Words as Anchors

Known Words Facilitate Statistical Learning

Toni Cunillera,[1,2] Estela Càmara,[3,4] Matti Laine,[2]
and Antoni Rodríguez-Fornells[4,5]

[1]Department of Basic Psychology, Faculty of Psychology, University of Barcelona, Barcelona, Spain
[2]Department of Psychology, Åbo Akademi University, Åbo, Finland
[3]Department of Neuropsychology, Otto-von Guericke University, Magdeburg, Germany
[4]Department of Physiology, Faculty of Medicine, Campus de Bellvitge – IDIBELL, University of Barcelona, Barcelona, Spain
[5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract. Can even a handful of newly learned words help to find further word candidates in a novel spoken language? This 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 appears to be present through adulthood and can thus contribute to second language learning.

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 difficulty of the segmentation task can also be highlighted by the comparison of spoken and written language: in the latter case, blank spaces clearly mark word boundaries. The acoustic signal, however, contains some reliable cues that can help to segment words in spoken input (e.g., Jusczyk, 1999; Kuhl, 2004).

Among other cues, the distributional properties of speech can be exploited for segmenting the speech stream into words. For example, words can be detected by computing the transitional probabilities of syllables, a process that is part of a more general learning mechanism coined as statistical learning (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). Statistical learning appears to be a domain-general mechanism (but see Conway & Christiansen, 2006) documented in different sensory modalities such as audition, vision, and touch (Conway & Christiansen, 2006; Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002). It has been utilized in a diverse set of learning paradigms in which statistical regularities can be exploited, such as in learning an artificial syntax (Gomez & Gerken, 1999) and segmenting out a tone or a speech sequence (Abla, Katahira, & Okanoya, 2008; Saffran, Aslin, et al., 1996; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, et al., 1996). Concerning word segmentation, the statistical learning view posits that words can be detected by computing the likelihood of appearance of syllables along the speech sequence (Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996). Here high transitional probabilities would indicate that the presence of the syllable *X* strongly predicts the occurrence of the syllable *Y*, and it is most likely to happen within a word than between words (i.e., *XY* could be a word or a part of it) because the pairs of contiguous syllables within a word are constrained by the lexicon and the phonotactics of the language. In contrast, low transitional probabilities (signaling a weak contingency between *X* and *Y*) are indicative of a word boundary.

Intuitively, one could assume that an interaction between known words and words to be learned may facilitate speech segmentation through statistical learning. In other words, one would expect that when the very first words are learned, further word segmentation could be facilitated by these words. It is a common experience that when traveling abroad, people may know a few words of the language of the foreign country. One may recognize such words when listening to a conversation between native speakers, although most of the speech message may sound as an uninterrupted stream of nonsense speech. But even in that situation new words may be detected when being adjacent to a known word that provides a boundary for segmentation. Thus far, the statistical learning approach (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996) has not considered as to how recently segmented words (based on the output of statistical learning)

Experimental Psychology 2010
DOI: 10.1027/1618-3169/a000017

could aid to isolate the remaining words from the speech stream.

The idea that isolated or familiar words can play a significant role in speech segmentation is not new (see, e.g., Peters, 1983; Pinker, 1994), although this hypothesis has been evaluated only in infant language learning. For example, Brent and Siskind (2001) explored the role of isolated words in the formation of infants' vocabulary. These authors demonstrated that isolated words are abundant in infant-directed speech. They found that about 30–50 words used by infants in their study (∼ 44% of the production from infants recorded at the age between 9 and 15 months) were words spoken to the infants in isolation by their mothers before infants used them. The authors suggested that initial words that compose the small vocabulary of infants may provide reliable cues that help infants with their vocabulary expansion observed during the second year of life (e.g., Bates, Bretherton, & Snyder, 1988; Fenson et al., 1994). Thus, new words may be segmented from multiwords utterances by recognizing adjacent known words. This strategy may be adopted by children with the first words that they master, and could happen early in development as Mandel, Jusczyk, and Pisoni (1995) demonstrated that infants recognize their own names already at the age of 4.5 months.

In another study, Bortfeld, Morgan, Golinkoff, and Rathbun (2005) exposed 6-month-old infants to a series of short utterances in which a familiar word (infant's own name or mother's name) or an unfamiliar one was followed by a new word (an object name unknown for the child). The results of this experiment proved that infants segmented new words from fluent speech only when new words were followed by a familiar name. This study demonstrated that the first words infants recognize become useful segmentation cues, probably acting as anchor points that indicate which sound sequence next in the sentence is a wordlike unit.

Segmenting fluent speech into words is a very different task in an already mastered language and in a new language to be learned. For a language already mastered a variety of lexical cues can be used, whereas for a new language only acoustical, perceptual, and distributional cues are available in the speech signal. 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, 1997; Brent & Cartwright, 1996; Dahan & Brent, 1999). Its main assumption is that speech segmentation might rely on the experience acquired with the words of a language. Another important feature of the model is that it brings together two properties of language at the earliest stages of language acquisition and lexicon formation, namely the distributional regularities and the knowledge of familiar words. As posited by Brent (1997), it was proposed as a top-down lexically driven segmentation model that is able to discover new words. It is worth noting that the INCDROP model differs from other proposals in asserting that experience with words of a language is the main determinant of segmentation at the earliest stages of language acquisition (e.g., Cole, Jakimik, & Cooper, 1980; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994). Statistical learning (based on word frequency and distributional regularities) is relegated to a secondary tier in the model. 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. It seems, therefore, that the inference of a new word is most likely to occur when it appears between a familiar word and a phrasal pause (see Dahan & Brent, 1999). Hence, the model predicts less accurate segmenting in utterances where the familiar word is found in the middle of the utterance.

Statistical learning (Saffran, Aslin, et al., 1996) predicts that word boundaries are expected when low transitional probabilities are detected in the speech sequence, irrespective whether they appear in the middle or at the edge of a sentence. However, as we mentioned above, it does not take into account the role of recently isolated words during the online segmentation process (as the INCDROP model does) or the role of acoustical cues as word-stress or allophonic variations. This apparently relates to the scope of the original statistical learning studies (to test the role of a single specific cue in segmentation) rather than to a theoretical stance. For example, it is interesting to note that Dahan and Brent (1999, p. 183) suggested that "transitional probabilities might work together with lexically driven segmentation". Although this hypothesis is plausible, no empirical evidence has been provided as to how these two processes, segmentation based on known words (anchors) and statistical learning (transitional probabilities), might interact with each other.

In this study, we examined whether familiar words can act as anchor words that aid adults to segment unknown words in a new language. In other words, we studied whether the presence of familiar words in a speech segmentation task facilitates statistical learning of novel words. Participants were exposed to a continuous speech stream of an artificial language that could be parsed into wordlike units only through statistical learning, that is, by computing transitional probabilities between syllables. Prior to the presentation of the speech stream, participants learned two novel words (anchor 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, would 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 nonanchor 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, Aslin, et al. (1996); Saffran, Newport, 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 nonwords were created for each language by recombining the syllables of the eight words composing the stream. Nonwords 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 eight 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 56 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 ~ 3 min. 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 28 to the nonanchor 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 asked 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 the fourth erroneous response was recorded, the slide show was replayed[1]. In the anchor word condition, the participants learned two of the eight words composing the novel language. In contrast, in the nonanchor 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

---

[1] We had to repeat the presentation of the slide show only in a very small number of cases.

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

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 or the nonanchor word condition (in both cases $t(26) < 1$). Consequently, in all subsequent analyses the data were collapsed across the two languages.

For the anchor word condition, the mean percentage of correctly segmented six novel words (excluding the two anchor words that were taught prior to the segmentation task) was 71.4 ± 14.8% (percentage different from chance level (50%), $t(27) = 7.7$, $p < .001$). For the nonanchor word condition, the mean percentage of correctly segmented eight novel words was 63.6 ± 9.8%. This percentage was also different from chance (50%), $t(27) = 7.3$, $p < .001$. The segmentation performance was significantly better for the anchor word condition than for the nonanchor word condition ($t(54) = 2.33$, $p = .02$, $d = .63$). When including the anchor words in the analysis, the difference in segmentation performance between the anchor word condition and the nonanchor word conditions remained essentially the same ($t(54) = 2.49$, $p = .01$, effect size: $d = .66$).

Even though the brief training phase ensured that all participants in the anchor word condition had learned to associate the isolated anchor words to their corresponding meanings, it was of interest to explore as to what extent the anchor words were explicitly segmented in the speech stream. For each participant, the criterion of at least three out of four correct responses in the test phase for each of the two anchor words was employed. Almost half of the participants (13/28) failed to reach this criterion. In order to examine whether or not their lack of explicit segmentation of the anchor words affected the overall segmentation performance of the study group, data from these 13 participants were substituted by data collected from new participants who fulfilled this criterion. The pattern of results remained very similar to the one reported above for the original anchor condition group: the mean percentage of correctly segmented words in the anchor condition was 71.5 ± 14.8%, a percentage that was significantly different from chance (50%), $p < .001$. A comparison of the segmentation results across conditions revealed significant differences (anchor words vs. nonanchor words: $t(54) = 2.33$, $p < .03$, $d = .62$).

The present results corroborate the hypothesis that the presence of anchor words facilitates the segmentation of novel words in a language stream. They also indicate that an explicit segmentation of the anchor words themselves is not necessary for this facilitation to occur. However, there are two potential confounds that need to be ruled out. First, it might be that the difference in segmentation performance between the anchor versus the nonanchor condition is due to interference created by the nonanchor words rather than facilitation by the anchor words in their corresponding conditions. Second, the overall segmentation load may have favored the anchor word condition. The participants in the nonanchor word condition had to segment eight totally novel words, whereas in the anchor word condition, two of the eight novel words had been shown in the training phase. In order to clarify these issues, we conducted two new experiments.

Experiment 2

A possible explanation for the significantly lower segmentation rate for the nonanchor word condition in Experiment 1 would be that recently learned words caused participants to use a detrimental mis-segmentation strategy, as the syllables that composed of these words were also present in the subsequent 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%), $t(27) = 5.3$, $p < .001$. A comparison of the segmentation results of Experiment 2 and the nonanchor word condition of Experiment 1 showed no differences ($t(54) = 0.7$, $p > .9$, $d = .02$). This indicates that it was irrelevant for the speech segmentation performance whether or not “nonanchors” consisted of syllables that were present in the language stream.

Moreover, when comparing the nonanchor word condition in this experiment with the anchor word condition in Experiment 1, we observed better segmentation for the anchor word condition ($t(54) = 1.98$, $p = .05$, $d = .53$).

Experiment 3

We ran another experiment in order to compare the segmentation rate in the anchor word condition with a nonanchor condition of a language composed of only six words. An intrinsic property of the anchor word condition in Experiment 1 was that although the streams consisted of eight words, only six of them were totally novel for the participants. Consequently, it could be argued that the significantly lower segmentation performances observed for the nonanchor word conditions in Experiments 1 and 2 were due to participants facing a more demanding task (segmenting eight words) in comparison with the anchor word condition (segmenting six words and recognizing the other two words either explicitly or implicitly).

In order to equate the number of words that needed to be segmented, in the present experiment we reduced the words composing the nonanchor 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 nonanchor word condition than in the previous nonanchor word conditions.

Method

Participants

Another 28 (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 1 min 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, six new nonwords for each language were created by recombining the syllables of the six words composing the language, yielding six syllable sequences with transitional probability equal to zero in the language stream. The two to-be-learned words used in the first phase of the experiment were taken from these nonwords. The slide show and the overall setup were the same as in Experiments 1 and 2.

Procedure

The procedure was the same as in Experiments 1 and 2. The same 2AFC speech segmentation test was administered to the participants as in Experiments 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 six 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 two languages ($t(26) < 1$) and thus the data were collapsed. The mean percentage of correctly segmented words was 63.1 ± 13.8% (see Figure 1), being different from chance level (50%), $t(27) = 5.0$, $p < .001$. When comparing the segmentation performance between Experiments 2 and 3, no difference was found ($t(54) < .3$). This indicates that anchor word facilitation effect observed in Experiment 1 was not due to a difference in segmentation load between the anchor versus nonanchor condition. When comparing the nonanchor word conditions between Experiments 1 and 3, no significant difference was observed either ($t(54) < .2$).

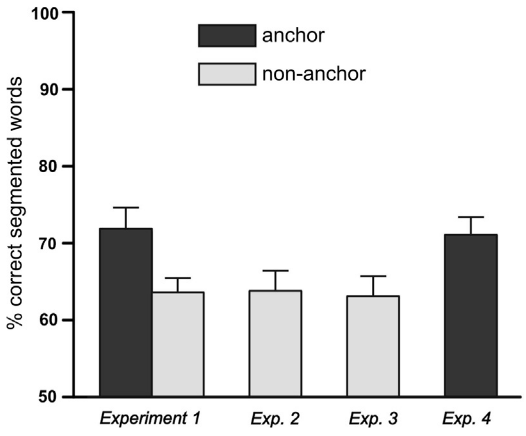

Figure 1. Mean percentage (±*SE*) of correctly segmented words in the auditory 2AFC test performed at the end of Experiments 1–4.

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: $t(54) = 2.19$, $p < .04$, $d = .58$).

In summary, the present findings help to rule out the possibility that the lower segmentation performance in the nonanchor condition in Experiment 1 was due to a higher overall segmentation load as compared to the anchor word condition.

Experiment 4

Finally, we wanted to ensure that the facilitatory effect of anchor words on speech segmentation observed in Experiment 1 is reliable enough to be replicated. The present experiment thus attempted to replicate the anchor word condition in Experiment 1 with an identical setup except for some more training for the anchor words in the learning phase.

Method

Participants

An additional set of 28 (mean age 20.3 ± 2.48 *SD*) undergraduate psychology students at the University of Barcelona took part in the 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

All the stimuli (language streams, words, part-words, and the slide show) were the same as in Experiment 1.

Procedure

The whole procedure was the same as in Experiment 1, with the exception of a modified training setup used for the *word learning phase*. For the present experiment, we doubled the number of repetitions of the "to-be-learned words" in the learning phase. Each anchor word appeared thus six times in the slide show (the last three times simultaneously in spoken and written form), as compared with the three auditory exposures used in Experiment 1.

Results and Discussion

With this new training setup, the results were similar to those reported in Experiment 1 (70.1 ± 12.6%, a percentage significantly above chance levels (50%), $p < .001$). Importantly, when comparing the performance in this new experiment with the nonanchor conditions in Experiments 1 and 3, we observed the same advantage for the anchor condition (Exp. 1 nonanchor condition vs. Exp. 4: $t(54) = 2.14$, $p < .04$, $d = .57$; Exp. 3 vs. Exp. 4: $t(54) = 1.98$, $p = .053$, $d = .53$).

As in Experiment 1, we also explored how well the anchor words were explicitly segmented in the speech stream. Given the added anchor word training in the present experiment, the number of participants failing to explicitly segment the anchor words was expected to be lower. Indeed, only 5 of the 28 participants failed to fulfill the criterion of at least three out of four correct responses in the test phase for each of the two anchor words. When their results were replaced by data from five new participants who fulfilled this criterion, the results remained again essentially the same mean percentage of correctly segmented words: 71.1 ± 12.1%, *t* tests against chance (50%) $p < .001$; Exp. 1 (nonanchor condition) versus Exp. 4: $t(54) = 2.55$, $p < .02$, $d = .68$; Exp. 3 versus Exp. 4: $t(54) = 2.32$, $p < .03$, $d = .62$).

In sum, the present results replicate those obtained for the anchor word condition in Experiment 1. This gives further support to the hypothesis that recently learned words can facilitate speech segmentation.

General Discussion

In this study we explored how recently learned words affect statistical learning in a speech segmentation task. The results from Experiments 1 and 4 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. Additional analyses performed in Experiment 1 indicated that this facilitatory effect is presented irrespective whether all subjects explicitly segmented the anchor words or not. In other words, even for subjects who did not anymore consciously recognize the anchor words, these words still appeared to boost segmentation performance.

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 between syllables. There is, however, an alternative explanation on how speech segmentation is achieved by distributional cues in the speech input, and it has been successfully implemented in a computer model called PARSER (Perruchet & Vinter, 1998). Rather than computing transitional probabilities, PARSER is based on the formation of chunks (Anderson & Lebiere, 1998) positing it as the core principle for statistical learning in speech segmentation (see Perruchet & Pacton, 2006 for an interesting discussion of this controversy). The model makes the strong claim that for segmenting speech there is no need for computations. Instead, chunks are formed and shaped over time following the laws governing associative memory and the attentional capacity constraints that limit the processing of incoming information. In the same vein, Pacton and Perruchet (2008) have recently proposed a general associative learning model that asserts attention as the necessary and sufficient condition for associative learning and language chunking to occur. However, the interplay of different segmentation cues has not yet been implemented in the PARSER.

The present results thus reflect an interplay between a top-down process (lexical segmentation) and a bottom-up process (computation of transitional probabilities). Bortfeld et al. (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 benefit from the learned words although their experience with these words was minimal and even when not all of them could explicitly recognize the anchor words anymore at the final segmentation task. This demonstrates that lexical items can contribute to speech segmentation immediately after they are learned.

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, Newman, & Jusczyk, 2005; Hollich et al., 2000). A recent speech segmentation study also found a positive effect of combining intrasensory statistical regularities in speech and music (Schön et al., 2008). 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 suggest 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, Aslin, et al., 1996; Saffran, Newport, et al., 1996). 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. It is important to note that some of the properties of the infant-direct speech are found in a variety of languages like English, Italian, French, German, Japanese, and Chinese (Fernald et al., 1989). 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, and vowel lengthening; Hoff-Ginsberg & Shatz, 1982). Interestingly, some properties of infant-directed speech are observable in "foreigner talk", that is, in native speakers interacting with nonnatives (Snow, Vaneeden, & Muysken, 1981).

It thus seems plausible that when infants and adults are exposed to a new language, they both rely on the same top-down and bottom-up strategies to isolate new words. In fact 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 versus 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.

Acknowledgments

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 predoctoral stage grant to TC. We thank Irene Nogué for helping in data collection and Salvador Soto-Faraco for his comments on the earlier drafts of this paper.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*, 952–964.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Bates, E., Bretherton, I., & Snyder, L. S. (1988). *From first words to grammar*. New York: Cambridge University Press.

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.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science, 17*, 905–912.

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 *Conference Proceeding ICSLP'96, Philadelphia* (pp. 1393–1396).

Fenson, L., et al. (1994). Variability in early communicative development. In *Monographs of the Society of Research in Child Development, 59* (5, Serial No. 242).

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. A. (1989). Cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477–501.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 28*, 458–467.

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.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.

Hoff-Ginsberg, E., & Shatz, M. (1982). Linguistic input and the child's 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. *Monographs of the Society for Research in Child Development, 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.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35–B42.

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.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34*, 80–96.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233–238.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*, 246–263.

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

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

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

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.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition, 106*, 975–983.

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.

Received January 18, 2009
Revision received March 9, 2009
Accepted March 16, 2009

Toni Cunillera

Dept. Psicologia Bàsica
Facultat de Psicologia
Universitat de Barcelona
Passeig de la Vall d'Hebron
171 Barcelona 08035
Spain
Tel. +34 93 3125144
Fax +34 93 4021363
E-mail tcunillera@ub.edu