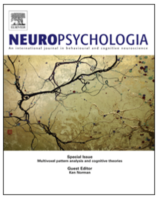
Neuropsychologia 82 (2016) 189–199

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Headstart for speech segmentation: a neural signature for the anchor word effect

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

[a] Department of Basic Psychology, Faculty of Psychology, University of Barcelona, 08035 Barcelona, Spain
[b] Department of Psychology, Abo Akademi University, FI-20500 Åbo, Finland
[c] Cognition and Brain Plasticity Group (Bellvitge Biomedical Research Institute) IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain
[d] Catalan Institution for Research and Advanced Studies, ICREA, Barcelona, Spain

ARTICLE INFO

Article history:
Received 2 June 2015
Received in revised form
7 January 2016
Accepted 10 January 2016
Available online 11 January 2016

Keywords:
Statistical language learning
Predictability in learning
ERPs
Speech segmentation
Lexical segmentation

ABSTRACT

Learning a new language is an incremental process that builds upon previously acquired information. To shed light on the mechanisms of this incremental process, we studied the on-line neurophysiological correlates of the so-called anchor word effect where newly learned words facilitate segmentation of novel words from continuous speech. Higher segmentation performance was observed for speech streams embedded with newly learned anchor words. The anchor words elicited an enhanced Stimulus-preceding negativity (SPN) component considered to be an index of expectation for incoming relevant information. Moreover, we confirmed a previously reported N400 amplitude increase for the to-be-segmented novel words, indicating a bottom-up learning process whereby new memory representations for the novel words emerge. We propose that the anchor word effect indexed by SPN reflects an expectation for an incoming novel word at the offset of the anchor word, thus facilitating the segmentation process.

1. Introduction

Mastering a foreign language is a lengthy and gradual process. The initial challenge facing the language learner is to identify words in running speech. The difficulty of this task lies in the fact that speech input does not provide reliable cues such as silent gaps between words that would indicate word boundaries. Nevertheless, some initial words are acquired by a beginning language learner, e.g. through their repeated use or their appearance in isolation. In turn, these familiar words can serve as "anchors" that facilitate language acquisition by helping to parse the new language further. For example, recognizing the word "*kiitos*" ("thanks") in Finnish could aid the language learner to segment the word "*ei*" ("no") in a very common utterance "*kiitos*ei" ("no thanks"), as the recognition of the first word reveals the onset of the second word. In the current study, we examined the brain signatures of this "anchor word" effect by measuring event-related brain potentials (ERPs) while young adults took part in a speech segmentation task where they tried to identify words of a new language in a continuous speech stream.

When one is exposed to a new language, novel words might be easily acquired after hearing them in isolation, which obviates the need for segmentation (i.e. lexical segmentation process, Marslen-Wilson and Welsh, 1978; McClelland and Elman, 1986; Norris, 1994). Following this idea, Brent and coauthors (Brent, 1997; Brent and Cartwright, 1996) proposed that if a particular number of words are heard in isolation and are successfully retained, those words then will allow the language learner to bootstrap a lexical segmentation device. This specific approach proposes a simple mechanism in which novel words are discovered by recognizing and extracting familiar words from an utterance and then treating the remaining contiguous phonemic string of the utterance as a candidate novel word.

Experimental evidence for the anchor word effect in infants was first provided by Bortfeld et al. (2005). The authors demonstrated that infants are able to use their first learned words to indicate the onset of a subsequent novel word. In their experiment, 6 month-old infants were exposed to a series of short utterances in which a familiar word (the infants' own name or the word "Mom") or an unfamiliar word was followed by a new object name unknown to the infant. The results proved that infants recognized only those new words that followed a familiar name, thus the presence of the familiar word helped them in segmenting a new word candidate. Dahan and Brent (1999), studying adults learning an artificial language, reported results that were in line with Bortfeld et al. (2005). In their study, participants were exposed to two kinds of nonsense utterances, short ones (2–3 syllables) and long ones (5 syllables), and the short utterances were embedded in the long ones. After a familiarization phase with the short and long syllabic

---

[*] Corresponding author.
E-mail addresses: tcunillera@ub.edu (T. Cunillera), matti.laine@abo.fi (M. Laine), arfornells@gmail.com (A. Rodríguez-Fornells).

http://dx.doi.org/10.1016/j.neuropsychologia.2016.01.011

strings (novel nonsense words), the participants faced a test that examined whether they considered new sequences of syllables (the non-studied fragment from the long utterance) as new potential words. The results showed that adults treated as new words those fragments which were part of the longer studied utterances, but only when the fragments were located at the edge of the utterance. The results of these two studies demonstrate that it is possible to use recently learned words as "anchors" to discover new adjacent words.

Importantly, at the word level, Conway et al. (2010) observed that adults' implicit (statistical) learning abilities correlated with their ability to predict an incoming word in a sentence context under demanding perceptual conditions. Thus, participants who were good at learning a grammar – which simply specified the probability, given a particular element, of the next element to occur – were also good in predicting the last word of high cloze-probability sentences that were acoustically degraded. Other cognitive variables that have been associated with linguistic knowledge (e.g., working memory capacity and intelligence) did not account for this positive correlation. The authors concluded that superior implicit learning abilities lead to more robust long-term memory representations for word order probabilities, which in turn show up as improved top-down processing that helps to implicitly predict incoming words in sentence processing. Following these ideas, in the current study, we measured event-related brain potentials (ERPs) while participants listened to a set of new languages in which, for each language, unknown words were mixed with some recently learned words in a continuous stream. Thus, we pursued to investigate whether the appearance of a known word could aid segmenting the other words in a language stream by predicting the appearance of an incoming new word. More specifically, participants were exposed to several streams and were asked each time to find the words that composed the new languages. Prior to each exposition phase, participants learned two new words that were or were not part of the language stream to be segmented. Based on earlier behavioral evidence with a similar setup (Cunillera et al. 2010a), we expected to observe an improvement in the capacity to segment new words in those streams where the two recently learned words (the anchor words) appeared.

Our expectation concerning the ERP signature of the anchor word effect is based on the assumption that this benefit observed in speech segmentation (Cunillera et al., 2010a) represents a top-down lexical segmentation process, where previously acquired words that appear in the stream help to reveal the onset of a subsequent novel word by mobilizing attentional resources. The idea of the crucial role of attention in language processing is highlighted in the dynamic attending theory (DAT; Large and Jones, 1999) and more specifically for statistical language learning in De Diego-Balaguer et al. (2007) and in the computational model PARSER (Perruchet and Vinter, 1998). The DAT states that when listening to a linguistic sequence, attention may be distributed with attentional cycles, with external cues (e.g., word stress) acting as a key attentional-temporal factor allowing for the creation of temporal expectations about the incoming of significant information and, consequently, facilitating auditory sequencing. Likewise, PARSER is built on the assumption that statistical learning results from the interaction between the placement of the attentional focus, general principles of memory, and the structural regularities of the linguistic input. Moreover, a recent study by De Diego-Balaguer et al. (2015, 2007) indicated that the presence of subtle prosodic pauses in the auditory stream modulated the ERP (N100/P200) components involved in word segmentation and rule learning. An N100 enhancement was encountered when participants needed to allocate their attention to syllable onsets to be able to segment continuous language streams. When subtle prosodic pauses were introduced between novel words, segmentation was not required and a clear attenuation of the N100 component for the onset syllable was observed. Similarly, Sanders et al. (2002) showed that with continuous non-segmented streams, high-performing participants displayed greater N100 amplitudes for the first syllable of the word. This increase in the N100 marker for selective attention to word onset during speech segmentation is consistent with studies indicating that listeners need to dynamically allocate their attention in time to manage crucial information from speech input during language comprehension (Astheimer and Sanders, 2009). Thus, in line with these two theories we hypothesized that if an anchor word acts as an anticipatory cue signaling the onset of a new word candidate, a redirection of attention to the immediately following novel word should take place. However, if the attentional modulation of the N100 is related to the predictability of an incoming word, we expected that the amplitude of the N100 elicited by the first syllable of a word will not be affected by the condition or word type, as in the current study the appearance of the different words along the stream was completely unpredictable.

A plausible ERP correlate for this effect would be the Stimulus-preceding negativity (SPN). This component is progressively built up in time as a slow negative long-lasting potential with a typical frontal distribution prior the presentation of relevant and expected information (Donkers et al., 2005; Morís et al., 2013; van Boxtel and Böckers, 2004). Therefore, we predicted that a modulation of the SPN would be observed in the anchor words reflecting the neural signature of the expectation of an incoming new word.

Although several studies have investigated on-line speech segmentation processes using ERP measures (Buiatti et al., 2009; Cunillera et al., 2006, 2008, 2009; De Diego-Balaguer et al., 2007; Sanders et al., 2002), none of these studies have addressed the question of how recently learned words could be used as attentional top-down signals that allows the learner to build up a temporal prediction cue of the exact moment in which the incoming new-word will appear. Previous ERP studies have so far converged on the identification of a negative N400-like frontal ERP component associated to speech segmentation (Abla et al., 2008; Buiatti et al., 2009; Cunillera et al., 2006, 2009). Similarly, word-learning studies with infants have shown a frontal negative deflection associated to the fast learning of object-word mappings (Friedrich and Friederici 2008; Mills et al., 2005) or when comparing known words against unknown words (Conboy and Mills, 2006) (for a review, see Rodríguez-Fornells et al., 2009). Interestingly, Abla et al. (2008) found that the amplitude of the N400-like component for triplets of tones decreased after an early period of exposition, but only in the group of high learners, indicating that the decline of this component was related to the success in learning the statistical structure of the tone stream. Considering all these findings, we expected to observe an amplitude decrease of the N400-like component for the anchor condition in comparison with the non-anchor one, as learning would be easier when anchor words are present in the continuous language stream.

2. Materials and methods

2.1. Participants

Twenty-three undergraduate psychology students at the University of Barcelona participated in the experiment. All participants were right-handed, reported no hearing deficits or language learning impairment, and were paid for their participation in the experiment. The experiment was approved by the local ethics committee, and written informed consent was obtained from each subject prior to the experiment. Data from 5 participants were discarded due to excessive EEG artifacts, leaving 18 participants for the analysis [mean age 19.9 ± 2.2 (SD)].

2.2. Stimuli

2.2.1. The artificial language streams

The stimuli were synthesized with the MBROLI tool of the MBROLA text-to-speech synthesizer, using a Spanish male diphone database at 16 kHz (Dutoit et al., 1996). Sixty different consonant–vowel syllables were combined to build up nine artificial language streams (see Appendix A1). All syllables had the same duration (232 ms, 116 each phoneme) and fundamental frequency (200 Hz; equal pitch rise and fall, with pitch maximum at 50% of the phoneme). All nine language streams were composed by combining 24 syllables, although depending on the condition, this combination differed. Thus, for the three language streams in the anchor and non-anchor conditions, eight trisyllabic nonsense words, each with a duration of 696 ms (hereafter called "words"), were combined, whereas in the remaining three streams (random condition), the 24 syllables were randomly organized. The idea for these random streams was to create a baseline condition in which participants could not extract or segment any word.

Across languages, each syllable – from the initial pool of 64 syllables – was used two or three times (see Appendix A1 and A2). Subsequently, words of each language were concatenated to form a nonstop speech stream in a way that each word in the stream was followed by each of the other words (never the same word) the same number of times along the stream, ensuring that the appearance of all words were equally unpredictable. A written excerpt from the 8-word speech stream is as follows: "demuri / senige / somepo / kotusa / tokuda / piruta / furake / bagoli / senige / tokuda…", with slashes denoting word boundaries. In order to control for possible perceptual biases, we increased the variability of word order within a stream by concatenating words in each stream in four different ways (see Appendix A3). Furthermore, to avoid introducing a perceivable word onset/offset cue in the stream-initial and stream-final positions, three syllables belonging to the language (the middle syllable of 3 different words) were added to the beginning and the end of each stream. Subsequently, the stream was faded in and out with an increasing and decreasing ramp during the first and last 696 msec of the stream (corresponding to the duration of the 3 added syllables). Finally, in order to equate the length of the different streams at millisecond level for ERP triggering[1], we used the compression tool of Cool Edit 2000 software to slightly adjust the duration of the audio files. For each language stream the exact duration was set to 5 min 22 s 944 ms, divided into eight short faded streams of 40 s 368 ms and intermixed with 5 s pauses. We opted to introduce pauses along the streams to allow participants to have short breaks and to blink normally.

The use of an artificial language enables full control over the potential segmentation cues that listeners can exploit; here the only reliable cue for word boundaries was the statistical structure (Transitional Probabilities) of the language (Saffran et al., 1996). In all 8-word streams (anchor and non-anchor conditions) TP of the syllables forming a word was 1.0, while for syllables spanning word boundaries the TP was 0.14. For the random streams, TP was $\sim$0.04 for each syllabic pair, which made the syllabic sequence unsegmentable.

Before being exposed to each language stream, participants were taught two words (anchor words) that were then used in the three conditions in different ways. In each language stream for the *Anchor condition*, the anchor words were simply 2 of the 8 words composing the language stream. Therefore, anchor words were heard in the language as part of the set of words composing the stream. For the *Non-anchor condition*, the two words that were taught never appeared in the language stream and were substituted by a set of 14 non-words[2], each one appearing the same number of times along the stream. The 14 non-words were created by recombining the three syllables of the two anchor words taught in the previous phase. Finally, in the *Random condition*, the two recently learned words were the only recognizable items among random sequences of syllables (see Section 2.3 below for a more detailed information). Importantly, the two words taught in the anchor word learning phase, irrespective of whether they appeared or not as part of the stream during segmentation phase, were always presented in the subsequent auditory two-alternative-forced-choice (2AFC) test. Lastly, it is worth to mention that in order to control for possible perceptual biases, i) we used three different language streams in each condition, ii) the assignment of anchor words was counterbalanced across participants, and iii) the meaning assignment for each anchor word was varied across participants.

For the purpose of testing participants' segmentation performance, sets of non-word foils were created for each language stream by recombining the syllables of the words comprising the languages. Thus, for each 8-word language stream (anchor and non-anchor conditions), 8 non-words were created. Non-words were sequences of three syllables that never formed a string in the language stream (TP=0). For the random condition, 16 different trisyllabic sequences were used for each stream.

2.2.2. *Anchor words learning phase*

Before being exposed to each language, the participants were taught two novel words by seeing pictures together with an auditorily presented narration in Spanish (Cunillera et al., 2010a). Subsequently, the participants were exposed to different slideshows in which they were taught two novel words corresponding to two objects (see Supplementary material and Appendix B).

2.3. *Procedure*

The participants underwent the following three consecutive phases altogether nine times: 1) *Anchor word learning phase*: learning two novel words, 2) *Segmentation phase*: an on-line segmentation task, and 3) *Test phase*: assessment of the participants' segmentation performance (see Fig. 1 and Cunillera et al., 2010a). The participants were first instructed to pay attention to the slideshows and to learn the new words that would be presented. The experiment began by playing the introductory slideshow. After that, the slideshow teaching two novel "alien" words (the *anchor word learning phase*) was presented. To ensure that the participants had learned the novel words, they were asked to write down in a notebook the "alien" names of the two objects presented in the slideshow. The notebook was given to the participants only after the presentation of each slideshow and was taken off immediately after they wrote down the two words. In case a participant wrote an erroneous response for any of the two words, the slideshow was replayed[3].

Each participant saw the same slideshows but with different word-object combinations, so that the word-object pairing was counterbalanced across participants. Immediately after successful completion of the *Anchor word learning phase*, the participants were requested to listen carefully to a language stream and to try to discover the words of the novel language (*Segmentation phase*). We used such explicit instructions aiming to ensure that

---

[1] Because words are manipulated as a single auditory file, ERP triggers were defined as dummy number appearing every 696 ms in the EEG signal along the presentation of the auditory file. Consequently, for a correct ERP triggering in this paradigm it was important to ensure that there was a precise millisecond-level correspondence between the onsets of all words (a trisyllabic string) along the auditory stream and the ERP triggers sent.

[2] In a basic sequence (a stream composed of 56 words, i.e., 8 words repeated 7 times each one) each of the 14 non-words appeared just once. As a basic sequence was repeated eight consecutive times during the exposure phase, the word/non-word proportion was 1/7.

[3] One slideshow was replayed once along the experiment for 10/23 participants, and twice for one participant. Only one subject needed three replays of one slideshow and an extra replay for two other slideshows. All these participants failed for one of the two words taught in the slideshow. Importantly, there was no systematic pattern of errors, and each participant failed on different words presented in different slideshows. Only for the word "BOSIRU" there were three participants misspelling the last vowel.

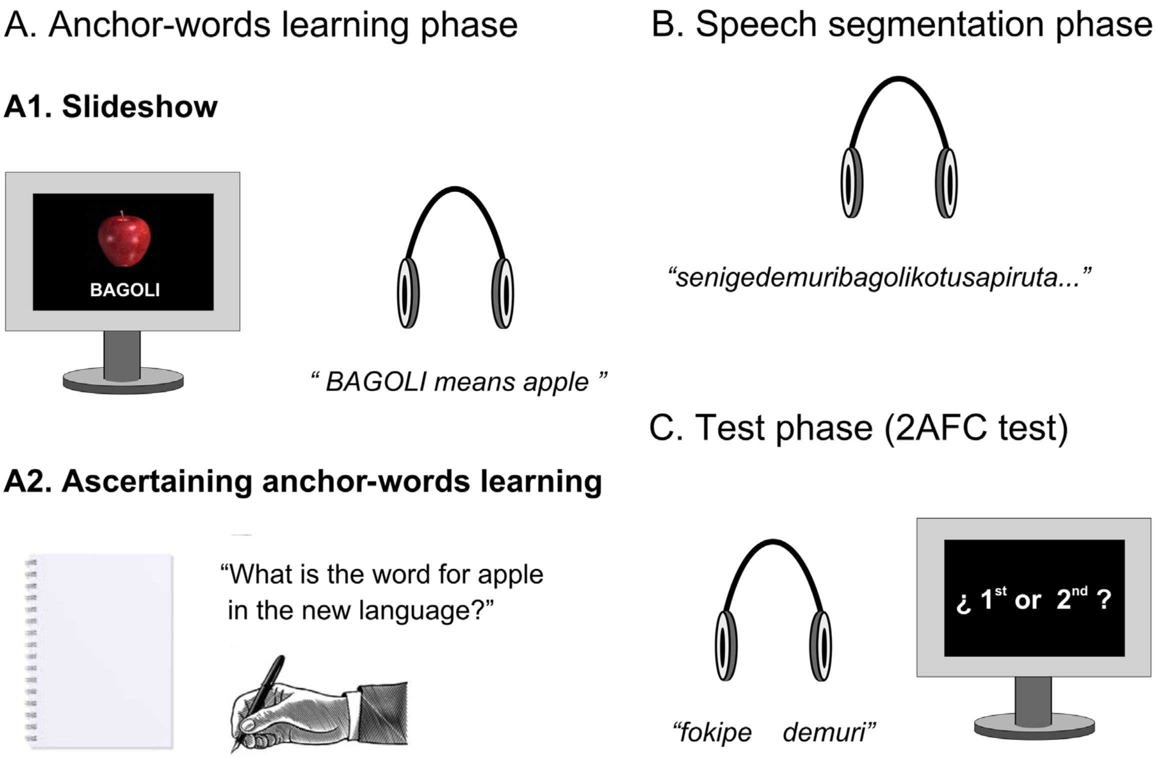


Fig. 1. Illustration of the procedure used in the study. A1. Participants were exposed to different slideshows in which they were taught two novel words corresponding to two objects (see Appendix B). An example of the word *BAGOLI* corresponding in this case to a picture of an "apple" is shown. A2. To ensure that participants had learned the novel words, they were asked to write down the learned names of the two objects presented in the slideshow. B. In the speech segmentation phase, participants were exposed to a continuous speech stream and were instructed to listen carefully and to try to discover the words of the novel language. In three of the streams (anchor condition) the two learned words in the slideshow (anchor words) were part of the language stream, whereas in other three streams the two learned words were not inserted in the streams and each one was substituted by a non-word (non-anchor condition). Finally, for the three streams in the random condition, anchor words were the only true items in the language streams because syllables in those streams were organized in a random order. Participants were blind to this manipulation. C. An example of a trial in the auditory two-alternative-forced-choice test is shown. A word from the stream and a non-word were presented (see Appendix A3) and the participant had to decide which one corresponded to a word presented in the stream.

participants pay attention to the language streams[4]. They were informed that a final test would be presented at the end of the language stream (see Fig. 1). Importantly, they were not informed about the presence or absence of the two recently learned words in the language stream. Furthermore, the participants were encouraged to blink as little as possible and to minimize face and eye movements to avoid introducing artefacts to the EEG signal that was measured while they listened to the language streams. The presentation of the streams was quasi-randomized, with the constraint that no more than one stream of each condition (anchor, non-anchor or random) followed each other.

Immediately after the language stream, a test phase with a standard auditory 2AFC test was delivered to the participants. Test items comprised the eight words of each stream and eight non-words, combined in a way that each word was paired with three different non-words but each of the eight non-words appeared equally often. More specifically, for the anchor and non-anchor conditions, test items comprised the eight words of each stream (including the two anchor words) and eight non-words[5] (see the Section 2.2 and Appendix A2), whereas for the random condition, test items were composed of the two anchor words plus 14 non-words. This procedure rendered a total of 24 word - non-word pairs in which, in all three conditions, the words taught in the previous anchor word learning phase were presented in 6 of the 24 item pairs.

Auditory presentation of the items of a pair was separated by a 696 msec pause and was followed 500 ms later by a visual display (¿ 1 o 2 ?) centered on the computer monitor that prompted participants to choose a response. This visual display remained on the screen for 5000 ms or until a response was given. 500 ms later a vertical array of five asterisks (*****) centered on the computer monitor (duration 2000 ms) indicated to the participant that she/he was free to blink. A fixed interval of 1000 ms was adopted between the offset of the blinking prompt and the occurrence of the next auditory stimulus pair. After hearing each pair, 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. The participants were allowed to rest as much as they needed after each test. The next part of the experiment began with the presentation of another slideshow. The whole experiment lasted from 2 to 2.5 hours.

2.4. *Electrophysiological recordings*

The ERPs were recorded from the scalp using tin electrodes mounted in an elastic cap (Electro-Cap, International 10–20 System locations) and located at 29 standard positions (Fp1/2, Fz, F7/8, F3/4, Fc1/2 Fc5/6, Cz, C3/4, T3/4, Cp1/2, Cp5/6, Pz, P3/4, T5/6, Po1/2, O1/2). Biosignals were referenced on-line to an electrode placed in the outer canthus of the right eye and then rereferenced off-line to the mean of the activity at the two mastoids processes. Electrode impedances were kept below 5 kΩ. The electrophysiological signals were filtered with a band-pass of 0.1–50 Hz (half-amplitude cutoffs) and digitized at a rate of 250 Hz. Vertical

---

[4] In a recently published study, Batterink et al. (2015) compared participants' performance in a speech segmentation task under explicit vs. implicit task instructions. The authors found no differences between the two approaches.

[5] The non-words used in the test phase for the non-anchor condition were different from those non-words presented for such condition in the segmentation phase. In our previous study (Cunillera et al., 2010a) using a similar paradigm but a purely behavioral approach, we used part-words instead of non-words and obtained similar results.

eye movements were monitored with an electrode at the infra-orbital ridge of the right eye. A blind source separation (BSS), a signal-processing automated methodology based on independent component analysis (ICA), was used off-line when necessary to eliminate eye blinks and eye movements (Joyce et al., 2004). Subsequently, the remaining trials that were identified with base-to-peak electro-oculogram (EOG) amplitude of more than 50 μV, amplifier saturation, or a baseline shift exceeding 200 μV/s, were automatically rejected (mean percentage of rejection was 18.1%).

2.5. ERP data analysis

Stimulus-locked ERPs for artifact-free trials corresponding to word onset stimuli were averaged for epochs of 700 msec for the segmentation phase, epochs of 1024 ms for the test phase, and epochs of 1400 ms for measuring the anchor word effect. A minimum of 164 epochs per subject in each condition was averaged. A baseline placed 50 ms prior to the stimulus was used in all analyses, which were performed for each participant separately for the three different conditions.

Following previous studies (Astheimer and Aslin, 2011; Cunillera et al., 2006; 2008; 2009), mean amplitude measures were taken in different time-windows (TW) encompassing the major ERP components of interest for the current study, i.e., the SPN, N400-like, and the N100 components. For measuring the SPN for anchor words in the anchor condition, an extensive TW (464–928 ms) was used, which encompassed the third syllable of the anchor word and the first syllable of the following word. For the N400-like, a TW at the time range 350–550 ms was taken for the analyses. Finally, for the N100 the TW entered into the analysis (106–166 ms) encompassed a 60 ms time range around the peak located on the grand average waveform.

These mean amplitude measures were then submitted to separate repeated measures ANOVAs always with a within-subjects factor **Word condition** (*Anchor* vs. *Non-anchor vs. Random*) and **Electrode** (15 levels: Fz, F7/8, F3/4, Cz, C3/4, T3/4, Pz, P3/4, T5/6). For the analysis of the anchor word effect in the anchor condition, the factor **Word sequence** (*Words after Anchors* vs. *Words after Words*) was considered for the analysis.

Twelve of the 15 selected electrodes were used for topographical analysis. This analysis was conceived to decompose significant interactions in which the electrode factor was involved. The 12 selected electrodes (F7/8, F3/4, T3/4, C3/4, T5/T6, P3/4) were divided according to three factors: **Hemisphere** [*right* (F7, F3, T3, C3, T5, P3) vs. *left* (F8, F4, T4, C4, T6, P4)], **Anterior-Posterior** [*anterior* (F7, F3, F8, F4) vs. *central* (T3, C3, T4, C4) vs. *posterior* (T5, P3, T6, P4)], and **Laterality** [*lateral* (F7, T3, T5, F8, T4, T6) vs. *medial* (F3, C3, P3, F4, C4, P4)]. For all statistical effects involving two or more degrees of freedom in the numerator, the Huynh-Feldt epsilon was used to correct for possible violations of the sphericity assumption (Jennings and Wood, 1976). The uncorrected degrees of freedom and adjusted *p*-values after the correction are reported. The effect size for all the experimental results is reported as Partial Eta squared (η_p^2), when computing repeated measures ANOVAs, as Cohen's f^2 when calculating one-way ANOVAs, and as Cohen's *d* when computing *t*-tests. For illustrative purposes only, a 0.5–6 Hz band-pass filter was applied to the grand-average ERPs for the on-line segmentation data. For better capturing the slow negative potential regarding the effect of anchor words in the online segmentation phase, a 0.5–3 Hz band-pass filter (Starr et al., 1997) was employed in the 1400 ms epochs encompassing the processing of two consecutive words.

3. Results

3.1. Behavioral results

When participants were asked to identify the novel word in the two-alternative forced-choice test, the 18 participants' mean percentage of correctly segmented words was 64.9 ± 10.4% for the anchor condition and 54.5 ± 9.5% for the non-anchor condition (the 6 out of 24 pairs of items that included an anchor word were excluded from the analysis). These percentages are significantly above chance level or at the border of significance, as revealed by a one-sample *t*-test (two-tailed) with chance level placed at 50% (Anchor condition: $t(17)=6.0$, $p<.0001$, $d=2.9$; Non-anchor condition: $t(17)=2.0$, $p=.061$, $d=.97$). A comparison of the percentage of hits in the two conditions revealed a significantly higher segmentation rate[6] in the anchor condition ($t(17)=3.4$, $p<.01$, $d=1.0$). Importantly, no differences were found among the three language streams within each condition, as revealed by the results of the one-way ANOVA (anchor condition: $F(2,53)=.1$, $p>.8$, $f^2=.004$; non-anchor condition: $F(2,53)=2.6$, $p=.08$, $f^2=.11$). This indicates that the participants did not show any preferences for the specific artificial languages employed in the experiment.

An additional analysis with binomial tests was conducted to identify the participants who performed better than expected by chance in the anchor and non-anchor conditions[7]. Chance level was first determined for a *p*-value < 0.05 and for the 108 test-items that comprised the test in each condition. Accordingly, performance at or above 56.4% in the average of all tests was categorized as significantly better than chance. Twelve participants fulfilled this criterion. For these twelve participants, the mean percentage of correctly segmented words was 68.7 ± 9.9% in the anchor condition, and 58.6 ± 7.3% in the non-anchor condition. These percentages indicated that the performance level was significantly better in the anchor than in the non-anchor condition ($t(11)=2.7$, $p = .02$, $d = 1.2$).

Finally, we calculated the percentage of correctly detected anchor words in the test for the three different conditions. Thus, the accuracy rate was 92.3 ± 10.6% for the anchor words in the anchor condition, 83.3 ± 10.4% in the non-anchor condition and 89.2 ± 10.6% in the random condition. The results of the ANOVA revealed a significant difference among the three conditions ($F(2,34)=3.5$, $p<.05$, $\eta_p^2=.17$). Further *t*-tests showed that the percentage of the anchor words detected at the test phase was higher in the anchor than in the non-anchor condition ($t(17)=2.9$, $p<.01$, $d=.85$), but there were neither differences between the anchor and the random conditions ($t(17)=0.9$, $p>.3$, $d=.26$) nor between the non-anchor and the random conditions ($t(17)=-1.5$, $p>.1$, $d=-.5$).

3.2. ERP results

3.2.1. On-line word segmentation

Fig. 2 depicts the grand average ERP signatures for words during the learning phase in the anchor and non-anchor conditions and non-words in the random condition, based on the data from all participants (n=18). The P50-N100-P200 complex of the auditory evoked components was clearly identifiable in the three conditions (see Cunillera et al., 2006). For non-words in the random condition, a slightly diminished amplitude of N100 component was observed mainly on central-parietal electrodes. The results of the ANOVA, however, showed that this reduction of the N100 component did not reach statistical significance ($F(2,34)=0.5$, $p>.5$, $\eta_p^2=.03$; Word

---

[6] In the current study the concept of word segmentation should be considered equivalent to pattern familiarity, as a proficient performance in a 2AFC test can be explained as a simple detection of a familiar pattern as well as a whole word form recognition performance (see Franco et al., 2011).

[7] Test items that comprised anchor words were excluded from the analysis testing participants' performance on word segmentation. Thus 6 item pairs from the pool of 24 pairs in each language stream were not included in the analysis. It should be noted that although word recognition and speech segmentation can be considered as equivalent concepts, in the current study word recognition refers to the identification of already acquired words, whereas speech segmentation refers to the identification of possible new words.

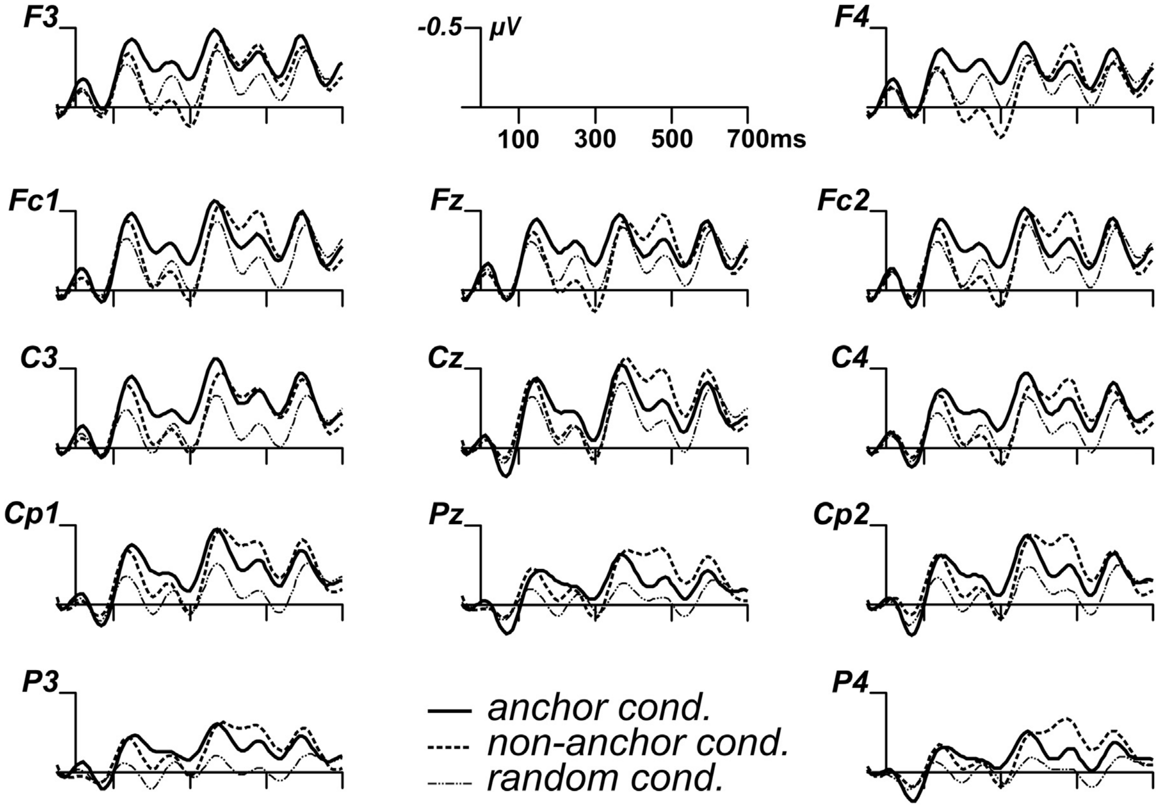


Fig. 2. Grand average ERPs of the speech segmentation phase for words in the anchor (solid lines), non-anchor (dashed-thick lines) and random (dashed-thin lines) conditions are depicted. Different electrode positions are shown covering frontal, central and parasagittal locations. Negativity is plotted upwards.

condition × Electrode, $F(28,476)=0.8$, $p>.5$, $\eta_p^2=.05$). As expected, these early components were followed by a central broadly distributed N400-like negativity in the 350–550 ms time-range, most notably observed for words in the anchor and non-anchor conditions. For this component, the ANOVA showed a significant main effect of Word condition ($F(2,34)=3.6$, $p<.04$, $\eta_p^2=.18$). This effect was even stronger when the mean amplitudes were computed in a narrower time-window (TW: 400–550 ms ($F(2,34)=4.6$, $p<.02$, $\eta_p^2=.21$). We proceeded by comparing words in the three conditions. The analyses revealed a clearly significant difference in the amplitude of the N400-like component for the non-anchor condition when compared to the random one ($F(1,17)=7.1$, $p<.02$, $\eta_p^2=.29$; Word condition × Electrode, $F(14,238)=2.0$, $p=.08$, $\eta_p^2=.11$), and a non-significant difference between the anchor and the random condition ($F(1,17)=2.7$ $p=.1$, $\eta_p^2=.14$), as well as between the anchor and the non-anchor condition ($F(1,17)=0.7$ $p>.4$, $\eta_p^2=.4$). The N400-like amplitude increase for words in the non-anchor condition was localized on the medial and central-posterior scalp region, as revealed by the significant Word condition × Laterality × Anterior-Posterior ($F(2,34)=4.8$, $p<.02$, $\eta_p^2=.22$) and Word condition × Laterality interactions [($F(1,17)=6.7$, $p<.02$, $\eta_p^2=.28$; Mean voltage difference (non-anchor minus random) at lateral: −0.09 µV; medial: −.18 µV; anterior: −0.09 µV; central: −0.14 µV; posterior: −0.18 µV)].

3.2.2. *Anchor word effect during the learning phase*

Here we analyzed two-word sequences, with the first word of the sequence being either an anchor word or a word (see Fig. 3A). By narrowing the signal into a low frequency range (0.5–3 Hz), a clear slow and long-lasting negative deflection was observed at frontal locations for anchor words in the anchor condition, interpreted as the SPN component. This SPN effect started after the onset of the second syllable of the anchor word, reached its maximum before the end of the anchor word, and declined before the onset of the second syllable of the next word. The results of the ANOVA revealed a significant Word sequence (*Words after Anchors* vs. *Words after Words*) × Electrode interaction ($F(14,238)=3.7$, $p<.01$, $\eta_p^2=.18$). The topographical analysis revealed a significant Word sequence × Anterior–Posterior interaction ($F(2,34)=8.9$, $p<.01$, $\eta_p^2=.34$), indicating a specific frontal distribution for the long-lasting negative deflection related with the processing of anchor words in the anchor condition (see topographical maps in Fig. 3B; mean voltage difference (words-after-anchor minus word-after-word) at anterior: −0,50 µV; central: −0,18 µV; posterior: −0,01 µV). Further *t*-tests revealed that these difference were significant at frontal-lateral electrodes (F3: $t(17)=-2.34$, $p<0.04$, $d=-0.8$; Fz: $t(17)=-1.91$, $p=0.07$, $d=-0.6$; F4: $t(17)=-2.16$, $p<0.05$, $d=-0.6$).

We also studied the temporal evolution of the SPN during the segmentation phase. For this purpose, we divided the ERP signal into four consecutive 1 min 20 s 736 ms blocks. Then, we selected a set of frontal electrodes as a Region of Interest for the analysis (ROI: Fp1/2, F3/4, F7/8, and Fz), and finally computed the difference between Word sequences (*Words after Anchors* minus *Words after Words*) to isolate the SPN. The results suggested that the SPN did not vary along the segmentation phase (main effect of Block: $F<1$; Block × Electrode: $F<1$). However, further exploratory analyses revealed that the SPN at the F3 electrode, where the SPN effect was largest, increased from the 1st to the 2nd block and vanished afterwards (see Fig. 5; 1st vs. 2nd block: $t(17)=2.1$, $p=.05$, $d=0.6$; 2nd vs. 3rd block: $t(17)=-1.6$, $p>.01$, $d=0.4$; 3rd vs. 4th block: $t(17)<.1$).

3.2.3. *SPN effect associated to individual learning performance*

We further studied the importance of the anchor word effect and word expectancy by correlating the mean amplitude of the SPN in the frontal region – where it reached its maximum (at F3 electrode) – with the word recognition performance at the test phase. This analysis revealed a strong correlation ($r(16) = 0.62$, $p < 0.01$; see Fig. 4) indicating that a lower amplitude of the SPN was associated with a higher subsequent word recognition performance .

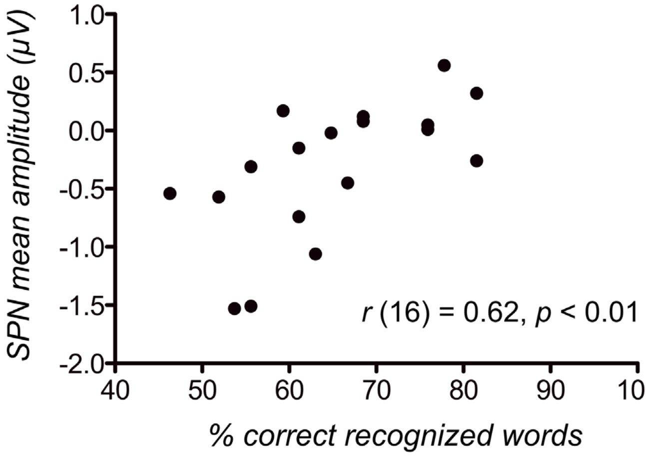


Fig. 4. Brain-behavior correlation analysis showing a significant association between the mean amplitude of the SPN component at F3 electrode during the segmentation phase and the subsequent behavioral performance in the anchor word condition (percentage of words recognized at the test phase).

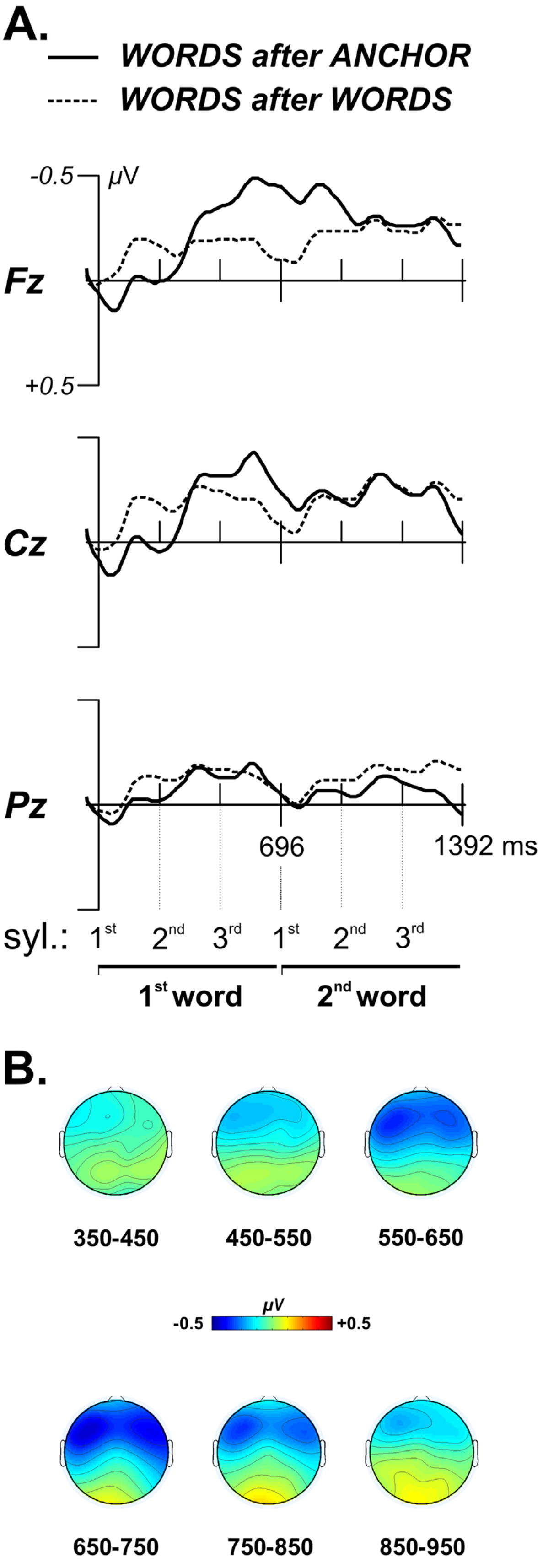


Fig. 3. A. Grand average ERPs at midline electrode locations (Fz, Cz, and Pz) for anchor-word sequences (e.g., *BAGOLI-KOTUSA*, in which *BAGOLI* is the anchor word) compared with word-word sequences (e.g., *FURAKE-SOMEPO*) in the anchor condition during the speech segmentation phase. A clear modulation of the SPN component is elicited specifically by anchor words in frontal locations. Superimposed is depicted the time of the onset of each syllable in these sequence of an anchor word (*BAGOLI*) and a following word in the stream (in this case, *KOTUSA*) corresponding to the “words after anchor” ERP waveform. Note how the SPN begins with the processing of the second syllable of the anchor word and decays at the end of the first syllable of following word in the stream. B. The topographical maps (isovoltage mapping with spherical spline interpolation) represent the time evolution of the difference waveform (mean amplitude of the anchor-word minus word-word sequence) in 100 ms steps starting at 350 ms. This time evolution illustrates the frontal distribution of the SPN elicited by the anchor words.

4. Discussion

This study sought to highlight the incremental language learning process by investigating the neurophysiological correlates of the anchor word effect, i.e., how newly learned words facilitate statistical learning when mixed with novel words during a speech segmentation task. In line with previous studies, the behavioral results indicated a higher segmentation performance in those streams where the anchor words were embedded (Bortfeld et al., 2005; Cunillera et al., 2010a). Importantly, our study reveals that in the anchor word condition, the anchor words elicited a clear fronto-central negativity, identified as an SPN modulation, presumably reflecting the temporal expectancy of the appearance of a new word immediately following the anchor word. Moreover, a strong correlation was found between the amplitude of the SPN and subsequent behavioral performance, indicating that low expectancy may be indicative of an improvement in word segmentation and word learning. Finally, the largest N400-like component was found for the non-anchor condition. We interpret this as a rapid diminishment of the N400-like component in the anchor condition during the exposure phase due to more effective word learning in this condition.

4.1. Behavioral results

The present findings provide further support to the view that the very first words acquired in a new language facilitate the discovery of further words. In other words, the anchors aid the statistical learning process based on transitional probabilities in

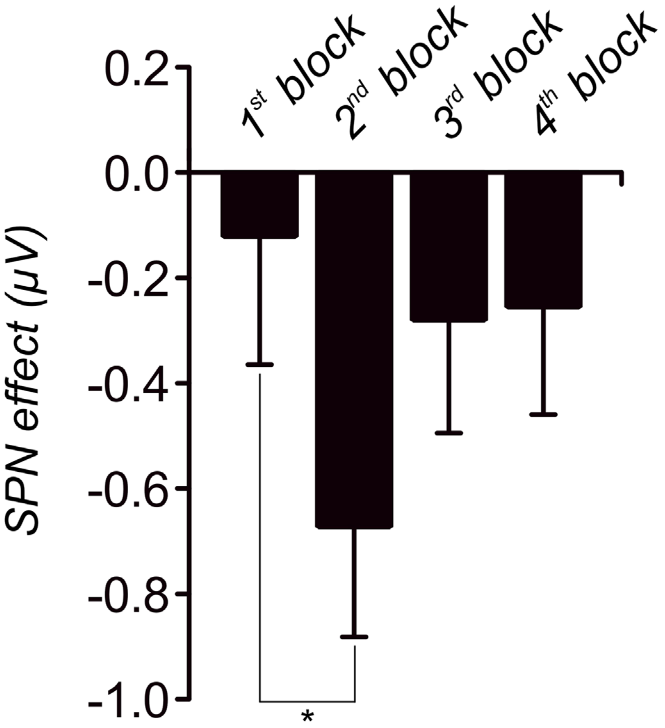


Fig. 5. Bar plot corresponding to the mean amplitude (μV) of the SPN effect at F3 electrode, computed as the difference between Word sequences (*Words after Anchors* minus *Words after Words*) in the T.W. 464–928 ms for the four consecutive blocks in which the segmentation phase was divided to study the time-evolution of the SPN. The * denotes a p-value < 0.05.

segmenting out a novel language (Saffran et al., 1996). Similar results showing that adult listeners can combine statistical learning with other segmentation cues available in speech are found in the literature. For instance, Schön et al. (2008; see also Schön and François, 2011) found a positive effect of combining redundant intrasensory statistical regularities (speech and music) in a speech segmentation task. In a similar vein, other studies have explored the capacity of learners to extract and combine regularities across the visual and the auditory modalities (Cunillera et al., 2010b, 2010c; Glicksohn and Cohen, 2013; Mitchel and Weiss, 2011).

An interesting account on speech segmentation by Davis and Johnsrude (2007) emphasizes the interaction between lexical knowledge and sublexical cues. These authors postulate that the complete account of perception of spoken language (in which speech segmentation is a preliminary step) requires an interactive mechanism involving top-down (lexical knowledge) and bottom-up (sublexical knowledge) processes where multiple, parallel representations of the speech input make distinct contributions to the comprehension of speech. The facilitatory effects of the merging of lexical and sublexical cues for segmenting speech have also been documented in computer simulations of language listeners who already have an internal lexicon (e.g., Grossberg and Myers, 2000; Norris et al., 1997). This issue has recently been addressed by Räsänen and Rasilo (2015) in a computational model that integrates segmentation and word-reference mapping during the creation of the lexicon. Moreover, models based on a chunking mechanism, as PARSER (Perruchet and Vinter, 1998), INCDROP (Brent, 1997), and MDLChunker (Robinet et al., 2011), should be able to accommodate anchor words in the input to improve learning, as all these three models are based on an incremental approach with the first segmented words guiding the discovery of the other words embedded in the input data (see Perruchet and Tillmann, 2010).

Participants in our study were able to benefit from the learned words although their experience with these words was minimal. This finding demonstrates that lexical items can contribute to speech segmentation immediately after they are learned (Bortfeld et al., 2005). In this vein, recognizing a word in the middle of a nonsense stream may lead the learner to anticipate an incoming new possible word.

4.2. ERP findings

Considering the importance of context and predictability in language processing, it is plausible to assume that word predictability can be crucial in successful language learning (Conway et al., 2010). Following this argumentation, the recent lexical status acquired by the anchor words in our study may facilitate the temporal tracking of the appearance of incoming words along the language stream. Thus, the recently learned anchor words could be acting as an attentional temporal cue that allows the language learner to focus attention on the immediately following novel word in the auditory stream. Supporting the view of attention as a key aspect of a successful segmentation, it has been proposed that similar attentional facilitation might take place in auditory stream composed by units of the same length. This unit length regularity is considered to act as an additional attentional cue that facilitates word segmentation (Hoch et al., 2013). This attentional hypothesis is supported by the modulation of the SPN in anchor words, which we observed only on those streams in which the anchor word was mixed with other words in the stream. This suggests that the learners at first focused on familiar parts (i.e., the anchor words) in the auditory stream. These landmarks enhanced their attentional focus that remained active for the processing of the following word in the stream, thus enhancing learning. In this sense it is possible that familiar words might became a cue that facilitates the temporal predictability of the incoming novel word, reflected as increased (negative) activation at frontal brain areas. Supporting evidence is found in a recent study on adults' audiovisual word learning that reported that visual attention was directed towards familiar patterns, demonstrating that selective attention was crucially related to the cognitive mechanism supporting cross-situation learning (Yu et al., 2012). A recent study conducted by Morís et al. (2013) demonstrated that SPN could be a plausible index of expectation during learning. The authors analyzed the temporal evolution of the SPN during an associative learning task and found that the SPN amplitude decreased as a function of learning, i.e., their results showed a clear decrease of the SPN amplitude while learning progressed. In this particular case it was found that, as soon as associations were created, participants decreased their attention load devoted to the incoming feedback informing about the learning process. The SPN has been considered as the anticipatory index underlying the consequences of actions (Brunia et al., 2011), but it cannot be ignored that the SPN can also represent the modulation of expectations per se, as in the case of the current task where learning occurs in the absence of feedback or a clear action. In the current study, we found a strong positive correlation between the amplitude of the SPN and learning outcome with a low SPN predicting better subsequent word recognition performance, indicating that a diminishment of word expectancy may signals progress in word learning. Interestingly, parallel results were found by Astheimer and Sanders (2011) who observed an enhancement in the amplitude of the N100 component elicited by word onsets for unpredicted vs. predicted words. In their study, participants were exposed to different continuous language streams in which words were arranged in pairs in the stream, making it possible to predict the second word of a pair through cumulative exposures. These results support the claim that during the initial stages of learning, learners might attend to word onsets and their attention decreases as word learning progresses. The lack of modulation of the N100 component in the current study may stem from the fact that, in contrast to Astheimer and Sanders (2011), all words in our streams were equally unpredictable.

Another ERP finding in the current study was that words that were segmented elicited increased amplitude of the N400-like component. This effect was visible in both the non-anchor and anchor conditions, even though it reached statistical significance only in the former one. This confirms the findings of the N400-like component as a neurophysiological index of word segmentation (Abla et al., 2008; Buiatti et al., 2009; Cunillera et al., 2009). In adult and infant studies, a N400-like component, a negative polarity deflection in the time-range of 200–500 ms with a fronto-central topography, has been related to the word-learning process (for a review, see Rodríguez-Fornells et al., 2009), and numerous studies have reported that segmentation from a continuous stream of words or tones is indexed by larger amplitudes of this N400-like component (Abla et al., 2008; Buiatti et al., 2009; Cunillera et al., 2006, 2009; Francois and Schön, 2011a, 2011b; Sanders et al., 2002). The fact that a better segmentation performance here was not reflected in a larger amplitude of the N400-like component can be interpreted as an indication of word consolidation, reflected by a higher word recognition performance in the test. This interpretation is in line with the fact that the largest effect of the N400-like component is found during the early learning phase, and this effect declines once the segmentation is achieved (see Cunillera et al., 2009). Thus, similarly to what is observed here with the SPN, the low amplitude of the N400 effect may be interpreted as an index of word learning. Despite the apparent similarities of these two components concerning stimulus predictability, there is a crucial distinction between them. It is well known that the N400 varies as a function of semantic properties or word frequency, among others. However, when considering contextual information, the amplitude of the N400 is inversely correlated with the predictability (cloze probability) of a target word (Kutas and Federmeier, 2011). In this vein, the N400 is understood as an indirect index of anticipation, as it is elicited at the time when the previously anticipated target is processed. On the other hand, the SPN is thought to reflect anticipatory attention for incoming relevant information such as instructions, feedback or affective stimuli (van Boxtel and Böcker, 2004). This makes SPN a more direct index of expectancy, reflecting the deployment of attentional resources to anticipated incoming significant information. In the current experiment, the SPN may index the anticipation of the incoming onset of a new word, reflecting an attentional recruitment that is needed to facilitate its processing.

5. Conclusions

When segmenting novel words from a continuous speech input, the statistical learning mechanism is facilitated by the presence of previously acquired words. The present results confirm that this facilitation encompasses even words that have been learned right before the exposure to the novel speech stream. The new finding reported in the current study was that the anchor word effect was reflected by a modulation of the SPN, an ERP component indexing anticipation of incoming relevant information. Thus the anchor words created an expectancy for the immediately following to-be-segmented novel words. Confirming earlier findings, we also found an enhanced N400-like component presumably reflecting the on-line speech segmentation process based on the computation of transitional probabilities.

Acknowledgments

We thank Irene Nogué and Laura Sánchez-Blanco for their help during data collection. This study was supported by grants MICINN from the Spanish Government (TC: PSI2011-23624; ARF: PSI2011-29219) and by a grant from the Academy of Finland (project #260276) and from the Abo Akademi University Endowment to ML and from the Abo Akademi University Endowment (the BrainTrain project) to ML. Finally, we are grateful to two anonymous reviewers for their thoughtful comments.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2016.01.011.

Appendix A

See Appendix Tables A1 and A2.

A3. For each language stream (duration, 5 min 22 s 944 ms), the set of 8 pseudowords were combined in four different ways within the stream (duration, 40 s 368 ms). Thus, each language stream was composed by the part streams combined in the following order: Concatenation-1 + C-2 + C-3 + C-4 + C-1 + C-2 + C-3 + C-4. A written excerpt for the four variants of the same language stream is provided with slashes denoting word boundaries.

Language-A (piruta, bagoli, demuri, senige, kotusa, tokuda, furake, somepo).

Concatenation-1: demuri / senige / somepo / kotusa / tokuda / piruta / furake / bagoli / senige / tokuda / demuri / somepo / piruta / kotusa / furake / piruta / bagoli / somepo / tokuda / …

Concatenation-2: tokuda / piruta / demuri / bagoli / somepo / kotusa / senige / furake / piruta / somepo / tokuda / demuri / kotusa / bagoli / senige / kotusa / furake / demuri / somepo / …

Concatenation-3: kotusa / somepo / tokuda / furake / piruta / demuri / bagoli / senige / somepo / piruta / kotusa / tokuda / demuri / furake / bagoli / demuri / senige / tokuda / piruta / …

Concatenation-4: bagoli / tokuda / senige / piruta / demuri / furake / somepo / kotusa / tokuda / demuri / bagoli / senige / furake / piruta / somepo / furake / kotusa / senige / demuri / …

Table A1
The set of 12 consonants and 5 vowels used to create the 60 syllables that were combined to create the words in all language streams. In brackets after consonants and vowels, the phonological transcription (Spanish phonological alphabet) of the graphemes are illustrated.

Consonants/vowels	A [a]	E [e]	I [i]	O [o]	U [u]
B [b]	BA	BE	BI	BO	BU
D [d]	DA	DE	DI	DO	DU
F [f]	FA	FE	FI	FO	FU
G [g]	GA	GE	GI	GO	GU
K [k]	KA	KE	KI	KO	KU
L [l]	LA	LE	LI	LO	LU
M [m]	MA	ME	MI	MO	MU
N [n]	NA	NE	NI	NO	NU
P [p]	PA	PE	PI	PO	PU
R [rr]	RA	RE	RI	RO	RU
S [s]	SA	SE	SI	SO	SU
T [t]	TA	TE	TI	TO	TU

Table A2
Words of the artificial languages used in the anchor, non-anchor and random conditions. For three of the nine language streams, the participants learned two of the eight words composing the novel language (anchor condition). For another three streams, the participants learned two words that were not presented in the subsequent language stream (non-anchor condition). Finally, for the remaining 3 streams, the participants learned two words that were the only recognizable words in a syllabic random sequence (random condition). For the anchor and random conditions, the presentation of the anchor words was counterbalanced across participants and languages.

	Anchor condition				**Non-anchor condition**		
	Lang. 1	**Lang. 2**	**Lang. 3**		**Lang. 1**	**Lang. 2**	**Lang. 3**
Word-1	PIRUTA	PABELA	BAMOFI	**Word-1**	GAMIRE	BAKUMO	MOGAFU
Word-2	BAGOLI	LUFAGI	NULOPI	**Word-2**	SIRAKO	FEDALU	BELARI
Word-3	DEMURI	FOLETI	KELAFO	**Word-3**	MATEPU	FAGELI	POBUFE
Word-4	SENIGE	BULOTE	GIREDA	**Word-4**	LEPOTI	ROPENI	KANESO
Word-5	KOTUSA	DINEKA	PUSONE	**Word-5**	FUSENA	SUTOGI	MIGUDO
Word-6	TOKUDA	GUKIBO	MAGURO	**Word-6**	BODUFI	PIBUNO	DESIMU
Word-7	FURAKE	DUBIPE	BETAKI	**Non-word-1**	BESODO	DENUKI	DEKAGU
Word-8	SOMEPO	NUGADO	TIDESU	**Non-word-2**	BETUDO	DEPEKI	DESOFU
Anchor-1	BAGOLI	DINEKA	NULOPI	**Non-word-3**	DORISA	FODEPE	FUGUMI
Anchor-2	FURAKE	NUGADO	KELAFO	**Non-word-4**	DOSOSA	FOKIPE	FUKAMI
				…			
				Non-word-14	TURIBE	PEKIDE	SOMIKA
	Random condition						
	Syllables			**Anchor-1**	**Anchor-2**		
Lang. 1	BA-DE-FI-GO-KU-LA-ME-NI-PO-RU-DA-FE-GI-KO-LU-TA-SE-TO			FAGEBI	SAPEKA		
Lang. 2	BE-DI-FO-GU-KA-LE-MI-RO-PU-RA-FA-GE-KI-LO-MU-SA-NE-TI			LIPAFE	MEDOTU		
Lang. 3	BI-DO-FU-GA-KE-LI-MO-NU-PA-RE-NA-PE-RI-SO-TU-DU-PI-MA			BOSIRU	NODILU		

Appendix B

See Appendix Table B1.

Table B1
The table illustrates the word frequency per million words in Spanish (drawn from the Spanish lexical data-base LEXESP (Sebastián-Gallés et al., 2000), imageability, familiarity, and concreteness for the visual object pictures used in the anchor words learning phase of the experiment. The last three variables are rated by a 1–7 scale where 7 denotes highest imageability, familiarity, or concreteness.

	Frequency	**Imageability**	**Familiarity**	**Concreteness**
Manzana (apple)	11.1	6.6	6.5	6.1
Naranja (orange)	11.6	6.0	6.4	5.5
Combustible (fuel)	13.8	5.5	6.3	4.7
Pila (battery)	10.5	6.0	6.3	5.1
Chocolate (chocolate)	14.8	6.6	6.7	6.1
Leche (milk)	54.1	6.3	6.5	5.7
Agua (water)	295.4	6.4	6.8	5.9
Pan (bread)	54.6	6.8	6.6	6.7
Miel (honey)	18.4	5.3	5.5	6.6
Queso (cheese)	11.1	5.9	5.9	5.9
Mapa (map)	22.3	6.2	6.0	4.3
Linterna (flashlight)	8.2	5.8	6.0	6.3
Chaqueta (jacket)	27.1	6.5	6.6	5.8
Bota (boot)	13.0	5.6	6.4	6.5
Libreta (notebook)	3.6	6.4	6.3	5.4
Lápiz (pencil)	7.0	5.9	6.3	6.8
	36.0	**6.1**	**6.3**	**5.8**

References

Abla, D., Katahira, K., Okanoya, K., 2008. On-line assessment of statistical learning by event-related potentials. J. Cognit. Neurosci. 20, 952–964.

Astheimer, L.B., Sanders, L.D., 2009. Listeners modulate temporally selective attention during natural speech processing. Biol. Psychol. 80, 23–34.

Astheimer, L.B., Sanders, L.D., 2011. Predictability affects early perceptual processing of word onsets in continuous speech. Neuropsychologia 49, 3512–3516.

Batterink, L.J., Reber, P.J., Neville, H.J., Paller, K.A., 2015. Implicit and explicit contributions to statistical learning. J. Mem. Lang. 83, 62–78.

Bortfeld, H., Morgan, J.L., Golinkoff, R.M., Rathbun, K., 2005. Mommy and me – familiar names help launch babies into speech-stream segmentation. Psychol. Sci. 16, 298–304.

Brent, M.R., 1997. Toward a unified model of lexical acquisition and lexical access. J. Psycholinguist. Res. 26, 363–375.

Brent, M.R., Cartwright, T.A., 1996. Distributional regularity and phonotactic constraints are useful for segmentation. Cognition 61, 93–125.

Brunia, C.H., Hackley, S.A., van Boxtel, G.J., Kotani, Y., Ohgami, Y., 2011. Waiting to perceive: reward or punishment? Clin. Neurophysiol. 122, 858–868.

Buiatti, M., Peña, M., Haene-Lambertz, G., 2009. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. Neuroimage 44, 509–519.

Conboy, B.T., Mills, D.L., 2006. Two languages, one developing brain: event-related potentials to words in bilingual toddlers. Dev. Sci. 9, F1–F12.

Conway, C.M., Bauernschmidt, A., Huang, S.S., Pisoni, D.B., 2010. Implicit statistical learning in language processing: word predictability is the key. Cognition 114, 356–371.

Cunillera, T., Càmara, E., Laine, M., Rodríguez-Fornells, A., 2010a. Words as anchors known words facilitate statistical learning. Exp. Psychol. 57, 134–141.

Cunillera, T., Càmara, E., Laine, M., Rodríguez-Fornells, A., 2010b. Speech segmentation is facilitated by visual cues. Q. J. Exp. Psychol. 63, 260–274.

Cunillera, T., Càmara, E., Toro, J.M., Marco-Pallarès, J., Sebastián-Gallès, N., Ortiz, H., Pujol, J., Rodríguez-Fornells, A., 2009. Time course and functional neuroanatomy of speech segmentation in adults. NeuroImage 48, 541–553.

Cunillera, T., Gomila, A., Rodríguez-Fornells, A., 2008. Beneficial effects of word final stress in segmenting a new language: evidence from ERPs. BMC Neurosci. 9, 23.

Cunillera, T., Laine, M., Càmara, E., Rodríguez-Fornells, A., 2010c. Bridging the gap between speech segmentation and word-to-world mappings: evidence from an audiovisual statistical learning task. J. Mem. Lang. 63, 295–305.

Cunillera, T., Toro, J.M., Sebastián-Gallès, N., Rodríguez-Fornells, A., 2006. The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. Brain Res. 1123, 168–178.

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. J. Exp. Psychol.-Gen. 128, 165–185.

Davis, M.H., Johnsrude, I.S., 2007. Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear. Res. 229, 132–147.

De Diego-Balaguer, R., Rodríguez-Fornells, A., Bachoud-Levi, A.C., 2015. Prosodic cues enhance rule learning by changing speech segmentation mechanisms. Front. Psychol. 6, e01478.

De Diego-Balaguer, R., Toro, J.M., Rodríguez-Fornells, A., Bachoud-Levi, A.C., 2007. Different neurophysiological mechanisms underlying word and rule extraction from speech. PLoS One 2, e1175.

Donkers, F.C., Nieuwenhuis, S., van Boxtel, G.J., 2005. Mediofrontal negativities in the absence of responding. Cognit. Brain Res. 25, 777–787.

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. Philadelphia. pp. 1393–1396.

Franco, A., Cleeremans, A., Destrebecqz, A., 2011. Statistical learning of two artificial languages presented successively: how conscious? Front. Psychol. 2, 229.

Francois, C., Schön, D., 2011a. Musical expertise and statistical learning of musical and linguistic structures. Front. Psychol. 2, 167.

Francois, C., Schön, D., 2011b. Musical expertise boosts implicit learning of both musical and linguistic structures. Cereb. Cortex 21, 2357–2365.

Friedrich, M., Friederici, A.D., 2008. Neurophysiological correlates of online word learning in 14-month-old infants. Neuroreport 19, 1757–1761.

Glicksohn, A., Cohen, A., 2013. The role of cross-modal associations in statistical learning. Psychon. Bull. Rev. 20, 1161–1169.

Grossberg, S., Myers, C.W., 2000. The resonant dynamics of speech perception:

interword integration and duration-dependent backward effects. Psychol. Rev. 107, 735–767.

Jennings, J.R., Wood, C.C., 1976. Epsilon-adjustment procedure for repeated measures analyses of variance. Psychophysiology 13, 277–278.

Hoch, L., Tyler, M.D., Tillmann, B., 2013. Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. Psychon. Bull. Rev. 20, 142–147.

Joyce, C.A., Gorodnitsky, I.F., Kutas, M., 2004. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. Psychophysiology 41, 313–325.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647.

Large, E.W., Jones, M.R., 1999. The dynamics of attending: How people track time-varying events. Psychol. Rev. 106, 119–159.

Marslen-Wilson, W.D., Welsh, A., 1978. Processing interactions and lexical access during word recognition in continuous speech. Cognit. Psychol. 10, 29–63.

McClelland, J.L., Elman, J.L., 1986. The trace model of speech-perception. Cognit. Psychol. 18, 1–86.

Mills, D.L., Plunkett, K., Prat, C., Schafer, G., 2005. Watching the infant brain learn words: effects of vocabulary size and experience. Cognit. Dev. 20, 19–31.

Mitchel, A.D., Weiss, D.J., 2011. Learning across senses: cross-modal effects in multisensory statistical learning. J. Exp. Psychol. - Learn. Mem. Cogn. 37, 1081–1091.

Morís, J., Luque, D., Rodríguez-Fornells, A., 2013. Learning-induced modulations of the stimulus-preceding negativity. Psychophysiology 50, 931–939.

Norris, D., 1994. Shortlist – a connectionist model of continuous speech recognition. Cognition 52, 189–234.

Norris, D., McQueen, J.M., Cutler, A., Butterfield, S., 1997. The possible-word constraint in the segmentation of continuous speech. Cognit. Psychol. 34, 191–243.

Perruchet, P., Tillmann, B., 2010. Exploiting multiple sources of information in learning an artificial language: human data and modeling. Cognit. Sci. 34, 255–285.

Perruchet, P., Vinter, A., 1998. PARSER: a model for word segmentation. J. Mem. Lang. *39*, 246–263.

Räsänen, O., Rasilo, H., 2015. A joint model of word segmentation and meaning acquisition through cross-situational learning. Psychol. Rev. 122, 792–829.

Robinet, V., Lemaire, B., Gordon, M.B., 2011. MDLChunker: a MDL-based cognitive model of inductive learning. Cognit. Sci. 35, 1352–1389.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Misse, A., De Diego-Balaguer, R., 2009. Neurophysiological mechanisms involved in language learning in adults. Philos. Trans. R. Soc. B - Biol. Sci. 364, 3711–3735.

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

Sanders, L.D., Newport, E.L., Neville, H.J., 2002. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nat. Neurosci. 5, 700–703.

Sebastián-Gallés, N., Martí, M.A., Cuetos, F., Carreiras, M.F., 2000. LEXESP: Léxico informatizado del español [LEXESP: A Computerized Word Pool in Spanish]. Edicions de la Universitat de Barcelona, Barcelona.

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

Schön, D., François, C., 2011. Musical expertise and statistical learning of musical and linguistic structures. Front. Psychol. 2, 167.

Starr, A., Aguinaldo, T., Roe, M., Michalewski, H.J., 1997. Sequential changes of auditory processing during target detection: motor responding versus mental counting. Electroencephalogr. Clin. Neurophysiol. 105, 201–212.

van Boxtel, G.J., Böcker, K.B.E., 2004. Cortical measures of anticipation. J. Psychophysiol. 18, 61–76.

Yu, C., Zhong, Y., Fricker, D., 2012. Selective attention in cross-situational statistical learning: evidence from eye tracking. Front. Psychol. 3, e148.