Journal of Memory and Language 60 (2009) 351–367

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words

Ansgar D. Endress [a,*], Jacques Mehler [b]

[a] Department of Psychology, Harvard University, William James Hall 984, 33 Kirkland Street, Cambridge, MA 02138, USA
[b] International School for Advanced Studies, Trieste, Italy

ARTICLE INFO

Article history:
Received 22 January 2008
Revision received 12 October 2008
Available online 4 February 2009

Keywords:
Statistical learning
Transition probabilities
Word segmentation
Cue integration
Perceptual cues to word learning

ABSTRACT

Word-segmentation, that is, the extraction of words from fluent speech, is one of the first problems language learners have to master. It is generally believed that statistical processes, in particular those tracking "transitional probabilities" (TPs), are important to word-segmentation. However, there is evidence that word forms are stored in memory formats differing from those that can be constructed from TPs, i.e. in terms of the positions of phonemes and syllables within words. In line with this view, we show that TP-based processes leave learners no more familiar with items heard 600 times than with "phantom-words" not heard at all if the phantom-words have the same statistical structure as the occurring items. Moreover, participants are more familiar with phantom-words than with frequent syllable combinations. In contrast, minimal prosody-like perceptual cues allow learners to recognize actual items. TPs may well signal co-occurring syllables; this, however, does not seem to lead to the extraction of word-like units. We review other, in particular prosodic, cues to word-boundaries which may allow the construction of positional memories while not requiring language-specific knowledge, and suggest that their contributions to word-segmentation need to be reassessed.

Introduction

Speech comes as a continuous signal, with no reliable cues to signal word boundaries. Thus learners have not only to map the words of their native language to their meanings (which is in itself a difficult problem), but first they have to identify the sound stretches corresponding to words. Thus, they need mechanisms that allow them to memorize the phonological forms of the words they encounter in fluent speech. Here we ask what kinds of memory mechanisms they can employ for this purpose. It is generally accepted that statistical computations are well suited for segmenting words from fluent speech, and thus for memorizing phonological word-candidates (e.g., Aslin, Saffran, & Newport, 1998; Cairns, Shillcock, Levy, & Chater, 1997; Elman, 1990; Goodsitt, Morgan, & Kuhl, 1993; Hayes & Clark, 1970; Saffran, 2001b; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Swingley, 2005). However, as reviewed below in more detail, there is evidence, in particular from speech errors, that memory for words in fact appeals to different kinds of memory mechanisms, namely those encoding the *positions* of phonemes or syllables within words. We thus ask whether learners extract word-like units from fluent speech when just the aforementioned statistical cues are given, or whether they require other, possibly prosodic, cues that allow them to construct positional memories. Specifically, we presented participants with continuous speech streams containing statistically defined "words". These words were chosen such that there were statistically matched "phantom-words" that, despite having the same statistical structure as words, never occurred in the speech streams. If statistical cues lead to the extraction of words from fluent speech, participants should know that they have encountered words but not phantom-words during the speech streams. In contrast, if memory for words is positional, participants should fail to prefer words to

* Corresponding author. Fax: +1 617 495 3886.
E-mail address: ansgar.endress@m4x.org (A.D. Endress).

0749-596X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2008.10.003

phantom-words when only statistical information is given. Rather such a preference should arise only once cues are available that lead to the construction of positional memories.

Evidence for co-occurrence statistics as cues to word boundaries

Once they reach a certain age, learners can use many different cues to predict word boundaries (e.g., Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Cutler & Norris, 1988; Dahan & Brent, 1999; Jusczyk, Cutler, & Redanz, 1993; Mattys & Jusczyk, 2001; Shukla, Nespor, & Mehler, 2007; Suomi, McQueen, & Cutler, 1997; Thiessen & Saffran, 2003; Vroomen, Tuomainen, & de Gelder, 1998). However, many of these cues are language-specific, and thus have to be learned. For instance, if learners assume that strong syllables are word-initial, they will be right in Hungarian but wrong in French (where strong syllables are word-final), and to learn where stress falls in a word, they have to know the words in the first place. Hence, at least initially, language learners need to use cues to word-boundaries that do not require any language-specific knowledge.

Co-occurrence statistics such as transitional probabilities (TPs) among syllables are one such cue that is particularly well-attested. These statistics indicate how likely it is that two syllables will follow each other. More formally, TPs are conditional probabilities of encountering a syllable after having encountered another syllable. Conditional probabilities like $P(\sigma_{i+1} = \text{pet}|\sigma_i = \text{trum})$ (in the word trumpet) are high within words, and low between words (σ denotes a syllable in a speech stream). Dips in TPs may give cues to word boundaries, while high-TP transitions may indicate that words continue. That is, learners may postulate word boundaries between syllables that rarely follow each other. Saffran and collaborators (e.g., Aslin et al., 1998; Saffran et al., 1996) have shown that even young infants can deploy such statistical computations on continuous speech streams. After familiarization with speech streams in which dips in TPs were the only cue to word boundaries, 8-month-old infants were more familiar with items delimited by TP dips than with items that straddle such dips. Even more impressively, after such a familiarization, infants recognize the items delimited by dips in TPs in *new* English sentences pronounced by a *new* speaker (Saffran, 2001b), suggesting that TP-based segmentation procedures may lead infants to extract word-like units.

Results such as these have led to the widespread agreement that co-occurrence statistics are important for segmenting words from speech. Though not thought to be the only cues used for word-segmentation, they are thought to play a particularly prominent role because, unlike other cues, they can be used by infants without any knowledge of the properties of their native language (e.g., Thiessen & Saffran, 2003).[1] Moreover, similar computations have been observed with other auditory and visual stimuli (Fiser & Aslin, 2002; Saffran, Johnson, Aslin, & Newport, 1999; Turk-Browne, Jungé, & Scholl, 2005), and with other mammals (Hauser, Newport, & Aslin, 2001; Toro & Trobalón, 2005). Such computations may thus be domain- and species-general, stressing again the potential importance of such processes for a wide array of cognitive learning situations. Accordingly, some authors have proposed that these processes may be crucial not only for word-learning but also for other, more grammatical aspects of language acquisition (Bates & Elman, 1996; Saffran, 2001a; Thompson & Newport, 2007).

Surprisingly, however, there is no evidence that TP-based computations lead to the extraction of word-candidates. The experiments above have provided numerous demonstrations that participants are more familiar with items with stronger TPs than with items with weaker TPs. This, however, does not imply that the items with stronger TPs are represented as actual word-like units, or even that they have been extracted. For example, one may well find that a piece of cheese is more associated with a glass of wine than with a glass of beer, but this does not imply that the wine/cheese combination is represented as a unit for parsing the visual scene. Likewise, choosing items with stronger TPs (where the syllables have stronger associations) over items with weaker TPs does not imply either that the items with stronger TPs have been extracted as perceptual units.

The distinction between a preference for high-TP items and representing these items as perceptual units is well illustrated in Turk-Browne and Scholl (2009) studies of visual statistical learning. In these experiments, participants saw a continuous sequence of shapes. This sequence was composed of a concatenation of three-shape items (just as the experiments reviewed above used concatenations of three-syllable non-sense words). Following such a familiarization, participants were as good at discriminating high-TP items from low-TP items when the items were played forward (that is, in the temporal order in which they had been seen during familiarization) as when they were played backwards. If a preference for high-TP items implied that these items have been extracted and memorized, one would have to conclude that participants have extracted also the backwards items although they had never seen them. It thus seems that a preference for high-TP items does not imply that these items have been memorized.

There are also other reasons to doubt that TPs may play an important role in word-segmentation. One reason is that computational studies using TPs (or related statistics) for segmenting realistic corpora of child-directed speech have encountered mixed success at best (e.g., Swingley, 2005; Yang, 2004). At minimum, TPs thus have to be complemented with other cues. This seems highly plausible, given that one would certainly not expect a single cue to solve a highly complex problem such as speech-segmentation.

While the poor performance of word-segmentation mechanisms based on TPs can be improved by the inclusion of other cues, there is a second, more fundamental, reason for doubting that TPs play an important role in word-segmentation. This reason is related to the kinds of

[1] Other speech segmentation models track "chunks" that occur in the input (e.g., Batchelder, 2002; Perruchet & Vinter, 1998). However, as these models have received less experimental attention and make the same predictions for the purposes of the current experiments as the transitional probability-based models, we will not discuss them further.

representations that are formed of acoustic word-forms. Presumably, the purpose of word-segmentation is to store phonological word-candidates in long-term memory. As these are essentially sound sequences (or sequences of articulatory gestures according to a direct realist perspective), it is reasonable to ask whether research on sequential memory can constrain the kinds of cues that can be used for word-segmentation. This issue is addressed in the next section.

Memory mechanisms for acoustic word forms

Research on sequential memory has revealed (at least) two kinds of mechanisms for remembering sequences (for a review, see e.g., Henson, 1998). One mechanism is referred to as "chaining memory." When memorizing the sequence ABCD using such a mechanism, one would learn that A goes to B, B to C, and C to D. In other words, this mechanism is fundamentally similar to TPs. There is another mechanism, however, that appeals to the sequential *positions* of items. For example, people often remember the first and the last element of a sequence — but not the intervening items. Chaining memories do not easily account for such results — because the "chain" is broken in the sequence middle. Positional mechanisms, in contrast, readily account for such results: People may memorize the item that occurred in the first and the last position without remembering items in intervening positions. These (and, in fact, many other) results are thus readily explained if a distinction between positional and chaining memories is assumed (e.g., Conrad, 1960; Henson, 1998; Henson, 1999; Hicks, Hakes, & Young, 1966; Ng & Maybery, 2002; Schulz, 1955). This distinction has also been observed in artificial grammar learning experiments. In such experiments, TPs and positional regularities seem to require different kinds of cues, to have different time courses, and to break down under different conditions (Endress & Bonatti, 2007; Endress & Mehler, in press; Peña, Bonatti, Nespor, & Mehler, 2002). In these experiments, participants were familiarized to speech streams. The streams contained both chaining and positional regularities. Following familiarization, participants had to choose between items that instantiated the chaining regularity, the positional regularity or both. Most relevant to the current experiments, participants were sensitive to the positional regularity only when the familiarization stream contained prosodic-like cues such as silences between words. TPs, in contrast, were tracked also in the absence of such cues. It thus appears that both positional and chaining memories can be learned from speech streams by independent mechanisms, but that positional memories require additional, perhaps prosodic cues.

Interestingly, a similar distinction between positional and chaining information has been proposed in artificial grammar learning experiments in the tradition developed by Miller (1958) and Reber (1967,1969), (although these experiments typically use *simultaneously* presented letter *strings* rather than sequences). In such experiments, participants are exposed to consonant strings governed by a finite-state grammar, and then have to judge whether new strings are grammatical. It now seems clear that participants acquire distributional information of the consonants of various kinds, including legal bigrams (which, we would argue, corresponds to chaining information; see e.g., Cleeremans & McClelland, 1991; Dienes, Broadbent, & Berry, 1991; Kinder, 2000; Kinder & Assmann, 2000) and the *positions* of legal letters and bigrams within the strings (which may correspond to the positional information mentioned above; see e.g. Dienes et al., 1991; Johnstone & Shanks, 1999; Shanks, Johnstone, & Staggs, 1997, but see Perruchet & Pacteau, 1990). Whilst these experiments were not necessarily optimized to distinguish chaining and positional information, it is interesting to note that a similar distinction has also been proposed in this literature.

What kinds of memory mechanisms are used for words? There is some evidence from speech errors that word memory has at least a strong positional component. With the tip of the tongue experience, for instance, people often remember the first and the last phoneme of a word, but not the middle phonemes (e.g., Brown & McNeill, 1966; Brown, 1991; Kohn et al., 1987; Koriat & Lieblich, 1974; Koriat & Lieblich, 1975; Rubin, 1975; Tweney, Ryan, Tkacz, & Zaruba, 1975). Such observations are hard to explain if memory for words relies upon chaining memories, since such chains would be broken in the middles of words. In contrast, they naturally follow if one assumes that words rely on positional memories. Likewise, spoonerisms (that is, reversals in the order of phonemes such as in "**qu**eer old **d**ean", from "**d**ear old **qu**een") often conserve the serial position in words and syllables of the exchanged phonemes (e.g., MacKay, 1970). Again, this would be unexpected if words were remembered by virtue of chaining memories (because positions are not encoded in such memories), but it is easily explained if word memory has a positional component.

If memory for acoustic word forms is positional, cues to chaining memories such as TPs may not enable participants to extract words from fluent speech. Rather, learners may require other cues such as those that have triggered positional computations in other artificial language learning studies (Endress & Bonatti, 2007). Here, we thus return to the original motivation for TP-based processes, and examine their potential for the first step in word-learning, namely word-segmentation. (In the following, we will use word-learning and word-segmentation interchangeably. We thus hypothesize that the role of a word-segmentation mechanism is to provide candidates for phonological word forms, but are agnostic as to how such forms may become linked to meaning.) At the very least, if TP-based learning mechanisms are used for word-learning, one would expect the output of these mechanisms (that is, presumably phonological word candidates) to make learners more familiar with items they heard frequently than with items they never heard at all. After all, a word-segmentation mechanism should learn the words contained in its input, and not some syllable combination it has never encountered at all.

The current experiments

To test whether co-occurrence statistics would lead to the extraction of word-like units from fluent speech, we

used standard TP-learning procedures with adults. Participants were told that they would listen to a monologue in an unknown language (in "Martian"). They were then familiarized with a continuous speech stream. This stream was a monotonous concatenation of nonce words (hereafter "words") with no pauses between them. Six words were concatenated such that TPs among their syllables were identical to TPs among syllables of particular "items" that did not occur in the streams (i.e. "phantom-words"). Phantom-words are items that, even though they did not occur in the stream, could become familiar to learners who only track pairwise TPs among syllables (see Fig. 1 and "Materials and method" of Experiment 1a for the construction of words and phantom-words).

At test, participants had to choose which of two items was more likely to be a Martian word. One item was a word and the other a phantom-word. If participants extracted word-like units through TP computations, they should prefer words to phantom-words even though the TPs were the same (since words are units that occurred in the stream while phantom-words are not). In terms of word-learning, one would expect participants to store only words they actually have encountered, and not all possible combinations of syllables that occurred together in other words.

To assess whether participants tracked the statistical structure of the speech streams, we also asked them to choose between words and part-words. Part-words occurred in the stream, but straddled a word boundary; TPs between syllables in words are thus higher than in part-words. We thus expected to replicate the standard finding that participants prefer words to part-words (e.g., Aslin et al., 1998; Saffran et al., 1996; Saffran et al., 1996).

In Experiments 1a through 1d, participants were exposed to speech streams whose durations ranged from 5 to 40 min. Words in these streams had the same TPs as some non-occurring phantom-words. If participants can use TP-information for extracting words from speech, they should be more familiar with words than with phantom-words. The familiarization in Experiment 2 was the same as in Experiment 1, but participants had then to choose between phantom-words and part-words; we asked whether participants would be more familiar with phantom-words even though they did not occur in the stream. Experiments 3 and 4 were identical to Experiment 1, except that participants were given additional cues to word boundaries. These cues were 25 ms silences between words, and a lengthening of the final syllable in each word, respectively.

Of course, results with adult participants do not automatically hold also for infants. However, statistical learning experiments have never revealed a difference between adults and infants (except that adults can learn more words), and some of our manipulations (such as presenting participants with a 40-min stream) are just not feasible for infants. That is, infant language processing is clearly different from that of adults in many ways (e.g., Werker & Tees, 1984). However, such differences have never been observed in statistical learning experiments on speech segmentation. Moreover, our crucial result will be that adults *cannot* keep track of the items they have heard if the foils have the same TP-structure. As there is no reason to assume that the memory abilities of infants are more sophisticated than those of adults, it seems reasonable to assume that our results would hold also for infants, but it is an important topic for future studies to test whether this assumption actually holds.

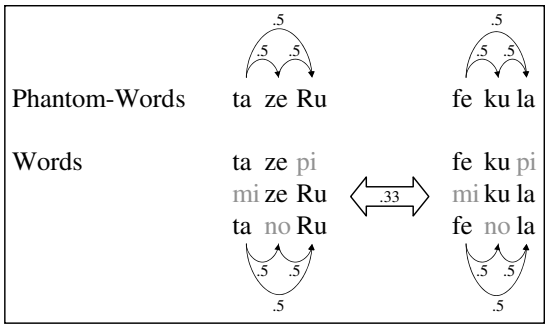

Fig. 1. Design of the current experiments. Participants were familiarized with continuous speech consisting of a concatenation of nonce words. These "words" were chosen such that TPs among syllables in words would be identical to TPs among syllables in "phantom-words", that is, in items that did not occur in the stream but had the same TPs as words. For each of the two phantom-words, there was a word sharing the first and the second syllable, a word sharing the second and third syllable, and a word sharing the first and third syllable. (The syllable that is *not* shared between a word and the corresponding phantom-word is printed in light gray characters in the figure.) In this way, TPs among adjacent and non-adjacent syllables within words and phantom-words were 0.5, and TPs among syllables across words 0.33.

Experiment 1a: Word-learning with 5-min exposure

Materials and method

Participants

Fourteen native speakers of Italian (7 women, 7 men, mean age 23.4, range 20–27) took part in this experiment.

Materials

The stimuli were synthesized with the MBROLA speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & van der Vreken, 1996), using the fr2 diphone database. (Pilot tests with native participants showed that Italian native speakers find synthesized speech with the fr2 voice more intelligible than speech synthesized with the available Italian diphone databases. We thus decided to use fr2. Obviously, all phonemes we selected also exist in Italian.) To avoid direct cues to word onsets, the stream was synthesized with increasing and decreasing amplitude ramps in the first and last 5 s, respectively. This ensured that the stream did not fade either in or out at any point corresponding to either words or part-words. Words had a mean length of 696 ms (mean phoneme duration 116 ms), and a fundamental frequency of 200 Hz. Test items were synthesized in the same way.

Pretraining

Before starting the experiment, participants completed a pre-training phase to get familiarized with the response keys. The pretraining consisted of 10 trials. In each trial, participants heard two syllables, one of which was 'so'.

Their task was to indicate whether 'so' was the initial or the final syllable. 'So' was the first syllable in half of the trials, and the second one in the other half.

Familiarization

Upon completion of the pretraining phase, participants were told that they would listen to a monologue in an unknown language ("Martian"), and were instructed to try to find the words in the monologue. This monologue was a concatenation of six trisyllabic nonce words (henceforth called just "words"). In the experiments presented here, words were constructed such that there would be items that would not appear in the streams, but that would have exactly the same transitional probabilities (TPs) between the first and the second, the second and the third, and the first and the third syllable as the words; for mnemonic purposes, we call these items phantom-words.

For constructing the words, we first selected two phantom words, and then chose the actually occurring words accordingly (see Fig. 1). For the phantom-word 'tazeRu', we included in the stream the three words 'tazeX', 'YzeRu' and 'taZRu', where X, Y, Z were the syllables /pi/, /mi/, and /no/, respectively; for the phantom-word 'fekula', we included the three words 'fekuX', 'Ykula' and 'feZla' with the same X, Y, and Z syllables. In this way, TPs between adjacent or non-adjacent syllables within words (and phantom-words) were 0.5, and TPs across word boundaries were 0.33 on average (range: 0.28–0.39). Each of the six resulting words occurred 75 times in the stream. Words in the stream occurred in random order, with the constraint that the same word could not occur twice in succession.

Test

After this familiarization, participants were presented with pairs of items, and were instructed to choose the one that sounded more Martian. Pairs could be of two types. In the first one, participants had to choose between words and phantom-words. Words and phantom-words could overlap either in their first and second, their first and third, or their second and their syllable; each overlap type was represented equally in the test pairs.

In the remaining trials, participants had to choose between words and part-words. Part-words could be of two types that were equally represented in the test pairs; they could either comprise one syllable of the first word and two syllables the next word (type CAB, if the speech stream is represented as a word sequence ABCABCABC...), or of two syllables of the first word and one syllable of the second word (type BCA). Most part-words shared two syllables with the word they were presented with. All test pairs are shown in Appendix A.

There were 6 word/phantom-word pairs, and 12 word/part-word pairs; each pair was presented twice in different word orders. Words and phantom-words appeared equally often during test. The resulting 36 test trials were administered in random order.

Validation of the materials

To make sure that our results would not be biased by item-specific factors, we familiarized 14 participants from the same pool as in the experiments reported below with a stream comprising the same syllables as in Experiment 1a, but arranged in random order. Following this familiarization, they completed the same test phase as in Experiment 1a. Words were preferred neither to phantom-words (preference for words: $M = 50.6\%$, $SD = 14.0\%$), $t(13) = 0.16$, $p = .876$, Cohen's $d = 0.042$, $CI_{.95} = 42.5\%, 58.7\%$, ns, nor to part-words, ($M = 56.8\%$, $SD = 19.0\%$), $t(13) = 1.35$, $p = .20$, Cohen's $d = 0.36$, $CI_{.95} = 45.9\%, 67.8\%$, ns. Below, we will report all statistical tests against a chance level of 50%. However, all significant tests would be significant also when tested relative to this validation experiment.

Results and discussion

As shown in Fig. 2, participants failed to prefer words to phantom-words (preference for words: $M = 50.6\%$, $SD = 15.1\%$), $t(13) = 0.1$, $p = .885$, Cohen's $d = 0.039$, $CI_{.95} = 41.9\%, 59.3\%$, ns; still, they tracked the TP-structure of the stream, as they preferred words to part-words ($M = 69.9\%$, $SD = 16.4\%$), $t(13) = 4.6$, $p < 0.001$, Cohen's $d = 1.2$, $CI_{.95} = 60.5\%, 79.4\%$. Hence, when two test items have the same TPs, participants do not seem to track whether an item actually occurred or not, and are as familiar with unheard phantom-words as with actual items.

Before concluding that TP computations did not allow participants to extract any word-candidates, we need to control for a number of confounds. First, many part-words in Experiment 1a contained syllables with identical vowels, which was never the case for words or phantom-words. Possibly, participants may reject part-words based on these vowel repetitions. Since phantom-words do not contain such identical vowels, they cannot be rejected on

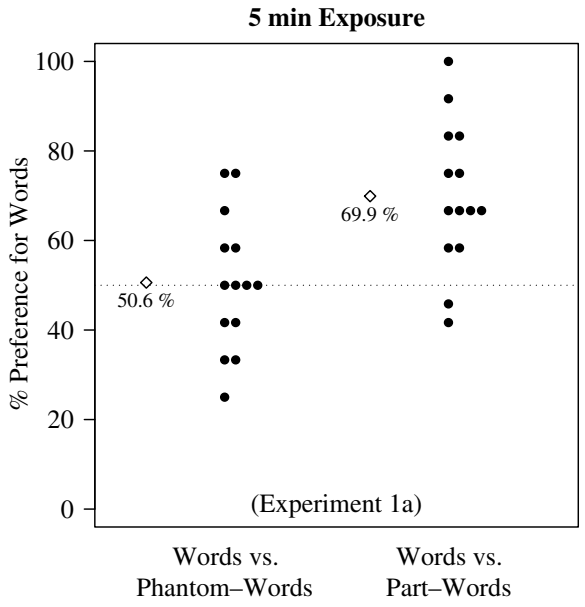

Fig. 2. Results of Experiment 1a. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. After a familiarization with a 5-min stream constructed according to Fig. 1, participants preferred words to part-words, but had no preference for words over phantom-words.

this basis. Hence, if participants' choices were based exclusively on the presence of identical vowels, one would expect a preference for words over part-words but not over phantom-words. Experiment 1b controls for this possibility by using stimuli such that words, phantom-words and part-words all contain identical vowels. If the vowel-repetitions are crucial to the results of Experiment 1a, we should expect participants to fail both on the word/phantom-word and the word/part-word comparison. (In pilot experiments, we controlled for this possibility also by eliminating vowel repetitions both in words and in part-words (see Appendix B for the materials). We replicated the results of Experiment 1a also in that case. Pooling over the different familiarization durations (ranging from 2 to 40 min), there was no preference for words to phantom-words ($M = 52.47\%$, $SD = 19.6\%$), $t(63) = 1.01$, $p = .316$, Cohen's $d = 0.13$, $CI_{.95} = 47.59\%, 57.36\%$, ns, but a preference for words over part-words ($M = 58.73\%$, $SD = 15.4\%$), $t(52) = 4.14$, $p = .0001$, Cohen's $d = 0.57$, $CI_{.95} = 54.49\%, 62.96\%$.)

Second, our familiarization stream comprised only six highly similar words; this may have made the word-segmentation unduly difficult. The overlap among words could have interfered with word segmentation in two different ways. First, participants could experience difficulties encoding TPs during familiarization. If this were the case, however, they should fail also on the word/part-word test pairs. Second, the overlap between *test items* may have led participants to confuse these items during the *test phase*. However, in each test pair, words shared two syllables with both phantom-words and part-words; if the failure to prefer words over phantom-words were due to this overlap and the resulting confusion during the test phase, one would expect a failure also for the word/part-word discriminations — since the overlap was exactly the same for both types of test pairs. We can thus reject the possibility that our results are due to excessive overlap among words or test items. Moreover, participants do not seem to experience difficulties with very similar materials such as in the experiments reported by Endress and Bonatti (2007) or Peña et al. (2002), where words and test items were equally similar. Still, even though these results suggest that it is highly unlikely that participants' failure to reject phantom-words is due to the similarity of words employed in our experiments, we further control for this possibility in Experiment 1c. In that experiment, we double the number of words in the stream; thus, for each word, there were six other words with no overlap at all. In addition to controlling for the overlap among words, we can also replicate the results of Experiments 1a with new materials and a familiarization stream with a different statistical structure.

Finally, the familiarization used in Experiment 1a may be just too short for participants to extract words, and participants would prefer words to phantom-words with more exposure. In Experiment 1d, we tested this possibility by exposing participants to 8 repetitions of the stream used in Experiment 1a, leading to a 40-min familiarization with 600 repetitions of each word. If the failure to discriminate words from phantom-words in Experiment 1a was due to insufficient exposure, we should observe successful discrimination under these conditions.

Experiment 1b: Word-learning with 14-min exposure and different words

Materials and methods

Experiment 1b was similar to Experiment 1a except that both words and part-words could have syllables with identical vowels. The words were *bagadu*, *togaso*, *bapiso*, *limudu*, *tomufe* and *lipife*; the phantom-words were *bagaso* and *limufe*. Moreover, each word was presented 200 times during familiarization (rather than 75 times as in Experiment 1a), yielding a 14-min familiarization. All test items are shown in Appendix C. 14 native speakers of Italian (12 women, 2 men, mean age: 23.4, range 19–33) took part in this experiment.

Results and discussion

The results of this experiment are shown in Fig. 3. As in Experiment 1a, participants failed to discriminate between words and phantom-words ($M = 54.8\%$, $SD = 15.9\%$), $t(13) = 1.12$, $p = .283$, Cohen's $d = 0.3$, $CI_{.95} = 45.6\%, 64.0\%$, ns, but they preferred words to part-words ($M = 81.0\%$, $SD = 15.0\%$), $t(13) = 7.75$, $p<0.001$, Cohen's $d = 2.1$, $CI_{.95} = 72.3\%, 89.6\%$. Neither the performance on word/phantom-word pairs nor that on word/part-word pairs differed from that observed in Experiment 1a, $F(1,26) = 0.5$, $p = .484$, $\eta^2 = 0.02$, ns, and $F(1,26) = 3.5$, $p = .0744$, $\eta^2 = 0.12$, ns, respectively. It thus seems safe to conclude that the presence of identical vowels in the part-words of Experiment

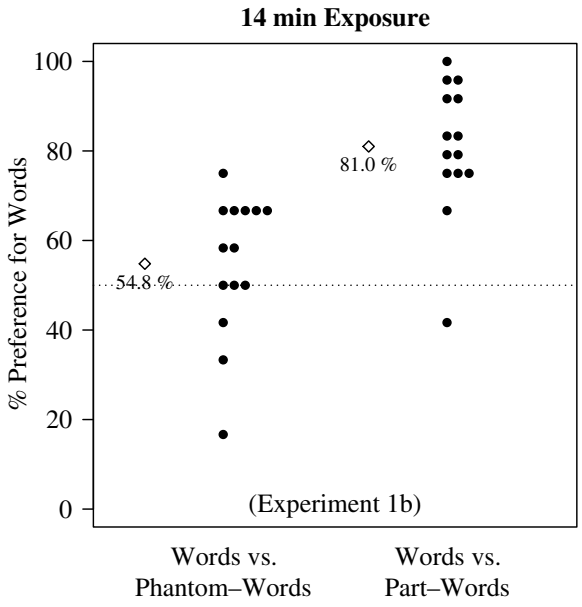

Fig. 3. Results of Experiment 1b. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. After a familiarization with a 14-min stream constructed according to Fig. 1, participants preferred words to part-words, but had no preference for words over phantom-words.

1a does not explain the failure to prefer words to phantom-words.

Experiment 1c: Word-learning with more variable words

Materials and methods

Experiment 1c was similar to Experiment 1a except that 12 words were used instead of 6. That is, we employed *two* sets of six words with the same statistical properties as the six-word set in Experiment 1a. We thus obtained 12 words in total (*paseti*, *Rosenu*, *pamonu*, *lekati*, *Rokafa*, *lemofa*, *bodesa*, *ludegi*, *bonagi*, *muRisa*, *luRivo* and *munavo*) and four phantom-words (*pasenu*, *lekafa*, *bodegi* and *muRivo*). TPs within words were 0.5; TPs across words were 0.187 on average (range: 0.033–0.313). All test pairs are shown in Appendix D. 14 native speakers of Italian (8 women, 6 men, mean age: 22.1, range 18–26) took part in this experiment.

Results and discussion

The results of this experiment are shown in Fig. 4. Again, participants failed to discriminate between words and phantom-words ($M = 49.4\%$, $SD = 14.8\%$), $t(13) = 0.15$, $p = .883$, Cohen's $d = 0.04$, $CI_{.95} = 40.9\%, 57.9\%$, ns, but they preferred words to part-words ($M = 57.7\%$, $SD = 10.6\%$), $t(13) = 2.74$, $p = .017$, Cohen's $d = 0.73$, $CI_{.95} = 51.6\%, 63.8\%$. While the performance on word/phantom-word pairs was not different from that in Experiment 1a, $F(1,26) = 0.04$, $p = .835$, $\eta^2 = 0.002$, ns, participants in Experiment 1a performed better on the word/part-word trials than in Experiment 1c, $F(1,26) = 5.50$, $p = .027$, $\eta^2 = 0.17$. The decrease in performance on the latter trials probably reflects the increased number of words. Importantly, however, the performance on the word/phantom-word pairs did not improve although the overlap among items was reduced. In line with previous work (Endress & Bonatti, 2007; Peña et al., 2002), we can thus rule out that the pattern of results in Experiment 1a was due to excessive overlap among words.

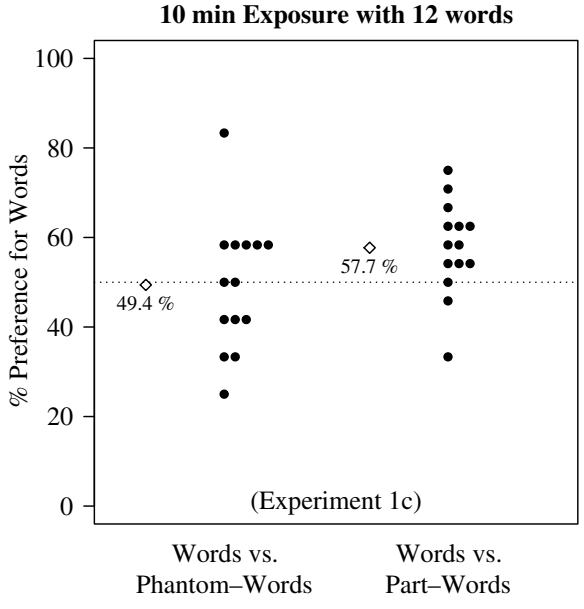

Fig. 4. Results of Experiment 1c. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. After a familiarization with a 10-min stream constructed according to Fig. 1 but with 12 words instead of six, participants preferred words to part-words, but had no preference for words over phantom-words.

Experiment 1d: Word-learning with 40-min exposure

Experiments 1a through 1c suggest that participants fail to prefer word over phantom-words also after controlling for a number of possible confounds. This is consistent with the hypothesis that participants do not extract word-candidates through TP computations. Another simple explanation, however, is that the familiarizations used in these experiments were simply too short for participants to extract words, and participants may require more exposure to prefer words over phantom-words. In addition to its a priori plausibility, this possibility receives support from previous experiments (Endress & Bonatti, 2007; Peña et al., 2002). In those experiments, participants were familiarized with a speech stream and had to choose between part-words and items that never occurred in the speech stream (but instantiated a regularity that is irrelevant for the current purposes). Participants preferred part-words only after long exposures, suggesting that TP-based statistical processes get consolidated over time. In Experiment 1d, we thus ask whether also a preference for words over phantom-words requires more exposure. We familiarized participants with eight repetitions of the stream used in Experiment 1a, leading to a 40-min familiarization with 600 repetitions of each word. We also increased the statistical power of this experiment by doubling the number of participants.

Materials and methods

Experiment 1d was identical to Experiment 1a except that the familiarization stream was presented 8 times, yielding a 40-min familiarization. 28 native speakers of Italian (19 women, 9 men, mean age: 21.3, range 18–26) took part in this experiment. One additional participant was tested but excluded from analysis because he turned the volume off during the speech stream.

Results and discussion

The results of this experiment are shown in Fig. 5. Even after such a massive exposure, participants failed to discriminate between words and phantom-words ($M = 44.6\%$, $SD = 20.4\%$), $t(27) = 1.4$, $p = .177$, Cohen's $d = 0.26$, $CI_{.95} = 36.7\%, 52.6\%$, ns, but they preferred words to part-words ($M = 71.6\%$, $SD = 16.6\%$), $t(27) = 6.9$, $p < .001$, Cohen's $d = 1.3$, $CI_{.95} = 65.1\%, 78.0\%$. The performance on neither type of trial differed from that observed in Experiment 1a, $F(1,40) = 1.0$, $p = .341$, $\eta^2 = 0.02$, ns and $F(1,40) = 0.1$, $p = .764$, $\eta^2 = 0.002$, ns, respectively. Hence, also a longer exposure does not seem to allow participants to extract word-like units when TPs are the only cues.

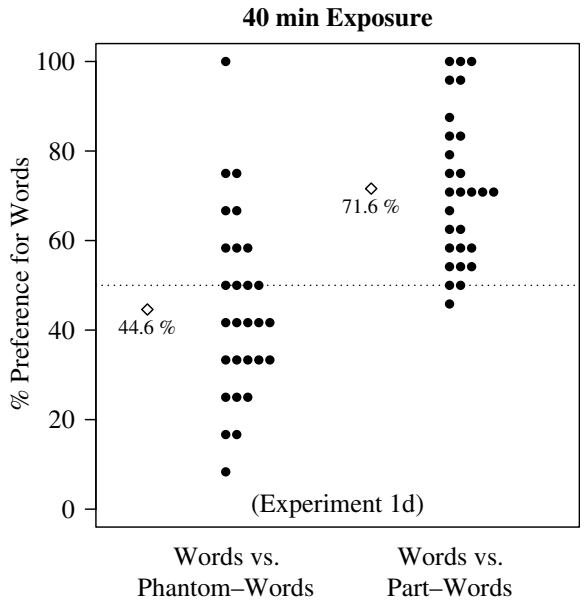

Fig. 5. Results of Experiment 1d. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. After a familiarization with a 40-min stream constructed according to Fig. 1, participants preferred words to part-words, but had no preference for words over phantom-words.

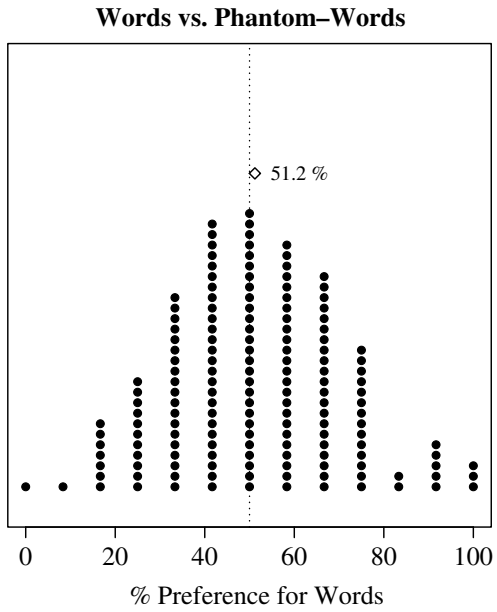

Fig. 6. Collapsed results of participants familiarized with streams with the same statistical structure as before, but different durations and syllables. Dots represent the means of individual participants, the diamond the sample average, and the dotted line the chance level of 50%. Even with 161 participants, no trend towards a preference for words with respect to phantom-words emerged.

We can also exclude that the failure to prefer words to phantom-words may have arisen due to insufficient statistical power. Surprised by the participants' failure to prefer words to phantom-words, we ran several additional experiments with the same statistical structure as in Experiments 1a but with different syllables and stream durations ranging from 2 to 40 min. (The most frequently used test items are shown in Appendix B.) We never observed any preference for words to phantom-words. The combined results are shown in Fig. 6. Even when collapsing all 161 participants who took part in the different experiments, no preference for words to phantom-words emerged (M = 51.2%, SD = 19.4%), $t(160) = 0.8$, $p = .438$, Cohen's $d = 0.061$, $CI_{.95}$ = 48.2%, 54.2%, ns; the preference for words compared to part-words, in contrast, was highly significant (M = 67.4%, SD = 17.5%), $t(149) = 12.2$, $p < .001$, Cohen's $d = 1.0$, $CI_{.95}$ = 64.6%, 70.3% (data not shown). (In some experiments, participants were not asked to choose between words and part-words; there are thus more participants comparing words to phantom-words than comparing words to part-words.)

These results suggest that participants learn the TP-structure of the streams. They may prefer words to part-words either by recognizing the high-TP bigrams in words, by rejecting the lower-TP bigrams in part-words, or by noting that syllables in words are more associated with each other than syllables in part-words (just as one may notice that a mouse is more associated with a piece of cheese than a cat). This ability, however, does not seem to allow them to extract any word-like units; participants are as familiar with words they have heard 600 times as with phantom-words they have not encountered at all. Apparently, their choices are based just on the frequency of co-occurrence of syllable pairs, and they do not consider at all the frequency of the overall, word-like unit. If so, they also should prefer phantom-words to part-words, even though part-words did occur in the stream. This was tested in Experiment 2.

Experiment 2: Can unheard phantom-words be preferred to actually encountered items?

Materials and methods

The familiarization in Experiment 2 was identical to the one from Experiment 1a. At test, participants were presented with two other kinds of trials. They had to choose between phantom-words and *part-words* (rather than between words and phantom-words as in Experiment 1a); the part-word types (see above) were equally represented in these test pairs. In the remaining trials, participants had to choose between words and part-words; again, the part-word types were equally represented in the test pairs.

There were 4 phantom-word/part-word pairs, and 12 word/part-word pairs; each pair was presented twice in different word orders. The resulting 32 test trials were administered in random order. All test pairs are shown in Appendix E.

Fourteen native speakers of Italian (10 women, 4 men, mean age: 23.6, range 19–36) took part in this experiment.

Results

As shown in Fig. 7, participants preferred phantom-words to part-words (M = 79.5%, SD = 26.2%), $t(13) = 4.2$, $p = .001$, Cohen's $d = 1.1$, $CI_{.95}$ = 64.3%, 94.6%. Again, they also preferred words to part-words (M = 71.4%, SD =

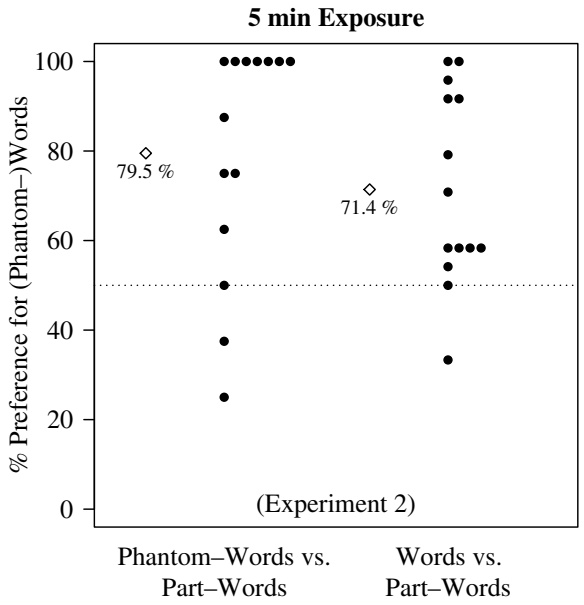

Fig. 7. Results of Experiment 2. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. After a familiarization with a 5-min stream constructed according to Fig. 1, participants preferred phantom-words to part-words, and words to part-words, even though phantom-words did not occur in the stream while words did.

21.5%), $t(13) = 3.7$, $p = .002$, Cohen's $d = 1.0$, $CI_{.95} = 59.0\%$, 83.8%, and the preference for phantom-words was not any stronger than that for words, $F(1,13) = 1.88$, $p > 0.2$, $\eta^2 = 0.03$, ns (repeated-measure ANOVA).

Discussion

In Experiment 2, participants preferred phantom-words to part-words although phantom-words never occurred in the stream. Such a result is expected if participants just track TPs among syllables without extracting any word-like units. In this case, they should be unable to distinguish between two items when their TPs are the same, and they should be more familiar with items they have not heard at all than with items that occurred frequently but that had lower TPs. This is exactly the pattern of results we found in Experiments 1 and 2. Such a result is problematic if TP-based computations have a prominent role in word-segmentation; after all, the units children store in the mental lexicon are presumably real words and not items they have never encountered at all. It is thus plausible to conclude that TP-based computations do not lead to the extraction of word-like units.

A possible conclusion of our experiment is that, when TPs are the only available cues, learners are limited to use "bigram statistics" that is, TPs between syllables *pairs*. While this seems to be a correct description of our results, it also seems to imply that TPs cannot be used to extract at least some words from fluent speech. Indeed, if learners have to acquire a trisyllabic word ABC (where the letters stand for the syllables), and their statistical machinery allows them just to track pair-wise relations such as AB, BC and A...C, then this machinery would simply not allow them to extract any trisyllabic unit. They would learn that A goes with B, that B goes with C, and that A goes also with C (at some distance). However, for extracting trisyllabic units, one would have to track how often all three syllables occur together, which goes beyond pair-wise relations. Moreover, if our hypothesis is correct that memories for word forms are positional, then bigram statistics would not lead to the extraction of units either. However, as our experiments do not use bisyllabic items, they do not directly speak to this issue.

While our results seem to suggest that TPs do not lead to the extraction of units from fluent speech, an alternative interpretation is that TPs may be such a strong cue to words that participants actually "*induce*" that they must have heard "words" due to their TP-structure even if they did not encounter them; indeed, participants are highly familiar with phantom-words, since they prefer them to part-words. Moreover, learners are sensitive to TP differences even if the frequency of the test items is matched (Aslin et al., 1998); TPs are thus known to have an important effect on participants' choices. Still, one of the hallmarks of psycholinguistic experiments is a sensitivity to word frequency (e.g., Cattell, 1886; Forster & Chambers, 1973; Solomon & Postman, 1952). If the words from the stream and the phantom-words were word-like entities, one would expect participants to track their frequency just as they do with real words. However, in our experiments, the participants' choices are not influenced *at all* by the frequency of the test items: They are as familiar with highly frequent words as with phantom-words that did not occur at all, suggesting that words from the stream and phantom-words may not behave as word-like units. Despite the general agreement that TP-based computations are crucial for word-learning, other cues seem to be required for actually extracting word-like units.

Still, there is another possibility why participants may not discriminate between words and phantom-words. For example, they may learn that words start with /ta/, /mi/ or /fe/. Possibly, participants may learn that words start and end with certain syllables. Phantom-words, however, start and end with the same syllables. If this were the case, one would thus expect them to be unable to choose between words and phantom-words. This possibility, however, is ruled out by Endress and Bonatti's (2007) finding that such positional information is not learned from continuous speech streams (as in the experiments presented here), but only when participants are familiarized with speech streams containing segmentation cues (that is, small silences between words). Since participants do not learn such positional information from continuous speech streams, their inability to discriminate between words and phantom-words cannot be due to their knowledge of positional information either.

A final interpretation of our results is that the preference for words over part-words may be based on the rejection of part-words — rather than on the endorsement of words. Indeed, if participants in statistical learning experiments just learn to reject part-words due to their low-TP transitions, one would expect a successful discrimination between words and part-words, but not between words and phantom-words; furthermore, phantom-words may be preferred to part-words since the latter are rejected.

This interpretation would thus reconcile the previous word segmentation literature with our results. Note, however, that also in this case, participants would not extract word-candidates due to TP computations, since they would just learn that certain syllable transitions cannot occur inside a word. To extract actual units, they thus would need to rely on other cues.

While this is a possible interpretation of our data and the previous literature, we believe that some results make it less plausible. First, in other experiments, participants *endorse* part-words, at least after long exposures (e.g., Endress & Bonatti, 2007; Endress & Mehler, in press; Peña et al., 2002). Since part-words are endorsed under some conditions, they thus cannot be intrinsically rejected, but only relative to the alternative choices of the two-alternative forced choice tasks usually employed in such experiments. Other cues, in contrast, seem to lead directly to the rejection of certain items. In Shukla et al.'s (2007) Experiment 5, for instance, items with high TPs but inconsistent with prosodic cues were rejected even in favor of *unheard* items. While we cannot rule out that part-words are rejected in our experiment, the rejection of part-words seems to be much milder than that of prosodic violations (since part-words are sometimes preferred). We thus suggest that a more plausible reason for which phantom-words are not dispreferred is that participants just compute pair-wise statistics among syllables without extracting any word-candidates.

Possibly, participants require cues other than TPs to tag some elements for extraction; this would certainly follow from the hypothesis that memory for acoustic word forms is positional in nature. Recall that there are (at least) two kinds of mechanisms for memorizing sequences such as words. One kind of memory consists of chaining memories fundamentally similar to TPs. There is another kind of memory, however, that appeals to the *position* of elements in a sequence (e.g., Henson, 1998). It seems that phonological word forms are, at least in part, encoded using the latter kind of memories. If so, participants should extract word candidates only when given appropriate cues for constructing positional memories. To illustrate this point, we will use a very subtle, possibly subliminal cue that has been shown to lead to positional memories in previous experiments (Endress & Bonatti, 2007; Peña et al., 2002), namely a 25-ms silence between words. In Experiment 3, participants were familiarized with the same 5 min stream as in Experiment 1a except that it contained 25-ms silences between words. If these cues are sufficient to signal to learners that the sound stretches delimited by these silences function as units, participants may prefer words to phantom-words.

Experiment 3: Word-learning with silences between words

Materials and methods

Experiment 3 was identical to Experiment 1a except that words in the familiarization stream were separated by 25-ms silences. 14 native speakers of Italian (7 women, 7 men, mean age: 24.6, range 19–35) took part in this experiment.

Results and discussion

As shown in Fig. 8, participants preferred words to phantom-words ($M = 66.1\%$, $SD = 18.0\%$), $t(13) = 3.3$, $p < .001$, Cohen's $d = 0.89$, $CI_{.95} = 55.7\%$, 76.5% (and also words to part-words, $M = 84.2$, $SD = 17.8\%$, $t(13) = 7.3$, $p < .001$, Cohen's $d = 2.0$, $CI_{.95} = 74.5\%$, 95.1%). An ANOVA comparing Experiments 1a and 3 as between-subjects factor and the trial type (words vs. phantom-words or part-words) as within-subject factor revealed main effects of experiment, $F(1,26) = 7.60$, $p = .011$, $\eta_p^2 = 0.23$, and trial type, $F(1,26) = 35.0$, $p < .0001$, $\eta_p^2 = 0.57$, but no interaction ($F < 1$). Separating words by very subtle silences was thus sufficient to enable participants to extract units.

While words in natural speech are clearly not separated by silences as in Experiment 3, cues leading to word extraction may be provided by prosody. Indeed, it has been shown that different prosodic cues modulate how speech is segmented (e.g., Jusczyk, Houston, & Newsome, 1999; Shukla et al., 2007; Thiessen & Saffran, 2003). Moreover, some prosodic cues may be universal across languages; learners thus may not need to acquire language-specific knowledge to use them. One such universal cue may be a lengthening of the final syllable of different units (e.g., Fon et al., 2002; Hoequist, 1983a; Hoequist, 1983b; Vassière, 1983). To illustrate that such cues are effective for word extraction, we familiarized participants with the same stream as in Experiment 1a, except that the final syllable of each word was lengthened by 50%. While previous research suggests that final lengthening aids in word segmentation (Saffran et al., 1996), here we ask whether it would allow participants to discriminate words from phantom-words.

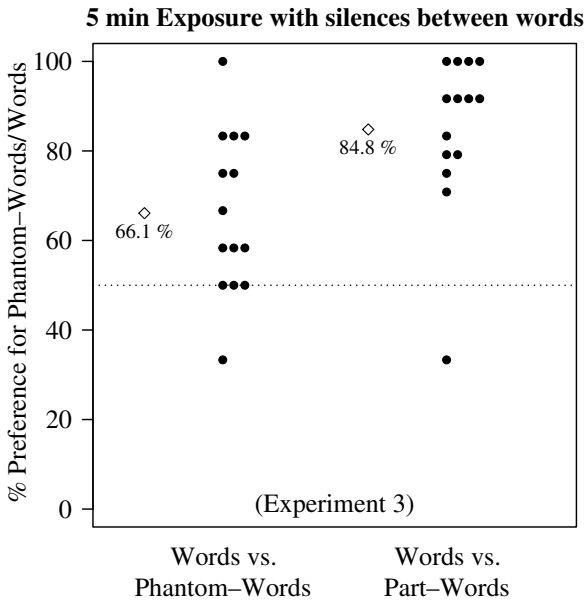

Fig. 8. Results of Experiment 3. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. After a familiarization with a 5-min stream (again constructed according to Fig. 1) in which words were separated by 25-ms silences, participants preferred words to both phantom-words and part-words.

Experiment 4: Word-learning with final lengthening

Materials and methods

Experiment 4 was identical to Experiment 1a except that the duration of the final vowel of each word was doubled to 232 ms during familiarization (corresponding to a lengthening of the final syllable by 50%); the test items were (physically) identical to those used in Experiment 1a. As the preference for words to phantom-words was marginal with 14 participants ($p = .048$), we added six participants to be sure that the results would remain stable. In total, 20 native speakers of Italian (15 women, 5 men, mean age: 22.4, range 18–27) took part in this experiment.

Results and discussion

As shown in Fig. 9, participants preferred words to phantom-words ($M = 69.2\%$, $SD = 22.6\%$), $t(19) = 3.8$, $p = 0.001$, Cohen's $d = 0.85$, $CI_{.95} = 58.6\%$, 79.8%, and also words to part-words ($M = 89.0\%$, $SD = 15.7\%$), $t(19) = 11.1$, $p < 0.001$, Cohen's $d = 2.5$, $CI_{.95} = 81.6\%$, 96.3%. An ANOVA comparing Experiments 1a and 4 as between-subjects factor and the trial type (words vs. phantom-words or part-words) as within-subject factor revealed main effects of experiment, $F(1,32) = 12.5$, $p = .001$, $\eta_p^2 = 0.28$, and trial type, $F(1,32) = 35.0$, $p < .0001$, $\eta_p^2 = 0.52$, but no interaction between these factors ($F < 1$).

Hence, like the silences in Experiment 3, prosodic cues such as final lengthening were effective for enabling participants to extract words. As the same kind of prosodic cues led to the extraction of positional memories in previous experiments (Endress & Bonatti, 2007), it thus seems reasonable to conclude that such memories are required for extracting word-like units from fluent speech.

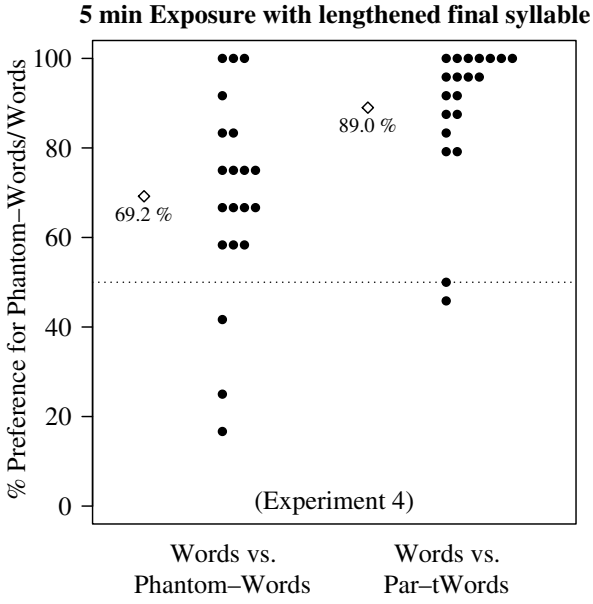

Fig. 9. Results of Experiment 4. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. After a familiarization with a 5-min stream (again constructed according to Fig. 1) in which the last syllable of each word was lengthened, participants preferred words to both phantom-words and part-words.

Interestingly, as evidenced by the lack of an interaction between experiment and trial type, lengthening the final syllable of words increased the preference for words over phantom-words and over part-words to the same extent; it thus seems that the inclusion of prosodic cues had an additive effect on both trial types. This result fits well with the idea that chaining information and positional information are tracked independently and in parallel in speech streams (e.g., Endress & Bonatti, 2007; Endress & Mehler, in press). Since words are favored to both phantom-words and part-words in terms of positional information, the availability of positional information should not only establish a preference for words over phantom-words, but it should also increase the preference for words over part-words. This is because positional information gives a second reason to prefer words over part-words in addition to their difference in TPs. The additive effect of the inclusion of prosodic cues may thus be taken as evidence that participants track chaining and positional memories in parallel by independent mechanisms.

The results of Experiments 3 and 4 raise three important problems. First, one may wonder how often phantom-words are likely to arise in naturalistic situations (e.g., in infant directed speech), and how relevant our results are for natural language acquisition. We believe, however, that this question applies equally to artificial language learning studies in general. Indeed, statistical learning experiments such as the ones presented here are usually performed with unnatural stimuli (e.g., with speech streams stripped of prosody). While such research has yielded numerous important insights, our results suggest that it will be important to test the cues that infants use for word-segmentation in more naturalistic conditions.

Second, the results of Experiments 1 and 2 suggest that one has to be cautious before concluding that participants extracted some units. Indeed, most previous statistical learning experiments concluded from a preference for words over part-words that participants had extracted words as units; the finding that words are not preferred to phantom-words under the same familiarization conditions suggests that participants may not have extracted any units at all. Mutatis mutandis, the same problem applies to the results of Experiments 3 and 4. Potentially, these results do not imply either that participants extracted any units even when prosody-like cues were given, and one may need still other familiarization conditions for successful extraction. We thus can conclude only that adding positional cues led participants closer to extracting units, as this reinstated at least a sensitivity to the frequency of the items — since participants preferred words to phantom-words. Further research is needed, however, to determine whether participants really extracted any units.

The third problem relates to our use of adult subjects for an experiment that is supposed to model language acquisition. While it is plausible to assume that also infants would fail to prefer words to phantom-words (since there

is no reason to think that the memory abilities of infants are any more sophisticated than those of adults), such an assumption is more problematic for Experiments 3 and 4. In fact, some prosodic cues such as stress seem to be used for segmentation only by relatively old infants (e.g., Thiessen & Saffran, 2003, but see, e.g., Johnson & Jusczyk, 2001). Infants thus have to learn about some of the prosodic properties of their native language before they can use them for acquiring language. If the cues used in Experiments 3 and 4 fall into this class of language-specific cues, it would be almost miraculous how infants could ever segment words from fluent speech: At least adult do not extract any units due to TP computations, while infants may not even be able to use prosodic-like cues such as those that made participants in our experiments prefer words to phantom-words.

In General discussion, we will discuss some prosodic cues that seem to be language-universal, and to which even neonates are sensitive. It thus seems that, at least in certain limits, a prosodic-based word-segmentation strategy may be feasible. Of course, it is likely to have important limitations, but one should also keep in mind that TP-based mechanisms yield a rather poor segmentation performance when tested in realistic settings (e.g., Swingley, 2005; Yang, 2004). Here, we just note that at least adults may extract acoustical word-candidates when they are given the possibility to construct positional memories, and that it will be important to test whether this is also true for infants.

General discussion

In the experiments presented here, we examined the potential of TP-based computations for word-segmentation. A basic prediction is that, if such computations are used for word-segmentation, they should make participants more familiar with items heard frequently than with unheard items. In contrast to this prediction, participants did not track at all whether or not they had encountered items when TPs in these items were matched; this remained true even after arbitrarily long periods of familiarizations. Moreover, participants were more familiar with unheard phantom-words than with frequently encountered part-words. When other cues (such as pauses between words, or the lengthening of word-final syllables) were given, in contrast, participants readily preferred words to phantom-words. Hence, to learn words from fluent speech, participants cannot predominantly rely on TPs but have to use also other cues. When such cues are given, learners track which items they actually heard; in the absence of such cues, however, learners seem to just compute associations among syllables without extracting any words.

A case for multiple-cue integration?

A possible interpretation of our results is that a single cue such as TPs is not sufficient for extracting auditory word-candidates; rather, learners may need multiple, converging cues for word-segmentation (e.g., Christiansen, Allen, & Seidenberg, 1998). While plausible, this interpretation of our results seems incorrect. The results provided by Shukla et al. (2007) are a case in point. These authors investigated how prosodic contours interact with statistical computations. They showed that participants do not recognize (and sometimes even reject) statistically well-formed items that straddle prosodic breakpoints. However, when participants were tested with *written* test items (after using the same auditory familiarization as before), they also recognized straddling items. Hence, participants must have also computed co-occurrence statistics across word boundaries (otherwise they could not recognize them in the written modality), but they seem to reject them because of their prosodic ill-formedness. Shukla et al. (2007) suggested that these results were due to two independent mechanisms: One mechanism may track co-occurrence statistics (such as TPs) irrespective of prosodic information. The other one may identify prosodically well-formed units that, we would argue, may provide word-candidates. Such results are rather problematic for a multiple-cue integration perspective. Indeed, when tested with auditory stimuli, learners seem to be guided by the prosodic cues irrespective of whether the co-occurrence statistics are compatible with them or not. When tested with visual stimuli (to which prosodic information apparently cannot be linked), participants respond exclusively on the basis of co-occurrence statistics irrespective of the prosodic information. In these experiments, participants thus responded predominantly based on prosodic information, while TPs played a secondary role at best.

In line with this conclusion, further research revealed that participants readily recognize items provided by prosodic cues even when statistical cues are made entirely uninformative (Endress & Hauser, under review). In these experiments, all syllable transitions had the same TPs, and the only cues to word boundaries were provided by prosody. Participants recognized the prosodically defined items irrespective of whether word-level prosody or sentence-level prosody was used, and irrespective of whether the prosody came from the participants' native language or an entirely unfamiliar language with radically different prosodic properties (that is, French, Hungarian and Turkish for monolingual English speakers). Hence, prosodic information alone was sufficient to make participants recognize certain items from fluent speech.

Another case in point for the conclusion that items can be recognized in the absence of TPs comes from the study of computations that can be performed on consonants and vowels. It seems that it is difficult or even impossible to track TPs over vowels (Bonatti, Peña, Nespor, & Mehler, 2005). (Whilst Newport & Aslin, 2004 proposed that TPs can also be computed over vowels, Bonatti et al., 2005 replicated their results, and showed that vowel TPs can only be tracked under particular conditions, namely when some vowels are repeated.) Toro, Mehler, Nespor & Bonatti (2008) then showed that, in the presence of prosodic-like cues, participants also recognize vowel sequences even though they do not compute TPs over vowels; that is, if "words" are defined by vowel sequences (such that consonants or syllables cannot be used to identify words), participants recognize these words when prosodic-like cues are

given even though they cannot compute TPs over vowels. It thus seems that prosodic and statistical cues draw on different kinds of mechanisms, and that prosodic but not statistical cues may readily identify word-candidates. Hence, the failure to prefer words over phantom-words in our experiments is unlikely to be due to the need for multiple cues to extract word candidates; rather, it seems that certain kinds of cues are required, and that co-occurrence statistics such as TPs are not among them.

That is, we are entirely open to the possibility that, in realistic settings, learners would use many different cues to extract words from fluent speech, including TPs. If memory for words is positional, however, certain kinds of cues will be *required* for word extraction, while others may be less important. In the next section, we will explore what kinds of cues may identify word-candidates from fluent speech, and then explain why prosodic cues may be particularly well-suited to fulfill this function.

What kinds of cues can be used for word-segmentation?

Presumably, learners have to memorize the output that the word-segmentation mechanisms provide, and thus need mechanisms to encode such sound-sequences. As mentioned in the introduction, there are (at least) two kinds of such mechanisms (e.g., Henson, 1998), namely "chaining" memories and "positional" memories. (Technically, what we call "positional" memory is called "ordinal" memory by Henson (1998), but this distinction is irrelevant for present purposes.) While chaining memories are fundamentally similar to TPs, the other mechanism may encode the sequential *positions* of phonemes or syllables within words. Moreover, evidence from speech errors suggests that word memory has at least a strong positional component (e.g., Brown & McNeill, 1966; Brown, 1991; Kohn et al., 1987; Koriat & Lieblich, 1974; Koriat & Lieblich, 1975; MacKay, 1970; Rubin, 1975; Tweney et al., 1975).

If memory for words is positional, one can easily explain why words are not preferred over phantom-words when only TPs are given as cues, and why words seem to be extracted whenever prosodic cues are given. To see why this is the case, consider the experiments by Peña et al. (2002) and Endress and Bonatti (2007). In these experiments, participants had to learn that words had to start with certain syllables and to end with others (although this was not the way the experiments were described in Peña et al., 2002; for still another interpretation, see Perruchet, Tyler, Galland, & Peereman, 2004); in other words, they had to learn that certain syllables had to occur in certain positions, namely the first and the last. After familiarization with a speech stream, participants had to choose between items that adhered to this positional regularity but did not occur in the familiarization stream, and items that occurred in the familiarization stream (and thus had non-zero TPs) but did not adhere to the regularity. When familiarized with a continuous speech stream, participants did not choose the items respecting the positional regularity; rather, at least after long familiarizations, they chose the items that occurred in the stream. When words were separated by 25-ms silences as in Experiment 3, in contrast, participants chose the items respecting the positional regularity. They preferred these items for familiarization durations of up to 10 min; for longer familiarization durations, they chose the items that occurred in the stream. If one assumes that the items adhering to the positional regularity are extracted through positional memory mechanisms, while the items that occurred in the stream without respecting this regularity were tracked essentially through chaining mechanisms, one can easily explain this pattern of results. Positional memories may require other cues than TPs, such as the 25-ms silences; hence they are observed only when such cues are given. Chaining memories, in contrast, may be available irrespective of the presence of other cues; in fact, this is precisely what previous speech segmentation experiments have demonstrated (e.g., Aslin et al., 1998; Saffran et al., 1996). Moreover, chaining memories may take more time to build up; hence, participants should choose items with weak TPs only after moderately long exposures.

The same kinds of mechanisms may explain also the results presented here. As words are preferred to part-words in all of our experiments, chaining memories such as TPs seem to be computed irrespective of other cues. When only TPs are computed, however, participants should be unable to compute anything more than "bigram statistics", that is, co-occurrence statistics among syllable pairs; this is because chaining memories fundamentally are associative links among pairs of items (that is, in our case among syllables). If so, phantom-words should be as "familiar" as words since they are equated in terms of "bigram statistics", and more familiar than part-words, since the latter have weaker TPs. Indeed, this is exactly what we observed.

Prosodic-like cues, in contrast, may give positional information — because they define the edges of domains. That is, a learner may know that the element at the beginning of a prosodically defined domain is the first element of a sequence, and the element at the end of the domain the last one. Such edge-cues are particularly suitable for positional memories, because positions of items in a sequence seem to be encoded relative to the sequence edge according to most current models of positional memory (although the specific implementations differ widely in these models; see e.g., Henson, 1998; Hitch, Burgess, Towse, & Culpin, 1996; Ng & Maybery, 2002; Page & Norris, 1998). For example, Henson's (1998) model contains a start marker (activated at the beginning of a sequence) and an end marker (whose activity increases towards the end of the sequence); the relative activity of these marker elements allows to determine the position of an element within a sequence. Since positional memories are edge-based, and since prosodic cues provide information about the edges of some units, such cues may be well suited for word segmentation. If so, one would expect learners to extract words as soon as these are indicated by prosodic-like cues; this seems indeed to be the case in Experiments 3 and 4. Hence, learners may not require multiple cues to extract words from speech, but, under some circumstances, a single cue may be sufficient so long as it allows the construction

of positional memories. In more realistic situations, however, we would expect learners to use many different kinds of cues.

What cues are available for word-segmentation?

It is generally assumed in the word-segmentation literature that there are no systematic cues in fluent speech indicating word boundaries. Of course, there are language-specific cues to word boundaries such as stress, and it is well known that adults and older infants use such cues for word segmentation (e.g., Bortfeld et al., 2005; Cutler & Norris, 1988; Bortfeld et al., 2005; Dahan & Brent, 1999; Jusczyk et al., 1993; Jusczyk, Hohne, & Bauman, 1999; Mattys & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999; Suomi et al., 1997; Thiessen & Saffran, 2003; Vroomen et al., 1998). However, in order to use the fact that stress in, say, Hungarian is word-initial, infants have to segment words in the first place; in fact, without knowing where word boundaries fall, they cannot know that stress is word-initial either. In contrast, such problems do not arise with TPs, as such cues can be used without any knowledge of one's native language.

However, more recent evidence points to other segmentation strategies that do not require any language-specific knowledge. For example, neonates are sensitive to prosodic cues to word boundaries — irrespective of whether they are taken from French (their future native language) or from Spanish (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Christophe, Mehler, & Sebastian-Galles, 2001). These authors presented infants with two-syllable items that either came from the same word or straddled a word boundary; the phonemes in these words, however, were identical. For instance, the phoneme sequence /mati/ was taken either from *mathé***mati***cien* (where the phonemes appear in the same word) or from *pyja***ma ti***gré* (where the phonemes straddle a word boundary). As neonates successfully discriminate these items, there seem to be some language-independent cues to word-boundaries (or at least to phonological phrase boundaries, a prosodic constituent that can be larger than single words). However, while adults (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Davis, Marslen-Wilson, & Gaskell, 2002) and older infants (Gout, Christophe, & Morgan, 2004) can use these discrimination abilities for segmenting words from fluent speech, it remains an open question whether this is also the case at the earliest stages of language acquisition.

Also Shukla et al. (2007) provided evidence for prosodic cues to word-boundaries that do not rely on language-specific knowledge. They showed that listeners segment words predominantly from the edges of intonational contours — irrespectively of whether the contours were taken from Italian (the participants' native language) or from Japanese (a language unknown to the participants; see also (Seidl & Johnson, 2006), for similar evidence with infants). As mentioned before, Shukla et al. (2007) also showed that items straddling contour boundaries are not considered word-candidates even when their TPs are high. It thus seems that also intonational contours can be used for segmenting words without any knowledge of the properties of a language. Using intonational contours may be particularly well-suited as a segmentation strategy because infant-directed speech uses short utterances with more salient intonational contours, and important words tend to be placed in perceptually salient, sentence-final positions (Fernald & Mazzie, 1991; Fernald & Simon, 1984).

Also the prosodic cues used in the experiments reported here may be language-universal. Indeed, unit-final syllable lengthening seems to occur in all languages (e.g., Fon et al., 2002; Hoequist, 1983a; Hoequist, 1983b; Vassière, 1983). Moreover, the lengthening of the final part of a sequence is not specific to language. Rather, the same is true of other auditory stimuli, where an increase in pitch (or intensity, with which pitch may be confounded) is perceived as a cue to an onset of a group, while lengthening of an element is perceived as an end cue (e.g., Hay & Diehl, 2007; Woodrow, 1909). It is thus unlikely that infants need to "learn" that the lengthening of a constituent signals its end (though this conjecture needs to be empirically grounded). While the lengthening effect may be more pronounced for prosodic units larger than words, this may well correlate with the finding that infants predominantly segment clause-final words in the presence of prosodic cues (Seidl & Johnson, 2006). While more research is needed to explore how infants use prosodic cues, it is at least plausible that these cues may help infants to start segmenting the short utterances characteristic of infant-directed speech (Fernald & Mazzie, 1991; Fernald & Simon, 1984; Snow, 1977).

Possibly, infants might be able to exploit such prosodic cues to segment speech without any knowledge of their native language. Unlike TPs, these cues may allow the construction of positional memories (because they define the beginnings and the ends of domains), and thus seem particularly well suited for speech-segmentation. Under such an interpretation, words were preferred to phantom-words in Experiments 3 and 4 because such cues were given; in contrast, when TPs are the only available cue (as in Experiments 1 and 2), one would not expect such a preference.

Of course, an exclusive reliance on prosodic cues creates other problems (such as the grouping of clitic groups, as in 'a dog'), but they may allow at least the construction of the kind of memories that seems to underlie acoustic word-forms. We thus suggest that speech segmentation can get started due to prosodic cues, but it remains an important topic for further investigation to confirm this conjecture in infants, and to determine what other cues are used. In particular, we are entirely open to the possibility that TPs may be used to consolidate word memories; in fact, given the overwhelming evidence that TPs are computed in many different situations (e.g., Aslin et al., 1998; Fiser & Aslin, 2002; Saffran et al., 1996), and that they can be computed simultaneously with positional memories (e.g., Endress & Bonatti, 2007; Endress & Mehler, in press), it would be strange if they were not used to link together syllables within words. Our results suggest, however, that this is possible only once positional memories are available.

There are also other, non-prosodic, segmentation strategies that may allow for the construction of positional memories, and that do not require language-specific knowledge. For instance, infants may preferentially learn words they hear in isolation (Brent & Siskind, 2001; Van de Weijer, 1999; but see Aslin, Woodward, LaMendola, & Bever, 1996).[2] They also seem to use known words to determine the edges of other words (e.g., Brent, 1997; Bortfeld et al., 2005; Dahan & Brent, 1999). For example, if infants recognize the word "dog" in fluent speech, they may know that a new word starts after "dog".

In sum, there may be a number of prosodic and non-prosodic cues in addition to TPs that may be used for speech segmentation without requiring language-specific knowledge. In contrast to TPs, these cues may allow the construction of positional memories, and may thus be well suited for extracting auditory word candidates. Once such cues are available, TPs may also have a role in word-segmentation, e.g., for strengthening the associations among syllables. In the absence of such cues, however, learners do not seem to register at all whether they have encountered an item or not, and are sometimes more familiar with spurious phantom-words than with well-attested syllable sequences. Apparently, participants failed to extract any word-candidates due to TP computations — because such computations may appeal to different kinds of memory mechanisms than memories for acoustic word forms. For understanding the first steps in word-learning, it will thus be important to explore both the kinds of representations learners form of acoustic word candidates, and the fundamental contributions of other, in particular prosodic, sources of information for the construction of such representations.

Acknowledgments

This research has been supported by McDonnell Foundation Grant 21002089 and the European Commission Special Targeted Project CALACEI (contract N° 12778 NEST). We are grateful to R. Aslin, L. Bonatti, D. Cahill, Á. Kovács, M. Nespor, M. Shukla, and J. Toro for helpful comments and discussions.

[2] These results also reconcile our claim that there is no evidence for the use of TPs in word-segmentation with Graf-Estes, Evans, Alibali, and Saffran's (2007) work. In these experiments, infants were first familiarized with a word-segmentation speech stream. Then they had to associate either words, non-words or part-words from the stream with visual pictures. Infants learned the picture-sound associations better for words than for either non-words or part-words. Graf-Estes et al. (2007) interpreted these results as evidence for a role of TPs in determining word candidates. However, there is a simpler alternative interpretation: In fact, during the word-learning phase, the auditory items were presented in isolation; this is likely to be sufficient to construct positional memories for these words. Once these memories are constructed, one may of course expect processing advantages for syllable sequences that are more familiar due to the previous segmentation phase; in this sense, TPs appear to have contributed to their results. We suggest, however, that the reason why the auditory items were extracted in the first place is that they were presented in isolation, and not because they were presented during the speech stream. We thus predict that phantom-words should be associated to pictures as easily as words, and more easily than part-words. On Graf-Estes et al.'s (2007) account, in contrast, one would predict that only words but not phantom-words are preferentially associated to pictures. Moreover, it should also be noted that Graf-Estes et al.'s (2007) stimuli were produced by a speaker rather than being synthesized. It is thus unclear whether possible prosodic cues introduced thereby contributed to their findings.

Appendix A. Test items used in Experiments 1a, 1d, 3 and 4

Table A1. Test pairs used in Experiments 1a, 1d, 3 and 4

Word/phantom-word trials		Word/part-word trials			
Word	Phantom-word	Word	Part-word (BCA)	Word	Part-word (CAB)
tazepi	tazeRu	tazepi	zepimi	tazepi	pitano
mizeRu	tazeRu	mizeRu	zeRufe	mizeRu	Rufeno
tanoRu	tazeRu	tanoRu	noRumi	tanoRu	Rufeku
fekupi	fekula	fekupi	kupita	fekupi	pimiku
mikula	fekula	mikula	kulafe	mikula	lataze
fenola	fekula	fenola	nolata	fenola	lamize

Appendix B. Test items used in some pilot experiments

Table B1. Test pairs used in some pilot experiments

Word/phantom-word trials		Word/part-word trials			
Word	Phantom-word	Word	Part-word (BCA)	Word	Part-word (CAB)
kesuti	kesuna	kesuti	sutiSo	kesuti	tikemo
Sosuna	kesuna	Sosuna	sunale	Sosuna	nalemo
kemona	kesuna	kemona	monaSo	kemona	nalepu
leputi	lepufa	leputi	putike	leputi	tiSopu
Sopufa	lepufa	Sopufa	pufale	Sopufa	fakesu
lemofa	lepufa	lemofa	mofake	lemofa	faSosu

Appendix C. Test Items Used in Experiment 1b

Table C1. Test pairs used in Experiment 1b

Word/phantom-word trials		Word/part-word trials			
Word	Phantom-word	Word	Part-word (BCA)	Word	Part-word (CAB)
bagadu	bagaso	bagadu	gaduto	bagadu	dubapi
togaso	bagaso	togaso	gasoli	togaso	solipi
bapiso	bagaso	bapiso	pisoto	bapiso	solimu
limudu	limufe	limudu	muduba	limudu	dutomu
tomufe	limufe	tomufe	mufeli	tomufe	febaga
lipife	limufe	lipife	pifeba	lipife	fetoga

Appendix D. Test Items Used in Experiment 1c

Table D1. Test pairs used in Experiment 1c

Word/phantom-word trials		Word/part-word trials			
Word	Phantom-word	Word	Part-word (BCA)	Word	Part-word (CAB)
Rosenu	pasenu	Rosenu	senule	Rosenu	saRose
Rokafa	lekafa	Rokafa	kafapa	Rokafa	giRoka
paseti	pasenu	paseti	setimu	paseti	gipase
pamonu	pasenu	pamonu	monubo	pamonu	vopamo
muRisa	muRivo	muRisa	Risapa	muRisa	famuRi
munavo	muRivo	munavo	navoRo	munavo	gimuna
luRivo	muRivo	luRivo	Rivole	luRivo	tiluRi
ludegi	bodegi	ludegi	degipa	ludegi	tilude
lemofa	lekafa	lemofa	mofalu	lemofa	salemo
lekati	lekafa	lekati	katilu	lekati	nuleka
bonagi	bodegi	bonagi	nagiRo	bonagi	nubona
bodesa	bodegi	bodesa	desale	bodesa	fabode

Appendix E. Test items used in Experiment 2

Table E1. Test pairs used in Experiment 2

Phantom-word/part-word trials			Word/part-word trials			
Phantom-word	Part-word	Part-word type	Word	Part-word (BCA)	Word	Part-word (CAB)
tazeRu	zeRufe	BCA	tazepi	zepimi	mizeRu	lamize
tazeRu	Rufeku	CAB	tazepi	kupita	tanoRu	Rufeno
fekula	zeRufe	BCA	mizeRu	zepimi	fekupi	pimiku
fekula	Rufeku	CAB	tanoRu	noRumi	mikula	pimiku
			fekupi	kupita	mikula	lamize
			fenola	noRumi	fenola	Rufeno

References

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

Aslin, R. N., Woodward, J., LaMendola, N., & Bever, T. (1996). Models of word segmentation in fluent maternal speech to infants. In K. Demuth & J. L. Morgan (Eds.), *Signal to Syntax. Bootstrapping from Speech to Grammar in early Acquisition* (pp. 117–134). Mahwah, NJ: Erlbaum.

Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition, 83*, 167–206.

Bates, E., & Elman, J. L. (1996). Learning rediscovered. *Science, 274*, 1849–1850.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*.

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.

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

Brent, M., & Siskind, J. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*, B33–44.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*, 204–223.

Brown, R., & McNeill, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325–337.

Cairns Shillcock Levy & Chater (1997). Bootstrapping word boundaries: A bottom–up corpus-based approach to speech segmentation. *Cognitive Psychology, 33*, 111–153.

Cattell, J. M. C. M. (1886). The time it takes to see and name objects. *Mind, 41*, 63–65.

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.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America, 95*, 1570–1580.

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy, 2*, 385–394.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access. I: Adult data. *Journal of Memory and Language, 51*, 523–547.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235–253.

Conrad, R. (1960). Serial order intrusions in immediate memory. *British Journal of Psychology, 51*, 45–48.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113–121.

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.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 218–244.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning Memory & Cognition, 17*, 875–887.

Dutoit, T., Pagel, V., Pierret, N., & Bataille, F., van der Vreken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *Proceedings of the fourth international conference on spoken language processing*, Philadelphia (Vol. 3, pp. 1393–1396).

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition, 105*, 247–299.

Endress, A. D., & Hauser, M. D. (in preparation). Cross-linguistic word segmentation without distributional cues.

Endress, A. D., & Mehler, J. (in press). Primitive computations in speech processing. *Quarterly Journal of Experimental Psychology*.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*, 104–113.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 15822–15826.

Fon, J. (2002). *A cross-linguistic study on syntactic and discourse boundary cues in spontaneous speech*. Unpublished Doctoral Dissertation, Ohio State University, Columbus, OH.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.

Goodsitt, J., Morgan, J. L., & Kuhl, P. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language, 20*, 229–252.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access. II: Infant data. *Journal of Memory and Language, 51*, 548–567.

Graf-Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*, 254–260.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78*, B53–64.

Hay, J. S. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics, 69*, 113–122.

Hayes, J. R., & Clark, H. H. (1970). Experiments in the segmentation of an artificial speech analog. In J. R. Hayes (Ed.), *Cognition and the Development of Language*. New York: Wiley.

Henson, R. (1998). Short-term memory for serial order: The start–end model. *Cognitive Psychology, 36*, 73–137.

Henson, R. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition, 27*, 915–927.

Hicks, R., Hakes, D., & Young, R. (1966). Generalization of serial position in rote serial learning. *Journal of Experimental Psychology, 71*, 916–917.

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*, 116–139.

Hoequist, C. (1983a). Durational correlates of linguistic rhythm categories. *Phonetica, 40*, 19–43.

Hoequist, C. (1983b). Syllable duration in stress-, syllable- and mora-timed language. *Phonetica, 40*, 203–237.

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.

Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 524–531.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*, 675–687.

Jusczyk, P. W., Hohne, E., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61*, 1465–1476.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*, 159–207.

Kinder, A. (2000). The knowledge acquired during artificial grammar learning: Testing the predictions of two connectionist models. *Psychological Research, 63*, 95–105.

Kinder, A., & Assmann, A. (2000). Learning artificial grammars: No evidence for the acquisition of rules. *Memory & Cognition, 28*, 1321–1332.

Kohn, S. E., Wingfield, A., Menn, L., Goodglass, H., Gleason, J. B., & Hyde, M. (1987). Lexical retrieval: The tip-of-the-tongue phenomenon. *Applied Psycholinguistics, 8*, 245–266.

Koriat, A., & Lieblich, I. (1974). What does a person in a 'TOT' state know that a person in a 'don't know' state doesn't know? *Memory & Cognition, 2*, 647–655.

Koriat, A., & Lieblich, I. (1975). Examination of the letter serial position effect in the TOT and the don't know states. *Bulletin of the Psychonomic Society, 6*, 539–541.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*, 323–350.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91–121.

Mattys, S. L., Jusczyk, P. W., Luce, P., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465–494.

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology, 56*, 485–491.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance. I: Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127–162.

Ng, H. L., & Maybery, M. T. (2002). Grouping in short-term verbal memory: Is position coded temporally? *Quarterly Journal of Experimental Psychology: Section A, 55*, 391–424.

Page, M. P., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761–781.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604–607.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119*, 264–275.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General, 133*, 573–583.

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

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*, 115–119.

Rubin, D. C. (1975). Within word structure in the tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior, 14*, 392–397.

Saffran, J. R. (2001a). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493–515.

Saffran, J. R. (2001b). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*, 149–169.

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., 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.

Schulz, R. W. (1955). Generalization of serial position in rote serial learning. *Journal of Experimental Psychology, 49*, 267–272.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science, 9*, 565–573.

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology A, 50*, 216–252.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology, 54*, 1–32.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language, 4*, 1–22.

Solomon, R. L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology, 43*, 195–201.

Suomi, K., McQueen, J., & Cutler, A. (1997). Vowel harmony and speech segmentation in finnish. *Journal of Memory and Language, 36*, 422–444.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*, 86–132.

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.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*, 1–42.

Toro, J. M., Bonatti, L., Nespor, M., & Mehler, J. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science, 19*, 137–144.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67*, 867–875.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134*, 552–564.

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 195–202.

Tweney, S., Ryan, D., Tkacz & Zaruba, S. (1975). Slips of the tongue and lexical storage. *Language and Speech, 18*, 388–396.

Vassière, J. (1983). Language-independent prosodic features. In A. Cutler & R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53–65). Hamburg, Germany: Springer.

Van de Weijer, J. (1999). Language input for word discovery. *Mpi series in psycholinguistics Max Plank Institute for Psycholinguistics, Nijmege, 9*.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*, 133–149.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Woodrow, H. S. (1909). A quantitative study of rhythm: The effect of variations in intensity, rate, and duration. *Archiv fur Psychologie, 14*, 1–66.

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences, 8*, 451–456.