Psychological Bulletin, 92,* 3-26.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L. et al. (2000). Breaking the language barrier: an emergentist coalition model for the origins of word learning. *Monogr Soc.Res.Child Dev., 65,* i-123.

Hollich, G. J., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development, 76,* 598-613.

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.

Jusczyk, P. W. (1999). How infants begin to extract words From speech. *Trends in Cognitive Sciences, 3,* 323-328.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831-843.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants recognition of the sound patterns of their own names. *Psychological Science, 6,* 314-317.

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29-63.

Mcclelland, J. L. & Elman, J. L. (1986). The trace model of speech-perception. *Cognitive Psychology, 18,* 1-86.

Norris, D. (1994). Shortlist - A connectionist model of continuous speech recognition. *Cognition, 52,* 189-234.

Peters, A. (1983). *The units of language acquisition*. New York: Cambridge University Press.

Pinker, S. (1994). *Language learnability and language development*. Cambrige, MA: Harvard University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606-621.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. Songs as an aid for language acquisition. *Cognition, In Press, Corrected Proof.*

Snow, C. E., Vaneeden, R., & Muysken, P. (1981). The interactional origins of foreigner talk - Municipal employees and foreign-workers. *International Journal of the Sociology of Language,* 81-91.

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.

Figure captions

Figure 1. Mean percentage (± s.e.) of correctly segmented words (wds) in the auditory 2AFC test performed at the end of Experiment 1, 2 and 3.