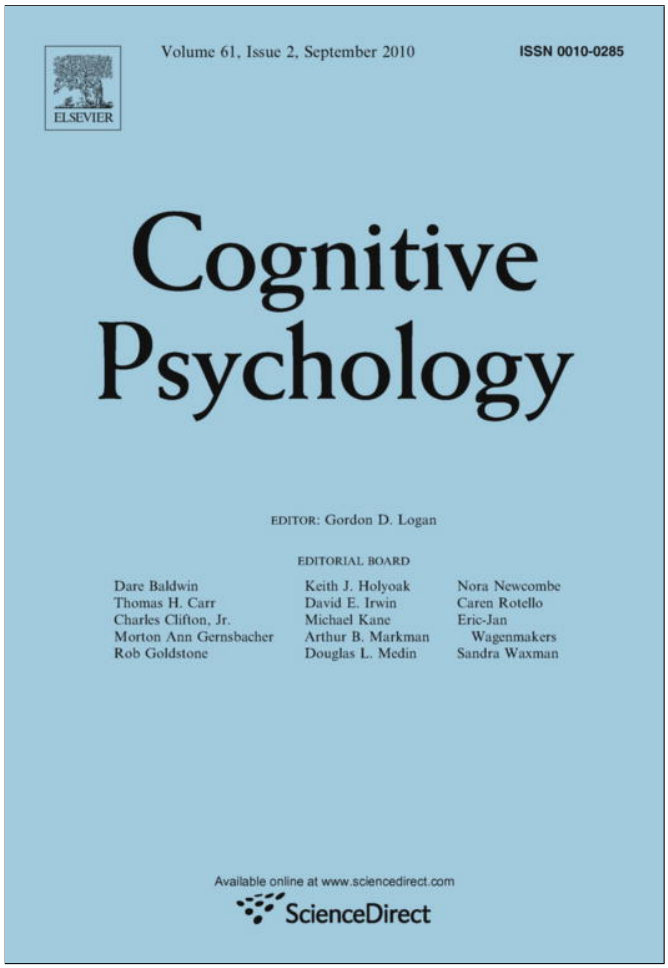

Cognitive Psychology 61 (2010) 177–199

Contents lists available at ScienceDirect

Cognitive Psychology

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

Word segmentation with universal prosodic cues

Ansgar D. Endress *, Marc D. Hauser

Department of Psychology, Harvard University, Cambridge, MA, USA

ARTICLE INFO

Article history:
Accepted 18 May 2010
Available online 22 June 2010

Keywords:
Word segmentation
Prosody
Language universals
Language acquisition
Statistical learning

ABSTRACT

When listening to speech from one's native language, words seem to be well separated from one another, like beads on a string. When listening to a foreign language, in contrast, words seem almost impossible to extract, as if there was only one bead on the same string. This contrast reveals that there are language-specific cues to segmentation. The puzzle, however, is that infants must be endowed with a language-independent mechanism for segmentation, as they ultimately solve the segmentation problem for any native language. Here, we approach the acquisition problem by asking whether there are language-independent cues to segmentation that might be available to even adult learners who have already acquired a native language. We show that adult learners recognize words in connected speech when only prosodic cues to word-boundaries are given from languages unfamiliar to the participants. In both artificial and natural speech, adult English speakers, with no prior exposure to the test languages, readily recognized words in natural languages with critically different prosodic patterns, including French, Turkish and Hungarian. We suggest that, even though languages differ in their sound structures, they carry universal prosodic characteristics. Further, these language-invariant prosodic cues provide a universally accessible mechanism for finding words in connected speech. These cues may enable infants to start acquiring words in any language even before they are fine-tuned to the sound structure of their native language.

1. Introduction

"Speech segmentation" refers to two distinct literatures. Traditionally, speech segmentation meant the processes used by native speakers of a language to access and identify words in speech from that

* Corresponding author. Address: Massachusetts Institute of Technology, 43 Vassar St., 46-4127, Cambridge, MA 02139, United States.
E-mail address: ansgar.endress@m4x.org (A.D. Endress).

0010-0285/$ - see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.cogpsych.2010.05.001

language; this literature has revealed a wealth of cues, often prosodic, that native speakers use to find word-boundaries and, at least in adulthood, to access words in fluent speech (e.g., Cutler, Mehler, Norris, & Segui, 1986; Cutler & Norris, 1988; Cutler & Mehler, 1993; Houston, Santelmann, & Jusczyk, 2004; Johnson & Jusczyk, 2001; Johnson, Jusczyk, Cutler, & Norris, 2003; Jusczyk, Cutler, & Redanz, 1993, 1999; Mattys & Samuel, 1997; Mattys, Jusczyk, Luce, & Morgan, 1999; McQueen, Otake, & Cutler, 2001; Nazzi, Iakimova, Bertoncini, FrTdonie, & Alcantara, 2006; Norris, McQueen, Cutler, & Butterfield, 1997). We will call this literature "native speech segmentation." In the last decade or so, however, "speech segmentation" has come to refer to a second set of processes, namely those used by infant learners (who do not have a native language yet) to discover (new) words from fluent speech (e.g., Aslin, Saffran, & Newport, 1998; Brent & Cartwright, 1996; Saffran, Newport, & Aslin, 1996; Saffran, Aslin, & Newport, 1996; Swingley, 2005). We call this literature "statistical word segmentation." In this literature, it has generally been assumed that fluent speech provides no systematic cues to word-boundaries. Furthermore, it is assumed that those cues discovered in the native speech segmentation literature differ from language to language, which seems to imply that they cannot be used by infant learners – who do not yet know which language they will end up speaking. Discovering words in fluent speech in the absence of systematic cues to word-boundaries is an obvious problem, and hence this problem has been aptly termed the "segmentation problem."

A potential solution to this problem, one that has received considerable empirical and theoretical attention in the last 10 years, is that infants may use statistical mechanisms to track distributional information of phonemes and syllables. The logic underlying these studies is that distributional information has the potential to provide cues to word-boundaries independently of the language that the developing child confronts (e.g., Aslin et al., 1998; Batchelder, 2002; Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997; Perruchet & Vinter, 1998; Saffran, Newport, et al., 1996; Saffran, Aslin, et al., 1996, 2005). However, though there is little doubt that human infants, and even non-human animals, can track distributional information (e.g., Hauser, Newport, & Aslin, 2001; Saffran, Aslin, et al., 1996; Swingley, 2005), we believe that there is no clear evidence that such abilities are used to segment words from fluent speech during the process of language acquisition. Said differently, though the current literature unambiguously shows that infants can extract distributional information, there are only two studies that provide any evidence that infants in fact use this information for the purposes of language acquisition (Graf-Estes, Evans, Alibali, & Saffran, 2007; Mirman, Magnuson, Estes, & Dixon, 2008). However, even for these studies, the exact contribution of transitional probabilities is unclear (see below for discussion). This distinction between extraction and use is, we believe, important, for at least three reasons. First, given that the capacity to compute statistics is unlikely to be specific to the domain of language, evidence that particular statistics can be tracked must be accompanied by additional experiments showing that such abilities are recruited in the process of acquiring a language. Second, there is evidence that such mechanisms may not be suitable for all languages – a claim that strikes at the problem of uncovering a universal segmentation mechanism (e.g., Yang, 2004). Third, some recent studies suggest that learners cannot segment words from fluent speech using distributional information even in situations where they demonstrably track such information (Endress & Mehler, 2009b).

Irrespective of whether distributional information can be used to segment words from fluent speech, learners would likely benefit from other cues to word-boundaries that do not rely on such information. That is, if distributional information cannot be used for learning words from fluent speech, learners need to have access to *some* cues that allow them to learn the words of their native language. On the other hand, even if future studies can confirm that distributional mechanisms are used for learning words from fluent speech, distributional information is by its very *probabilistic* nature not entirely reliable (e.g., Yang, 2004), suggesting that infants might need to combine such information with other cues to successfully learn words from fluent speech (Christiansen, Allen, & Seidenberg, 1998; Saffran, Newport, et al., 1996).

Here, we add to the growing literature on the segmentation problem by asking, most generally, whether there are other potential mechanisms that might allow learners to segment words from fluent speech, with the initial desiderata that it be a mechanism that can segment speech from any potential linguistic input, and as such, is designed as a universal acquisition mechanism. More specifically, we ask whether adult learners can segment a foreign language they have never been ex-

posed to in the absence of distributional or statistical cues. Though it is natural to use infants to explore developmental questions, here we use adults because it enables us to provide the strongest possible test of the universality question: if adults, who have already acquired a language, can segment words from an entirely unfamiliar non-native language, with the potential constraint that their own native language may actually aversely affect processing of non-native material, then adults must have access to some mechanism that enables this ability, and whatever this mechanism is, it must have the kinds of properties that would enable infants to learn whichever language they are exposed to in early development. Evidence that adults have such a mechanism does not, of course, imply that it is operative in infant development, nor that it is recruited. But if this mechanism is in place, it is at least a strong candidate, thus setting the stage for additional studies in infants.

We start by asking whether monolingual native speakers of American English can segment fluent speech when exposed to simplified artificial speech with French prosody or to natural sentences from French, Turkish and Hungarian. If participants can segment these stimuli, they must use language-universal, potentially prosodic, features.

1.1. The segmentation problem: a brief review

Fluent speech does not contain any pauses between words analogous to spaces in written texts, and it is commonly assumed that there are no systematic cues to word-boundaries (e.g., Aslin et al., 1998; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996). That is, adults and older infants can use different prosodic cues to word-boundaries (for word learning in infants, see e.g., Houston et al., 2004; Johnson & Jusczyk, 2001; Johnson et al., 2003; Jusczyk et al., 1993, Jusczyk, Houston, & Newsome, 1999; Mattys et al., 1999; Nazzi et al., 2006; for lexical access in adults, see e.g., Cutler et al., 1986; Cutler & Norris, 1988; Cutler & Mehler, 1993; Mattys & Samuel, 1997; McQueen et al., 2001; Norris et al., 1997). However, as many of these speech cues differ across languages, it is assumed that they are language-specific, and cannot be used by infant learners before they have learned at least some basic properties of their future native language (even though infants can acquire certain properties of their native language very quickly, see e.g., Maye, Werker, & Gerken, 2002).

In response to this problem, several authors have proposed word-segmentation mechanisms that rely on co-occurrence statistics. These mechanisms take a sequence of, say, syllables as input, and compute how often syllables occur together. Syllables that frequently co-occur are likely to be part of the same word, while syllables that occur rarely together are likely to span a word-boundary. While this general idea has been implemented in various forms, including minimum description length compression techniques (Brent & Cartwright, 1996), mutual information of syllables (Swingley, 2005), and various models that track syllable "chunks" in speech (e.g., Batchelder, 2002; Perruchet & Vinter, 1998), the most widely used measure of co-occurrence statistics are transitional probabilities (TPs). TPs are conditional probabilities of syllables following each other; dips in TPs may indicate word-boundaries, while peaks in TPs may indicate that a word continues. Rats, cotton-top tamarins and human infants can all track co-occurrence statistics on a variety of stimuli, including speech (e.g., Aslin et al., 1998; Fiser & Aslin, 2001; Hauser et al., 2001; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996; Saffran, Johnson, Aslin, & Newport, 1999; Toro & Trobalón, 2005). As a result, these mechanisms are readily available to the child learner during language acquisition, and may thus plausibly be deployed in the process of speech segmentation (e.g., Aslin et al., 1998; Saffran, Aslin, et al., 1996; Swingley, 2005). A prime advantage of such approaches is that co-occurrence statistics-based mechanisms are language-independent and, therefore, can be evaluated without any assumptions about the language a learner will end up speaking. As a result, if speech cues to word-boundaries are language-specific, such approaches appear to be the only viable avenue to speech segmentation. Further, simulation results have shown that, once a learner has acquired knowledge about language-specific speech cues (e.g., stress or phonotactic regularities), these cues can be used to complement purely distributional information (e.g., Brent & Cartwright, 1996; Christiansen et al., 1998). Co-occurrence statistics might thus provide a language-general mechanism to bootstrap word-segmentation that can be integrated with other cues once the learner knows to interpret them.

Despite the plausibility of co-occurrence statistics-based segmentation mechanisms, several authors have raised doubts about whether such mechanisms can be used by infant learners to seg-

ment words from fluent speech, and even if they can, there is little direct evidence that such mechanisms are used in the process of language acquisition. Even in the arguably strongest case so far, the role of statistical computations in word-segmentation is unclear. Specifically, Graf-Estes et al. (2007) (see also Mirman et al., 2008) presented infants with a continuous speech stream in which some syllable combinations had stronger TPs, and others weaker TPs. Following this, they presented these syllable combinations in isolation, and paired them with visual images. Results showed that infants were better at associating the images with items with stronger TPs compared to items with weaker TPs. Graf-Estes et al. (2007) proposed that these results show that TP-based processes play an important role in word learning.

However, there is an alternative interpretation, relating to the prosodic cues that were implicitly present in these experiments. Specifically, Graf-Estes et al. (2007) did provide infants with important prosodic information during the sound-picture association phase; after all, the items were presented in *isolation* during this phase, thus constituting a full utterance. This prosodic information might be crucial for infants to associate speech items with pictures. Infants might not learn any word-like chunks from the exposure to continuous speech at all (see Endress & Mehler, 2009b, and below). Rather, they might learn word-like chunks *exclusively* when the speech items are presented in isolation and, therefore, in the phase where they were paired with the pictures. On this view, the advantage of high-TP items over low-TP items would be due to a secondary process not directly involved in word learning: since the syllables in high-TP items have stronger associations than the syllables in low-TP items, it might simply be easier for infants to process the former items, leading to the high-TP advantage in the speech-picture association phase. Importantly, however, infants memorize the items as word-like chunks only because they were presented in isolation, and thus indicated by clear prosodic cues. We believe, therefore, that the available evidence does not warrant the conclusion that TPs play a necessary role in language acquisition.

In line with such concerns, some authors suggest that distributional mechanisms based on TP computations may not be successful in all languages infants may have to learn, throwing doubt on the universality issue (Yang, 2004). Other studies raise doubts whether learners actually use co-occurrence statistics for segmenting speech even when they can track them. For example, Endress and Mehler (2009b) familiarized participants with continuous speech streams where TPs were the only cues to word-boundaries. The streams were constructed such that (statistically defined) "words" had identical TPs to "phantom-words" that, in contrast to words, never occurred in the stream. That is, the syllables in the words that occurred in the speech stream were arranged so that TPs between them were identical to TPs between syllables in phantom-words, although the specific syllable combinations corresponding to phantom-words never occurred in the speech stream. It turned out that participants were unable to decide whether they had heard words or phantom-words even after hearing each word 600 times. Moreover, they believed that it was more likely that they had heard phantom-words (which they had not heard) than items that did occur in the speech stream but had weaker TPs. These results confirm that participants were indeed sensitive to TPs. However, they raise the possibility that participants may not use this ability to construct units that could be stored in memory; if they did, they should remember the words that occurred in the input, and not spurious syllable combinations that they have never heard at all.

Independently of whether the above mentioned concerns about TP-based mechanisms are valid, we focus here on the presumably non-controversial point that speech may carry additional cues to help the naïve learner solve the segmentation problem (e.g., Christiansen et al., 1998; Saffran, Newport, et al., 1996). In other words, given that TPs are, by definition, probabilistic, and thus may not guarantee reliable segmentation, it would not be surprising to find that learners take advantage of additional cues, recruiting mechanisms that increase reliability either in combination with TP-based mechanisms, or provide a sufficient basis to solve the segmentation problem on their own.

Prosody is a prime candidate cue to segmentation, for at least four reasons. First, as mentioned above, adult speakers use prosodic cues to access words in fluent speech once they have learned their native language; if language acquisition can exploit similar cues as language processing, prosodic cues might be used for finding words in fluent speech. Second, in the experiments where participants failed to use TPs to find words in fluent speech, prosodic cues reestablished a sensitivity to the words participants had heard, allowing them to discriminate words and phantom-

words (Endress & Mehler, 2009b). Third, in experiments in which prosodic information was pitted against statistical information, 8-month-old and 11-month-old infants seem to rely more on prosodic information than on statistical information (e.g., Johnson & Jusczyk, 2001; Johnson & Seidl, 2009), even though infants might use stress-based, prosodic information after they use statistical information (Thiessen & Saffran, 2003). Fourth, several authors have suggested that prosodic information is used to bootstrap many different aspects of language acquisition, including words and syntactic information, importantly even before infants know any words of their native language (e.g., Christophe, Nespor, Guasti, & Van Ooyen, 2003; Morgan & Demuth, 1996; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). For these reasons, it would seem plausible to think that prosody might also be used to bootstrap word-segmentation. However, the possibility that learners might use prosodic cues to find words in fluent speech also raises an important problem. If, as is generally assumed, prosodic cues are language-specific, they cannot be used by infant learners to kick-start word-segmentation as they do not yet know the relevant properties of their native language. In the next section, however, we discuss the possibility that there are language-independent cues to word-boundaries that do not rely on co-occurrence statistics.

1.2. The segmentation problem across languages

While prosody provides powerful cues to word-boundaries, those cues are believed to be largely language-specific, and speakers of different languages do indeed use different segmentation strategies, both when segmenting their native language and when segmenting a foreign language (e.g., Cutler et al., 1986; Cutler & Mehler, 1993; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Otake, Hatano, Cutler, & Mehler, 1993). However, prosodic cues may be less language-specific than commonly assumed. Consider, as an example, stress. While stress is (generally) word-initial in English, it is word-final in French (at least on words that bear stress due to sentence-level prosody). At first sight, one would thus expect English speakers to mis-segment French, as they should consider the last (stressed) syllables as initial syllables, since the initial position is where stress normally resides in English. However, this reasoning is not necessarily valid, as "stress" is not simply a binary property of syllables; that is, stressed syllables are not simply louder than unstressed syllables. In particular, stress is implemented using three different dimensions, namely the loudness, pitch, and duration of a syllable (e.g., Ashby & Maidment, 2005; Hayes, 1995). Stress is implemented using these dimensions in potentially language-universal ways. For example, languages that use word-initial stress (such as English) tend to rely more on the pitch and the loudness of a syllable to signal stressed syllables, while languages that use word-final stress (such as French) tend to use duration and loudness (e.g., Hayes, 1995). That is, the location of the stress differs between English and French, but so do the cues used to convey stress. The choice of cues, in contrast, may be universal for a given stress position. Indeed, even for auditory non-speech sequences, an increase in pitch is perceived as a group onset, while an increase in duration is perceived as a group-offset (see Hay & Diehl, 2007; Woodrow, 1909; but see Iversen, Patel, & Ohgushi, 2008, for evidence that these biases might not be universal). If these perceptual grouping mechanisms are innate, they would allow infants learners to use stress as a cue to word-boundaries without any knowledge of their future native language.[1]

Considering other prosodic cues to word-boundaries than stress also seems to suggest that such cues might be less language-specific than commonly assumed. Specifically, the prosodic structure of an utterance is hierarchically organized (e.g., Hayes, 1989; Nespor & Vogel, 1986). The units at

[1] While knowing that an increase in pitch or duration signals a word onset or offset, respectively, may be instrumental for learning words from fluent speech in many situations, a complete account will necessarily be more complex. Languages with both long and short vowels are a case in point. In German, for instance, the first vowel in words such as 'Biene' (bee) is longer than the second vowel (because the first vowel is a long vowel). Hence, although the stressed syllable is word-initial, it is longer than the word-final syllable; simply interpreting the longer syllable as word-final would, in this case, lead to mis-segmentation. Still, native speakers of English with no knowledge of German can use this kind of prosodic information to identify word-boundaries (unpublished data). Presumably, listeners interpret all three dimensions of stress – amplitude, duration and pitch – in an integrated fashion, a conclusion that also seems to follow from research on auditory perceptual grouping mechanisms (e.g., Fraisse, 1982).

the top of this hierarchy are whole utterances; these are then hierarchically decomposed into smaller constituents, with syllables (or, in some languages, moras) at the lowest levels.

Except for constituents at the lowest levels (i.e., moras, syllables and feet), boundaries of prosodic constituents are word-boundaries, because the edges of these constituents necessarily coincide with edges of words. To the extent that prosodic boundaries are signaled by phonetic cues, these cues are, therefore, also potential cues to word-boundaries. Moreover, at least at the highest levels, boundary cues are unlikely to require language-specific knowledge to be perceived. For example, utterances are simply bounded by silence, and thus by general perceptual end markers. At the next level, Intonational Phrases follow a complete pitch contour; Intonational Phrase boundaries are therefore marked by the transition between two pitch contours and, therefore, an abrupt change in pitch (e.g., Vaissière, 1983, 2005), a situation that might plausibly correspond to a general perceptual breakpoint. Infants seem to be sensitive to these perceptual boundaries (e.g., Hirsh-Pasek et al., 1987; Nazzi, Nelson, Jusczyk, & Jusczyk, 2000), preferentially learning words they hear in isolation (and thus, constitute their own utterance; Brent & Siskind, 2001; Van de Weijer, 1999; but see Aslin, Woodward, LaMendola, & Bever, 1996), as well as words that occur at utterance edges compared to utterance middles (Seidl & Johnson, 2006).

Utterances and Intonational Phrases might thus provide some cues to words boundaries that do not depend on language-specific knowledge. However, these prosodic constituents are typically larger than words, and therefore provide boundary information only for a subset of the words they contain. While this problem is reduced in infant-directed speech because of the short utterances characteristic of infant-directed speech (Fernald & Mazzie, 1991; Fernald & Simon, 1984; Snow, 1977), cues to the boundaries of smaller constituents might be more conducive to learning words from fluent speech, because smaller constituents contain fewer potential words.

There is some evidence that 6-month-old infants, exposed to a native English-speaking environment, are sensitive to boundary cues of some smaller prosodic units (Soderstrom et al., 2003). Moreover, previous research suggests that some cues signaling boundaries are similar in a variety of languages, although they tend to be stronger at higher-level prosodic units (e.g., Intonational Phrases) than at lower-level units. For example, syllables at the end of a prosodic unit tend to be lengthened in many if not all languages (e.g., Fon, 2002; Hoequist, 1983b; Vaissière, 1983; see also Christophe, Mehler, & Sebastian-Galles, 2001; Shattuck-Hufnagel & Turk, 1996 and references therein), a generalization that seems to hold even for sign languages (Liddell, 1978; Wilbur, 2009, as cited by Brentari, González, Seidl, & Wilbur, in press). Likewise, phonemes tend to be more strongly articulated at the beginning of prosodic units than in other positions within the units (Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2004). Further, as mentioned above, transitions between prosodic units such as Intonational Phrases are signaled by an abrupt change in pitch (e.g., Vaissière, 1983, 2005).

In line with these data, neonates are sensitive to prosodic cues to word boundaries independently of whether they are taken from French (their future native language) or from Spanish (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Christophe et al., 2001; see also Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Davis, Marslen-Wilson, & Gaskell, 2002; Gout, Christophe, & Morgan, 2004, for related work with adults). In these experiments, infants were presented with two-syllable items. These items either came from the same word (as the phoneme sequence /mati/ in *mathé***mati***cien*), or straddled a word boundary (as in *pyja***ma ti***gré*). Neonates successfully discriminated such items, independently of whether the words were taken from their future native language or from another language. Moreover, naïve English speakers judge *Korean* sentences as more natural when silences are inserted between syntactic constituents (that is, at the boundary of prosodic constituents) than when silences are inserted in other positions (e.g., between words of the same constituents, or within words; Pilon, 1981; but see Wakefield, Doughtie, & Yom, 1974, for evidence that these discriminations might require at least some minimal exposure to Korean). Further, in sign languages, even non-signers can detect sentence boundaries (Brentari et al., in press; Fenlon, Denmark, Campbell, & Woll, 2008), and they can make accurate inferences about whether two-sign combinations constitute one or two words (Brentari et al., in press). These results also suggest that some prosodic break-points can be perceived in the absence of language-specific knowledge.

The results reviewed so far suggest that some prosodic cues to word-boundaries are independent of language-specific knowledge (see the Section 7 for potentially universal non-prosodic cues to word

boundaries). Moreover, at least in adults, these cues are used for accessing known words. For example, words such as "hamster" have another word as their first syllable (i.e., "ham"). Adults speakers use the duration of the first syllable to decide whether they have heard "hamster" or "ham"; "ham" is shorter when it is part of "hamster" than when it is a word on its own (e.g., Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006b). Adults can use this cue even in newly learned words for which they have not experienced any duration variation at all (Shatzman & McQueen, 2006a). However, while these results demonstrate that infants and adults have a perceptual sensitivity to some potentially universal prosodic cues to word-boundaries, and that, in their native language, at least adults can use these cues to access words they have already learned, it is unclear whether learners would use such cues to discover *new* words in fluent speech. It is this question that we start addressing in the experiments presented below.

1.3. The current experiments

In the following experiments, we will ask whether learners can use prosodic cues to word-boundaries in the absence of language-specific knowledge. We start addressing this issue by asking whether adult, largely monolingual native speakers of American English can segment speech from completely unfamiliar languages, even when all other cues to word-boundaries (e.g., co-occurrence statistics) are eliminated from the input.

In all experiments, participants were native speakers of American English. In Experiments 1 and 2, participants were familiarized with artificial speech streams in which all syllable transitions had the same TPs, and in which syllable frequencies were controlled. In Experiment 1, we placed target words at the edge of an intonational contour recorded from French; in Experiment 2 we placed the target word in the middle of such a contour. While participants in Experiments 1 and 2 were exposed to artificial speech, we asked in Experiments 3–5 whether native speakers of American English can segment naturally recorded sentences from prosodically distinctive languages, specifically, French, Turkish and Hungarian, respectively.

Before presenting the experiments in detail, it is important to clarify what they can and cannot demonstrate. First, and most importantly, if English speakers can segment speech in a language unknown to them, they can do so for three possible reasons. First, we might have selected a set of languages that happen to be particularly easy to segment for English speakers, while they might fail to segment other languages that we did not test. While we cannot completely rule out this possibility, we attempted to minimize its likelihood by selecting a typologically diverse set of languages from three different language families and, importantly, with different prosodic characteristics than English. Second, the prosodic boundary cues might be realized similarly in all languages. That is not to say that all prosodic cues are realized in exactly the same way in all languages; for example, sign languages such as American Sign Language and tone languages such as Thai will realize prosodic constituents using different cues than a language such as English. Still, a subset of prosodic boundary cues might be similar across the languages of the world, and these cues might be sufficient for English speakers to segment foreign languages they have never heard before. Third, the cues might differ in the languages we tested, but learners might be equipped with perceptual mechanisms capable of dealing with such variation in prosody. Given the difficulty of obtaining experimental sentences in the typologically diverse set of languages we used, our stimuli do not allow us to make acoustic measurements to decide between these possibilities (because we could not obtain sentences with phonemically matched material in different prosodic positions). In either case, however, if adult, monolingual English speakers can segment words from an unknown language, there exist cues that allow them to detect word-boundaries in the absence of language-specific knowledge, calling for a reconsideration of a longstanding assumption in the statistical word-segmentation literature.

Second, the primary function of word-segmentation is presumably to allow learners to store acoustic or phonological word forms in memory. However, all of our experiments employ a two-alternative forced choice task, asking participants to select between items consistent with word boundaries and items straddling word-boundaries. In parallel with previous research showing that such tasks do not allow for conclusions about whether learners store items in memory (Endress & Mehler, 2009b), we are also not licensed to draw conclusions about memory storage from our experiments. The goal of

our experiments is, therefore, to show that participants can exploit prosodic cues to word-boundaries in languages different from their native language, leaving unresolved the question of whether they memorize any word forms.

Third, and as mentioned above, we tested an adult population as a means of exploring a developmental question. We based this decision on the assumption that experiments on adults provide a strong test of the question of whether language-universal cues to word-boundaries exist and can be exploited by non-native learners. Indeed, adults have already acquired a native language (and have well documented problems acquiring a second language; see e.g., Birdsong & Molis, 2001; Lenneberg, 1967; Johnson & Newport, 1989); hence, if they can detect word-boundaries in an unknown language, they must have access to some mechanism that enables this ability, and whatever this mechanism is, it is plausible that it might enable infants to learn whichever language they are exposed to in early development. Further, given the fact that acquisition of a native language imposes constraints on subsequent acquisition, especially in adulthood, evidence that adults can segment continuous speech from a non-native language reveals that the mechanism is in place despite prior perceptual constraints on the relevant segmentation dimensions. In other words, infants might show an even stronger proclivity to extract and use prosodic information than adults given their more plastic and less constrained capacity for learning.

2. Experiment 1: Word segmentation at edges of French Intonational Phrases

Experiment 1 asks whether participants can use sentence-level prosody for recognizing words, even when no statistical information is given, and when the prosody belongs to a language unfamiliar to the participants. In particular, we recorded the prosody of a French sentence, and placed the target words systematically at the end of this prosodic contour.

Previous work has shown that sentence-level prosody has an effect on the segmentation of words from fluent speech. For example, Shukla, Nespor, and Mehler (2007) showed that learners recognize statistically well-formed items only to the extent that they are consistent with prosodic phrase-boundaries. In their experiments, participants were familiarized with a speech stream containing "words" with high TPs. The speech stream was synthesized as a sequence of Intonational Phrases. When words straddled prosodic break-points, they were not recognized or even rejected in a subsequent test phase.

However, these results leave open the possibility that word-candidates need to be generated by a mechanism sensitive to the distributional properties of the input stream; prosodic information may act to filter out such statistically defined word-candidates once they are available. Alternatively, word-candidates may be defined through prosodic information irrespectively of any statistical information. In the following experiment, we address this second possibility by asking whether learners can recognize word-candidates that are defined *only* through prosodic cues in a context where distributional information has been made entirely uninformative.

2.1. Materials and method

2.1.1. Participants

Fourteen native speakers of English (eight females, mean age 20.4 years, range 17–31) took part in this experiment. In this and all other experiments presented here, they participated for course credit or monetary compensation, were recruited through the Harvard study pool and had no known auditory impairments. Most participants were monolingual or had very limited instruction in a second language (less than 2 years), but our population included a limited number of English–Spanish and English–Mandarin bilinguals as well.[2] In all experiments, participants were randomly assigned to one of two language conditions (see below). No participant reported any knowledge of French.

[2] Including all participants who reported to have some experience with some other language, irrespective of proficiency or amount of instruction, there were four non-monolingual speakers in each of Experiments 1–3, 10 in Experiment 4, and seven in Experiment 5. In no experiment did we observe any difference between participants as a function of language experience. In the following, we thus present only the pooled data.

2.1.2. Apparatus

This experiment and all other experiments were run using Psyscope X software (http://psy.ck.sissa.it). Stimuli were presented over headphones; responses were collected from pre-marked keys on a keyboard.

2.1.3. Stimuli

All stimuli were synthesized using the de7 voice of mbrola (Dutoit, Pagel, Pierret, Bataille, & van der Vreken, 1996) with a fundamental frequency of 235 Hz and a phoneme duration of 120 ms.[3]

2.1.4. Pre-training

Before starting the experiment, participants completed a pre-training phase in order to familiarize them to the response keys. The pre-training 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.

2.1.5. Familiarization

Participants were told that they would listen to a monolog in "Martian" (a made-up language), and were instructed to find the words in the monolog. Then they listened to the familiarization speech stream.

The familiarization stream consisted of a concatenation of trisyllabic non-sense items. To make TPs uninformative, each of the three syllable position in the items could be occupied by four different syllables. That is, the initial syllable could be either /dEI/, /fa:/, /li:/ or /ku:/ (in SAMPA notation); the second syllable could be /maU/, /pi:/, /wE:/ or /zaI/, while the third syllable could be /Sa:/, /nOY/, /gO/ or /rE:/. In total, we thus obtained $4 \times 4 \times 4 = 64$ words. These items were randomly concatenated into a speech stream with 20 repetitions per word; the stream lasted approximately 15 min. Participants were presented twice with this speech stream, yielding a total familiarization of 30 min.

TPs among adjacent syllables were 0.25 within and between words; second-order TPs were approximately 0.25 within and between words (range 0.22–0.30). As TPs were identical within words and across word-boundaries, they provided no cues to word-boundaries. While each syllable occurred only in a specific position within a word, such positional information is not tracked unless prosody-like boundary cues are given, and, when such cues are given, it is tracked by mechanisms that are independent of TP-based mechanisms (Endress & Bonatti, 2007; Endress & Mehler, 2009a).

Four words were selected as target words (hereafter, just "words"); these words were always marked with French prosody (see below). Two different groups of participants were exposed to one of two languages such that, during test, "correct" items for one group of participants were "incorrect" choices for the other group, and vice-versa. In Language 1, these words were /dEImaUSa:/, /fa:pi:nOY/, /li:wE:gO/ and /ku:zaIrE:/. In Language 2, the words were /fa:zaIgO/, /ku:pi:Sa/, /li:maUnOY/, /dEIwE:rE:/. The speech stream was otherwise monotonous with the parameters described above.

To mark the target words with French prosody, we recorded the sentence "Le taureau va miner Nicolas" ("The bull will mine Nicolas") from a female native speaker of Parisian French. We chose this sentence because it is composed exclusively of CV syllables (the final consonants in 'miner' and 'Nicolas' are not pronounced), and because it lacks any semantic reading. We then measured the duration and pitch of each segment using Praat. These duration and pitch values were then linearly transformed so that their average matched that of the other phonemes in our speech stream; that is, the average phoneme duration was set to 120 ms, while the average fundamental frequency pitch was set to 235 Hz. The pitch and duration values for each segment in the target word are shown in Table 1.

This normalized prosody was placed such that all target words but no other word occurred in the end of the sentence; the syllables not located in such a prosodic contour had flat prosody, that is, all duration and pitch cues were neutralized.

[3] We used the de7 voice because it is of much better quality than the available English voices. Pilot tests with native speakers of American English showed that they found the synthesized speech with the de7 voice highly intelligible. Obviously, all phonemes we selected exist in American English.

Table 1
Prosodic parameters of the French sentence used for Experiments 1 and 2. All phonemes are given in SAMPA transcription.

Segment	Average pitch/Hz	Duration/ms	Segment	Average pitch/hz	Duration/ms
l	224	39	n	255	153
9	220	98	i	260	104
t	226	152	k	260	144
O	224	120	o	229	96
R	236	89	l	186	77
o	264	170	A	164	197
v	252	94			
A	203	130			
m	212	114			
i	244	100			
n	246	120			
e	257	161			

2.1.6. Test

Following familiarization, we presented participants with pairs of items, requesting them to choose the one item of each pair that they considered to be in Martian. One item was always one of the target words (marked by prosody). The other items (hereafter called the "foils") had occurred equally often in the familiarization stream as the target words – but never carried prosody. Foils for participants exposed to Language 1 were the target words for participants to Language 2 (see above), and vice-versa. Test items were presented with "flat" prosody, that is, all duration and pitch cues were neutralized.

The foils were chosen so as to minimize the number of times they occurred immediately before or after the target words; they occurred once or twice in such positions. We also minimized their occurrence in positions separated by one or two trisyllabic units before target words, and by one trisyllabic unit after a prosodically defined word. Summing over all of these positions, foils occurred in these positions at most three times.

We paired all words with all foils, yielding 16 test pairs in total. These test pairs were presented twice with different item orders. Test trials were presented in random order with the constraint of not repeating items in consecutive trials, and to have at most three answers of the same kind in a row.

2.2. Results and discussion

As shown in Fig. 1, participants preferred words to foils ($M = 60.9\%$, $SD = 11.1\%$), $t(13) = 3.69$, $p = 0.003$, Cohen's $d = 0.99$, $CI_{.95} = 54.5\%, 67.3\%$, with no difference between the languages they were exposed to, $F(1,12) = 0.96$, $p = 0.347$, $\eta^2 = 0.07$, ns. The performance in the first half of the test phase ($M = 61.2\%$, $SD = 11.7\%$) was higher than in the second half of the test phase ($M = 56.7\%$, $SD = 14.6\%$), $F(1,13) = 4.84$, $p = 0.046$, $\eta_p^2 = 0.271$.

Results from Experiment 1 showed that participants could use non-native, sentence-level prosody to identify word-candidates in a situation where co-occurrence statistics were made uninformative. In this experiment, words were located at the end of an Intonational Phrase. This may have lead participants to recognize word-candidates in two different ways. On the one hand, words might be recognized just by virtue of being in an Intonational Phrase at all, irrespectively of *where* in the Intonational Phrase they are located. On the other hand, and in line with the observation that words are learned predominantly at sentence boundaries (e.g., Dahan & Brent, 1999; Seidl & Johnson, 2006, 2008; Shukla et al., 2007), word-candidates might be recognized only when they are located in contour-edges. Experiment 2 tests these possibilities by embedding target words in the middle of the Intonational Phrase.

3. Experiment 2: Word segmentation in the middle of French Intonational Phrases

Experiment 2 asks whether the presence of prosodic information in a speech stream is sufficient for recognizing words, or whether words have to be placed in certain positions within prosodic units for prosodic information to be useful for word-extraction. More specifically, target words in Experiment 1

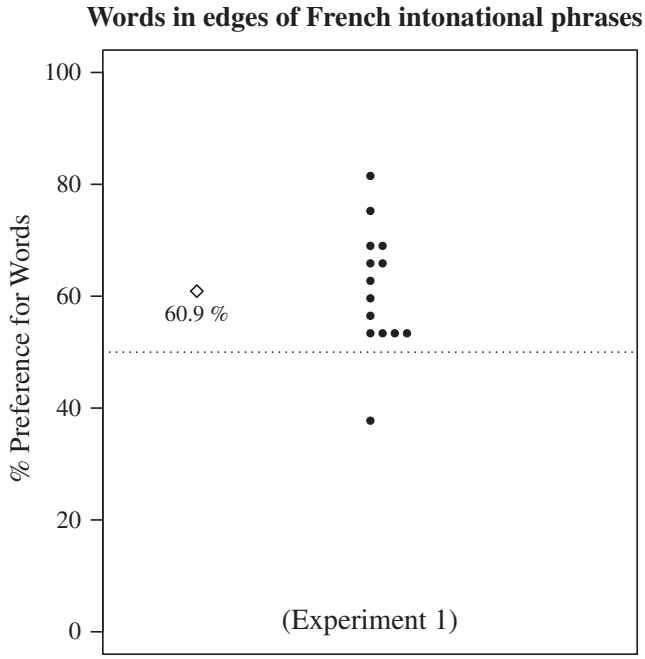

Fig. 1. Results of Experiment 1. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants preferred target words to equally frequent foils when the former were placed at the right edge of French Intonational Phrases, even though participants had no experience with French.

were placed at the ends of Intonational Phrases, a position that is known to be conducive for learning words (e.g., Dahan & Brent, 1999; Seidl & Johnson, 2006; Shukla et al., 2007). In Experiment 2, in contrast, we placed the target words in the middle of Intonational Phrases such that they straddled a word-boundary in the original French sentence; if the mere presence of prosodic structure is sufficient to allow learners to extract words, words should be preferred to foils in this case as well.

3.1. Materials and method

Experiment 2 was identical to Experiment 1 except that the Intonational Phrases were placed such that the first syllable of target words was the third syllable in the Intonational Phrases. In this way, target words were misaligned both with word-level and phrasal prosodic cues. The test phase was identical to that used in Experiment 1. Care was taken to minimize the occurrence of the foils in positions located one or two trisyllabic units before target words, and one trisyllabic unit after a prosodically defined word. This assured that foils would not be highlighted by prosodic information. Summing over all of these positions, foils occurred in these positions at most three times.

We tested 14 new native speakers of English (three females, mean age 19.1 years, range 18–21) in Experiment 2.

3.2. Results and discussion

As shown in Fig. 2, participants failed to prefer words over foils ($M = 50.4\%$, $SD = 7.1\%$), $t(13) = 0.23$, $p = 0.818$, Cohen's $d = 0.063$, $CI_{0.95} = 46.3\%, 54.6\%$, ns, with no difference between the languages they were exposed to, $F(1,12) = 0.21$, $p = 0.658$, $\eta^2 = 0.02$, ns. There was no difference in performance between the first half and the second half of the test phase, $F = 0$. Performance in Experiment 2 was, therefore, significantly different from that observed in Experiment 1, $F(1,26) = 8.9$, $p = 0.006$, $\eta^2 = 0.25$.

These results show that the mere presence of prosodic structure is not sufficient to allow people to extract words; rather, word-candidates seem to be recognized only to the extent that they are placed

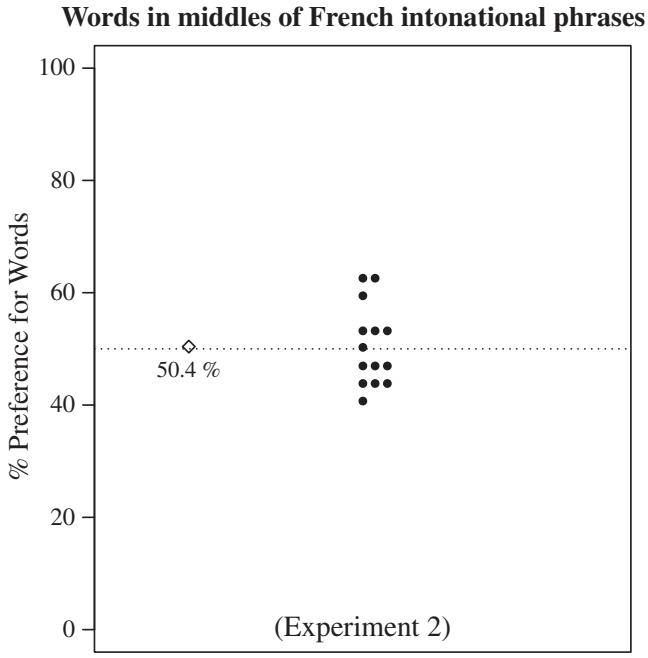

Fig. 2. Results of Experiment 2. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants did not prefer target words to equally frequent foils when the former were placed in the middle of French Intonational Phrases.

at the edges of prosodic contours. (While we embedded words only at the trailing edge of Intonational Phrases, prior experiments that used both the leading and the trailing edge suggest that we would observe an edge advantage for the trailing edge as well, see Seidl & Johnson, 2006; Shukla et al., 2007). Still, the Intonational Phrases in Experiment 1 (where participants successfully identified target words) were surrounded by syllables carrying "flat" prosody, a situation that might have facilitated the participants' task relative to more natural listening conditions; if so, participants might fail to use prosodic cues to word-boundaries in more natural situations. Experiments 3–5 address this possibility. In Experiment 3, we used real French sentences, and asked whether participants could segment words from these sentences, even if the words were placed in the interior of the sentences and not at their edges.

4. Experiment 3: Word segmentation from real French sentences

In the previous experiments, we used synthesized speech to investigate whether prosodic cues can be used to segment fluent speech, even though we used naturalistic prosody. While the stimuli we employed were artificial, this cannot explain why we obtained differences between Experiments 1 and 2, since both used synthetic speech. We are nonetheless left with the possibility that subjects might perform differently with natural speech. In Experiment 3, we addressed this possibility by presenting monolingual native speakers of American English with real French sentences.

In each trial, a French sentence was played three times. Then, participants heard two words, and had to choose which of these words had occurred in the sentence. One word always occurred in the sentence. The syllable sequence in the other word also occurred in the sentence, but straddled a word-boundary. For example, in the sentence "Deux cho*meurs trient* les dechets" ("two unemployed sort the trash"), participants would have to choose between "chomeurs" (unemployed) and "meutri" (wounded), which contains the same syllables as the second syllable of "chomeurs" and the "trient" (which is pronounced like 'tree'). Also, in this experiment co-occurrence statistics were uninformative, as each syllable occurred only once in each sentence.

4.1. Materials and method

4.1.1. Participants

Fourteen new native speakers of English (10 females, mean age 23.6 years, range 18–35) took part in this experiment. No participant reported any knowledge of French.[4]

4.1.2. Stimuli

All sentences and test items are listed in Appendix A. They were recorded from a female native speaker of Parisian French. Test items were recorded in isolation. Stimuli were recorded using a Sennheiser ME67 directional microphone connected to a Lenovo ThinkPad T61 computer running Adobe Audition 3.0, and saved in the aiff file format (44.1 kHz, 16 bit, mono).

In each sentence, syllables occurred only once. As is apparent from Appendix A, one test item associated with a sentence always occurred in that sentence, while the syllable sequence of the other test item occurred in the sentence, but straddled a word boundary. Prosodically, these word-boundaries correspond to Phonological Phrase boundaries or Prosodic Word boundaries.

4.1.3. Procedure

Participants were informed that they would hear French sentences played three times, and then two words, one of which had occurred in the sentence. We then instructed them to choose which of these words had occurred in the sentence. As mentioned above, the syllable sequences of both words occurred in the sentence. However, only one word actually occurred in the sentence, while the other straddled a (prosodic) word boundary.

Each of the 11 sentences/test item combinations was played twice with the test items in different orders. The order of the trials was randomized.

4.1.4. Baseline condition

In Experiments 3–5, we ran control conditions where participants had to decide which of our test items was more likely to be in French, Turkish or Hungarian, respectively, but without prior exposure to the sample sentence that the correct choice was taken from. That is, participants faced the same choices between "correct" items and foils as in Experiments 3–5, but without hearing the sentences from which these items were derived. These experiments reveal, therefore, the participants' baseline preference for correct items as opposed to foils. Fourteen participants completed each of the baseline experiments. In no baseline experiment did participants prefer correct items to foils (p's > 0.05); in the baseline to Experiment 5 (using Hungarian stimuli), however, participants had a statistically significant preference for *foils*. Moreover, the results of Experiments 3–5 were significantly different from those of their respective control experiments. Below, we test our results against a chance performance of 50%; we note, however, that testing against the performance in the baseline condition would yield very similar results.

4.2. Results and discussion

As shown in Fig. 3, participants successfully chose words over part-words (M = 65.3%, SD = 10.2%), $t(13) = 5.61$, $p < 0.0001$, Cohen's $d = 1.5$, $CI_{.95}$ = 59.4%, 71.1%. There were no differences in performance between the first half and the second half of the experiment, $F < 1$.

These results show that English speakers, with no familiarity to French, can nonetheless exploit French prosody to extract words from fluent speech. This shows that some prosodic cues, although characteristic of a specific language, can be extracted by non-native speakers. Prosodically, these cues correspond to Prosodic Word boundaries or Intonational Phrase boundaries, and might be realized in language-universal ways.

[4] Given that Spanish and French are closely related languages, we verified that Spanish-speaking participants ($N = 3$) did not perform differently from the remaining sample. While a sample size of three participants does not allow for robust conclusions concerning the relevance of language background, Spanish-speaking participants did not perform better than the remaining participants (in fact, their numeric scores were slightly worse than the sample average).

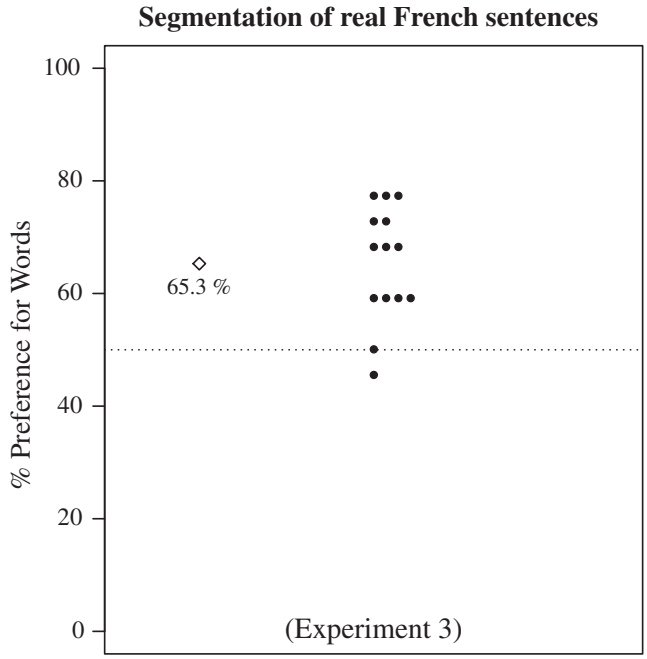

Fig. 3. Results of Experiment 3. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants prefer words to syllable sequences straddling word-boundaries when presented with real French sentences.

Moreover, the prosodic cues seemed to be highly efficient. In fact, the participants' performance after having heard a word only three times was comparable to that observed in typical segmentation experiments, where each word is often presented hundreds of times.

To further confirm that monolingual English speakers can segment speech from unknown languages, and especially languages with different prosodic properties, Experiments 4 and 5 represent attempts to replicate the results from Experiment 3 using sentences from Turkish and Hungarian, respectively.

5. Experiment 4: Word segmentation from real Turkish sentences

The results of Experiment 3 suggest that monolingual English speakers can use prosodic information from an unknown language to identify words in fluent speech. However, one may argue that, since the Norman invasion, English has been massively influenced by French. Consequently, English speakers may have some experience with French even if they (technically) never learned it as a language. Even though we made sure that the French words used in Experiment 3 were not related to English words, this may have allowed them to segment French, but not other languages. To control for this possibility, we replicated Experiment 3, but with Turkish sentences and words.

For the purposes of the current experiments, the prosodic properties of Turkish are very similar to those of French; both languages have word-final stress, and otherwise similar phonological properties (e.g., Christophe et al., 2003). One difference between French and Turkish is that Turkish has vowel-harmony. However, as both real words and foils were actual Turkish words, vowels in both types of test items harmonized, as did the vowels in the two words straddling the word boundary. Hence, while native speakers of a language with vowel-harmony can use this cue for word-segmentation (Suomi, McQueen, & Cutler, 1997; Vroomen, Tuomainen, & de Gelder, 1998), this cue was uninformative in the current experiments.

5.1. Materials and method

Experiment 4 was similar to Experiment 3, except that we used Turkish stimuli instead of French stimuli, and only 10 different sentences. All sentences and test items are given in Appendix B. They

were prepared by a native speaker of Turkish, and recorded from a different female native speaker of Turkish who was naïve regarding the purpose of the experiment.

We tested 14 new native speakers of English (eight females, mean age 23.7 years, range 15–32) in Experiment 4. No participant reported any knowledge of Turkish.

5.2. Results

As shown in Fig. 4, participants successfully chose words over part-words ($M = 63.5\%$, $SD = 11.4\%$), $t(12) = 4.24$, $p = 0.001$, Cohen's $d = 1.2$, $CI_{.95} = 56.6\%, 70.4\%$. There was no difference in performance between the first half and the second half of the experiment, $F(1,13) < 1.43$, $p = 0.252$, ns. The participants' performance did not differ between Experiments 3 and 4, $F < 1$.

5.3. Discussion

The results of Experiment 4 show that English speakers, lacking any experience with Turkish, can use Turkish prosody to find word-boundaries in fluent speech. This suggests that some cues to word-boundaries are not language-specific, and speakers can exploit those cues in the absence of any experience with a language.

Participants used the available prosodic cues despite the important differences between English and Turkish. More specifically, both Turkish and French have word-final stress, which contrasts with the predominant word-initial stress pattern in English. Further, and in contrast to head-initial languages such as English and French, Turkish is a head-final language. Prosodically, this implies that the main prominence within a Phonological Phrase comes first in Turkish and last in French (or English; see Nespor & Vogel, 1986); despite this difference, our English speakers segmented Turkish sentences. While our results do not allow us to determine exactly which cues were used by our participants, they do suggest that the prosodic correlates of the head-position are not among those cues, as these cues are different between English and Turkish. Possibly, participants might rely on Prosodic Word boundaries or Phonological Phrase boundaries to find words in an unknown language, as these seem to exist in all languages (e.g., Nespor & Vogel, 1986).

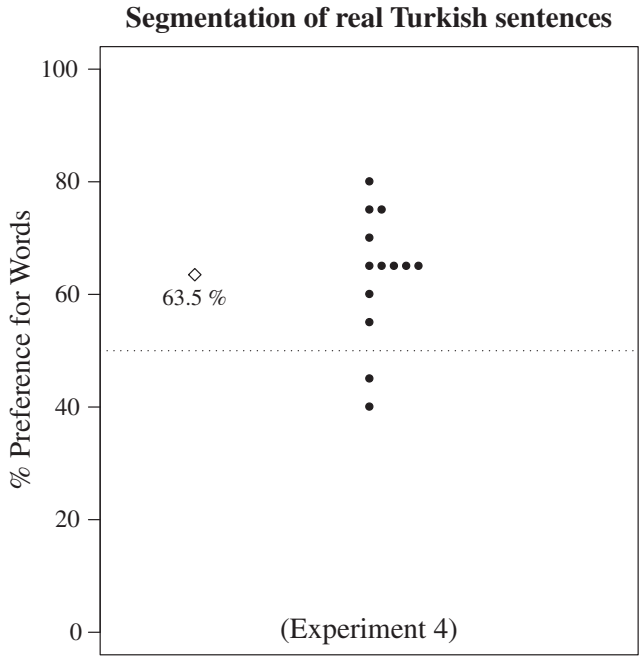

Fig. 4. Results of Experiment 4. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants prefer words to syllable sequences straddling word-boundaries when presented with real Turkish sentences.

In Experiment 5, we further extend this result by attempting a replication with Hungarian sentences; while Hungarian has word-initial stress like English, it is a head-final language like Turkish.

6. Experiment 5: Word segmentation from real Hungarian sentences

Experiment 5 further extends the results of Experiments 3 and 4 by asking whether native English speakers can segment Hungarian sentences. Like English, Hungarian has initial stress; if participants can segment Hungarian, then the results of Experiments 3 and 4 cannot be due to some obscure bias to assume word-final stress in unknown languages. Moreover, Hungarian would add a language from a third family to our sample of languages, thereby firming up the possibility that humans are endowed with a universal segmentation mechanism. While French is (as English) an Indo-European language, Turkish and Hungarian are Turkic and Finno-Ugric languages, respectively.

6.1. Materials and method

Experiment 5 was similar to Experiment 3, except that we used Hungarian stimuli instead of French stimuli, and as in Experiment 4, only 10 sentences. All sentences and test items are given in Appendix C. They were prepared by a native speaker of Hungarian, and recorded from a different male native speaker of Hungarian naïve regarding the purpose of the experiment. Fourteen new native speakers of English (eight females, mean age 23.3 years, range 18–30) took part in this experiment. No participant reported any knowledge of Hungarian.

6.2. Results

As shown in Fig. 5, participants successfully chose words over part-words (M = 64.3%, SD = 12.7%), $t(13) = 4.21$, $p = 0.001$, Cohen's $d = 1.1$, $CI_{.95}$ = 57.0%, 71.6%. There was no difference in performance between the first half and the second half of the experiment, $F < 1$. The performance in Experiment 6 differed neither from that in Experiment 3, $F < 1$, nor from that in Experiment 4, $F < 1$.

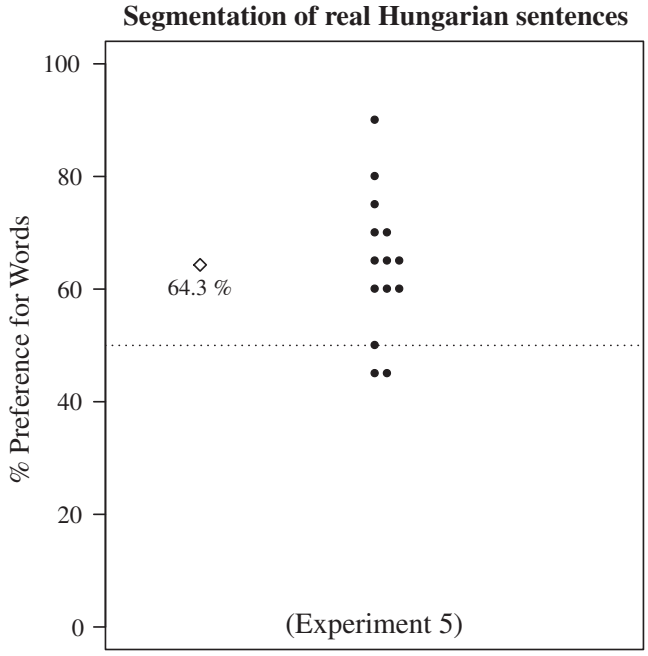

Fig. 5. Results of Experiment 5. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants prefer words to syllable sequences straddling word-boundaries when presented with real Hungarian sentences.

6.3. Discussion

The combined results of Experiments 3–5 demonstrate that largely monolingual speakers of English can identify word-boundaries in natural, fluent speech from three different language families. They do so in the absence of co-occurrence statistics (since each syllable occurred only once in each sentence). Hence, co-occurrence statistics such as TPs can be at most optional for word-segmentation. Conversely, participants seem to readily exploit prosodic cues from entirely unknown languages.

7. General discussion

Before infants can learn the meanings of the words of their native language, they have to locate words and their boundaries within the fluent speech to which they are exposed. To do so, infants (who do not yet know their future native language) must clearly rely on cues that do not require language-specific knowledge. However, while there are many speech cues to word-boundaries, and especially prosodic cues (e.g., Cutler et al., 1986; Cutler & Mehler, 1993; Cutler & Norris, 1988; Houston et al., 2004; Johnson & Jusczyk, 2001; Johnson et al., 2003; Jusczyk et al., 1993, 1999; Mattys & Samuel, 1997; Mattys et al., 1999; McQueen et al., 2001; Nazzi et al., 2006; Norris et al., 1997), it is commonly assumed that these cues are language-specific and cannot be used by infant learners (e.g., Aslin et al., 1998; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996; Swingley, 2005). For example, while stress is word-initial in English, it is word-final in Turkish. Assuming that words start with strong syllables would thus lead to successful word-segmentation in English but not in Turkish. One may think that infants could work around this problem by checking whether utterance-initial syllables are stressed; if they are, then stress is presumably word-initial. However, even in languages with relatively regular prosodic features such as English and French, this strategy would fail. For example, while most English words have initial stress, most English sentences do not. As sentences tend to start with determiners in English (e.g., articles such as "the"), the first stressed syllable is often the *second* syllable in an English utterance (e.g., on the first syllable of "donkey" in "The donkey pushed the cart"; see e.g., Nespor & Vogel, 1986). Even more dramatically, in a prosodically very regular, stress-final language such as French, the *initial* syllable of content-words is often stressed if these occur at the beginning of some prosodic constituents (e.g., Astésano, Bard, & Turk, 2007). To the extent that prosodic cues are language-specific, it is thus difficult to see how young infants could exploit them to extract words from fluent speech.

However, the fact that different languages have different prosodic properties does not necessarily imply that prosodic cues require language-specific knowledge to be exploited. In fact, it seems that the way in which languages differ in their prosodic properties does *not* depend on any given (spoken) language (some exceptions notwithstanding). Stress is a case in point. As discussed in the introduction, stress is implemented using three different dimensions, namely the loudness, pitch, and duration of a syllable (e.g., Ashby & Maidment, 2005; Hayes, 1995), and languages with word-initial stress (such as English) tend to rely more on the pitch and the loudness of a syllable to signal stressed syllables, while languages with word-final stress (such as French) tend to use duration and loudness (e.g., Hayes, 1995). Exactly the same correlation holds for auditory non-speech sequences: An increase in pitch is perceived as a group onset, while an increase in duration is perceived as a group-offset (Hay & Diehl, 2007; Woodrow, 1909; but see Iversen et al., 2008, for evidence that these biases might not be universal). It is thus possible that learners can interpret stress in an unknown language, because they know how to interpret the choice of cues that are used to signal stress.

Moreover, other prosodic properties than stress might be realized in very similar ways across the languages of the world. As mentioned in the introduction, prosody is hierarchically organized (e.g., Hayes, 1989; Nespor & Vogel, 1986). Prosodic constituents at the highest levels of the hierarchy are clearly language-universal; for example, in many cases, no language-specific knowledge is required to detect that an utterance has started or ended. The question is then whether learners can exploit prosodic boundaries at intermediate levels (i.e., at Phonological Phrase boundaries and Prosodic Word boundaries) to find word-boundaries in a language they do not know.

The experiments presented here address the possibility that learners may have the perceptual capacity to exploit prosodic cues from an unknown language. In Experiments 1 and 2, we familiarized mostly monolingual native speakers of English with speech streams where the TPs between all syllables were identical, and with no other distributional cues to word boundaries. Superimposed on these speech streams, we implemented prosodic cues by placing words in the end of French Intonational Phrases (Experiment 1); participants were more familiar with the prosodically cued items than with foils not so cued. Experiment 2 showed that the mere presence of prosodic structure was not sufficient for word extraction. In that experiment, target words were embedded in the middle of a French Intonational Phrase such that they were misaligned with both sentential and word-level prosody; under these conditions, participants did not discriminate the target words from foils. While Experiments 1 and 2 used artificial speech, Experiments 3–5 show that native speakers of American English can segment real French, Turkish and Hungarian sentences after substantially less exposure than in typical speech segmentation experiments. We believe that participants were able to exploit Prosodic Word boundaries or Phonological Phrase boundaries because some of the cues signaling such boundaries seem to be language-universal, providing a mechanism that can supervene over the acquired native language. Such a conclusion, however, needs to be confirmed by further experiments. Indeed, as mentioned in the introduction, there are several reasons for which English speakers might have the ability to segment the languages we tested. Leaving aside the (unlikely) possibility that we selected a set of languages that happened to realize prosodic boundary cues similarly to English, English speakers might segment foreign languages either because all languages share a subset of boundary cues that are sufficient for word-segmentation, or because adult speakers of English retain a sensitivity to boundary cues in foreign languages even if these cues differ from how they are realized in English. In either case, however, they are equipped with a prosodic mechanism that allows them to segment any language they might encounter.

In sum, learners can use prosody from entirely unfamiliar languages to segment words from fluent speech, even when the languages in question come from entirely unfamiliar language families (that is, Turkic and Finno-Ugric in the case of Turkish and Hungarian, respectively). Further, this ability is realized in adult learners who have already acquired a native language, a factor that clearly limits the acquisition of non-native linguistic material.

7.1. Relation to previous prosody-based word-segmentation mechanisms

We suggested above that learners may use universal prosodic cues for word-segmentation. However, several lines of evidence suggest that infants acquire rich knowledge about the prosodic structure of their native language (at least at the Prosodic Word level), and that they use this knowledge to learn words from fluent speech. For example, infants acquiring English as their native language prefer a strong-weak stress pattern (that is predominant in English) at 9 months of age, but not at 6 months of age (Jusczyk et al., 1993). Moreover, at 7.5 months, infants mis-segment weak-strong words from continuous passages. For example, they would extract the item /taris/ when they really heard the words "the guitar is" (because the last syllable of "guitar" is strong); at 10.5 months, in contrast, they can also segment weak-strong words (Jusczyk et al., 1999; see also Houston et al., 2004, for similar results with longer words). Related biases have been observed in English-speaking adults, who tend to segment words starting with strong initial syllables (e.g., Cutler & Norris, 1988). French-learning infants, in contrast, seem to start by segmenting stressed syllables (that are final in French), and show robust segmentation of bisyllabic words only by 16 months (Nazzi et al., 2006). That is, even though (European) French infants know a number of bisyllabic words at 11 months (e.g., Hallé & de Boysson-Bardies, 1994; Hallé & de Boysson-Bardies, 1996), they show a delay relative to their English learning peers in comparable experimental settings.

It is thus clear that infants learn the prosodic characteristics of their native language, and use this knowledge for segmenting (and sometimes mis-segmenting) words from fluent speech. This, however, does not imply that infants could not use apparently universal prosodic cues. First, as mentioned above, while prosodic properties like stress differ across languages, they might vary in language-universal ways, relying on basic perceptual grouping mechanisms. Second, cues to the boundaries of many prosodic constituents seem to be selected from a relatively small set of options across different

spoken languages (e.g., Christophe et al., 2001; Fon, 2002; Fougeron & Keating, 1997; Hoequist, 1983a; Hoequist, 1983b; Keating et al., 2004; Shattuck-Hufnagel & Turk, 1996; Vaissière, 1983), and, in sign language, even non-signers are sensitive to certain prosodic boundary cues (Brentari et al., in press; Fenlon et al., 2008). As a result, such universal cues might be useful for kick-starting word-segmentation. These cues might also allow infants to learn the more language-specific aspects of the prosodic organization of their native language.

The availability of these prosodic cues would be particularly important if, as suggested elsewhere, distributional word-segmentation cues such as transitional probabilities (e.g., Aslin et al., 1998; Saffran, Aslin, et al., 1996) cannot be used for word-segmentation because they rely on the wrong kinds of memory mechanisms (Endress & Mehler, 2009b), or because they are not reliable in many languages (Yang, 2004). If this is the case, then universal aspects of prosodic organization would provide, together with other potentially universal non-prosodic cues discussed below, one of the few candidate cues to bootstrap word-segmentation abilities, even if, or when distributional cues can be used.

7.2. Are prosodic cues sufficient to extract words from fluent speech?

Even though participants in our experiments were able to exploit prosodic cues, when these were the only available cues, it is unlikely that learners could rely exclusively on such cues; if they did, they would probably end up considering, say, clitic groups (such as "a dog") as words. This problem runs throughout our experiments, as participants most likely detected Phonological Phrase boundaries, and Phonological Phrases are often larger than single words (e.g., Nespor & Vogel, 1986). As mentioned above, our results thus share with previous word-segmentation experiments the problem that participants were clearly sensitive to word-boundaries; however, we have no evidence that they extracted any words and stored them as memory entries. Clearly, learners need to use some other cues to recover from segmentation errors such as considering clitic groups as words. From a computational point of view, TPs may well play an important role in the process of recovering from prosodic mis-segmentation; however, if, as we argued above, TPs feed into the wrong kind of memory mechanisms for storing words, then they cannot be used for this purpose either.

There may be other strategies that would allow learners to recover from such errors of segmentation. For example, infants seem to expect new words to start after the end of words they already know (e.g., Brent, 1997; Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Dahan & Brent, 1999); if they recognize, say, the word "dog" in fluent speech, they will expect a new word to start after "dog" should the speaker continue. Known words may thus be used to break up prosodic mis-segmentations such as "thedog."

If infants indeed use this (and other) strategies, one may ask how they could possibly learn words to use them to break up mis-segmentations. One possibility is suggested by the observation that infants preferentially learn words they hear in isolation (Brent & Siskind, 2001; Van de Weijer, 1999; but see Aslin et al., 1996); such words could evidently be used to break up mis-segmentations. However, it will be important to find out whether there are other strategies to recover from such mis-segmentations, and to segment words from fluent speech in the first place.

In sum, our results suggest that learners can exploit prosodic cues from entirely unknown languages to find some word-boundaries, even when co-occurrence statistics are made entirely uninformative. However, it is still unclear what the exact prosodic cues are that infants can exploit, and how infants recover from prosodic mis-segmentation. While our results provide a "feasibility proof" that prosodic cues can be used to segment entirely unfamiliar languages – and thus satisfy a general desiderata to uncover a universal segmentation process – answers to these questions will contribute to our understanding of how infants extract words from fluent speech.

Acknowledgments

Funding for this work was provided by MBB grants to M. Hauser, A. Nevins and A. Endress, as well as gifts from J. Epstein and S. Shuman. We are indebted to T. Beydola and Á. Kovács for preparing the Turkish and Hungarian sentences, respectively.

Appendix A. Phrases and test items used in Experiment 4

See Table A1.

Table A1
Phrases and test items used in Experiment 4. Translations are given in italicized characters.

Sentence	Word	Part-word
Mon cheval est mort	Cheval	Vallée
My horse is dead	*Horse*	*Valley*
Ce chapeau lisse la vie	Chapeau	Police
This hat smoothens life	*Hat*	*Police*
Une vipère dut s'éclater	Vipère	Perdu
A viper must have been blown up	Viper	Lost
Dix serpents sifflent dans le vent	Serpent	Pensif
Ten snakes whistle in the wind	*Snake*	*Thoughtful*
Notre courroux gît pour toujours	Couroux	Rougit
Our wrath is laid to rest forever	*Wrath*	*Become red*
Il faut compter les sous	Compter	Télé
One has to count the coins	*Count*	*Television*
Ce depart fait mal	Depart	Parfait
This departure is painful	*Departure*	*Perfect*
Vos bachots faisaient scandale	Bachot	Chauffait
Your skiffs created a scandal	*Skiff*	*Heatened*
Deux chomeurs trient les dechets	Chomeurs	Meurtri
Two unemployed sort the trash	*Unemployed*	*Wounded*
Trois defauts naissent chaque annee	Defaut	Faunesse
Three flaws are born each year	*Flaw*	*Feminine faune*
Il fuma chez les voisins	Fuma	Macher
He smoked at the neighbors' place	*Smoked*	*Chew*

Appendix B. Phrases and test items used in Experiment 5

See Table B1.

Table B1
Phrases and test items used in Experiment 5. Translations are given in italicized characters.

Sentence	Word	Part-word
Dün Özge mirastaki payını aldı	Özge	Gemi
Yesterday Özge got her share of the inheritance	*(Female name)*	*Ship*
Yarın elli mandalin siparis verelim	Mandalin	Limanda
Let us order fifty tangerines tomorrow	*Tangerines*	*Harbor*
Onun sayesinde ekşi şelale suyu içtim	EkşI	Şişe
I drank sour water from the waterfall because of him	*Sour*	*Bottle*
Sandalyeyi erikçi çekti	Çekti	Çiçek
The man who sells plums pulled the chair	*Pull*	*Flowers*
Boş oda varmış kızın evinde	VarmıŞ	Davar
There is an empty room in the girl's house	*Exist*	*Cattle*
Senin yüzüne hayatta bakmam	Bakmam	Tabak
I will not look at your face again	*Look*	*Plates*
Oldukça kıllı bir çocuk	Kıllı	Çakıl
He is a very hairy kid	*Hairy*	*Pebbles*
Yemek boyunca İlyas tıkladı durmadan	İlyas	Yastık
Ilyas tapped (on the table) continuously during the meal	*(Male name)*	*Pillows*
Bugün hain ekmekçi gelmemiş	Hain	Inek
Today the damned bread-man did not come	*Evil*	*Cow*
Ana musluğu değistir	Ana	Namus
Replace the main tap	*Main*	*Honor*

Appendix C. Phrases and test items used in Experiment 6

See Table C1.

Table C1
Phrases and test items used in Experiment 6. Translations are given in italicized characters.

Sentence	Word	Part-word
A két alma kacskaringós úton haladt tovább	Alma	Makacs
The two apples proceeded on the stubborn path	*Apple*	*Stubborn*
A zöld béka lapjától jobbra tettem le	Béka	Kalap
I put it to the right from the page of the green frog	*Frog*	*Hat*
A lassú séta poshadt vizekhez vezetett	Séta	Tapos
The slow walk led to stale waters	*Walk*	*Trample*
Minden olajos fekete ruha szondázásra került	Ruha	Haszon
All oily black clothes underwent a probe test	*Clothes*	*Benefit*
Hat fehér műanyag pohár masnira volt kötve	Pohár	Hármas
Six white plastic cups were made ribbons of	*Cup*	*Triple*
Csak a második piros talált párt máganak	Piros	Rosta
Only the second red found a pair for himself	*Red*	*Find*
A bíró kacsalábon forgó kastélyban lakott	Bíró	Róka
The judge lived in a castle that was on ducks legs	*Judge*	*Fox*
Aki betör pecsétet is visz magával	Betör	Törpe
Who breaks in also takes a seal with himself	*Break in*	*Dwarf*
Ahogy nyílik az ajtó csavarok esnek ki belőle	Ajtó	Tócsa
As the door opens screws fall out of it	*Door*	*Puddle*
A mértan árnyékában elenyészik a képzelet	Mértan	Tanár
Imagination vanishes in the shadow of geometry	*Geometry*	*Teacher*

References

Ashby, M., & Maidment, J. (2005). *Introducing phonetic science*. Cambridge, UK: Cambridge University Press.

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.

Astésano, C., Bard, E. G., & Turk, A. (2007). Structural influences on initial accent placement in French. *Language and Speech, 50*(3), 423–446.

Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition, 83*(2), 167–206.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language, 44*(2), 235–249.

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*(4), 298–304.

Brent, M. (1997). Toward a unified model of lexical acquisition and lexical access. *Journal of Psycholinguistic Research, 26*(3), 363–375.

Brentari, D., González, C., Seidl, A., & Wilbur, R. (in press). Sensitivity to visual prosodic cues in signers and nonsigners. *Language and Speech*.

Brent, M., & Cartwright, T. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*(1–2), 93–125.

Brent, M., & Siskind, J. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*(2), B33–B44.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33*(2), 111–153.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*(2–3), 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*(3), 1570–1580.

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy, 2*(3), 385–394.

Christophe, A., Nespor, M., Guasti, M. T., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science, 6*(2), 211–220.

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*(4), 523–547.

Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in japanese. *Journal of Memory and Language, 32*(2), 258–278.
Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of french and english. *Journal of Memory and Language, 25*(4), 385–400.
Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology. Human Perception and Performance, 14*(1), 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*(2), 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*(1), 218–244.
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 (Vol. 3, pp. 1393–1396). Philadelphia.
Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition, 105*(2), 247–299.
Endress, A. D., & Mehler, J. (2009a). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology, 62*(11), 2187–2209.
Endress, A. D., & Mehler, J. (2009b). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*(3), 351–367.
Fenlon, J., Denmark, T., Campbell, R., & Woll, B. (2008). Seeing sentence boundaries. *Sign Language & Linguistics, 10*(2), 177–200.
Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209–221.
Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104–113.
Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499–504.
Fon, J. (2002). A cross-linguistic study on syntactic and discourse boundary cues in spontaneous speech. Unpublished doctoral dissertation. Ohio State University, Columbus, OH.
Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101*(6), 3728–3740.
Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic.
Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access ii. infant data. *Journal of Memory and Language, 51*(4), 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*(3), 254–260.
Hallé, P. A., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior & Development, 17*(2), 119–129.
Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior & Development, 19*(4), 463–481.
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*(3), B53–B64.
Hay, J. S. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics, 69*(1), 113–122.
Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and phonology. Rhythm and meter* (vol. 1, pp. 201–260). Orlando, FL: Academic Press.
Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.
Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*(3), 269–286.
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.
Houston, D. M., Santelmann, L. M., & Jusczyk, P. W. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes, 19*(1), 97–136.
Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America, 124*(4), 2263–2271.
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*(4), 548–567.
Johnson, E. K., Jusczyk, P. W., Cutler, A., & Norris, D. (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology, 46*(1), 65–97.
Johnson, J., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60–99.
Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science, 12*(1), 131–141.
Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687.
Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3–4), 159–207.
Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2004). Domain-initial articulatory strengthening in four language. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation* (pp. 145–163). Cambridge, UK: Cambridge University Press.
Lenneberg, E. H. (1967). *Biological foundations of language*. New York: John Wiley and Sons.
Liddell, S. K. (1978). Nonmanual signals and relative clauses in American sign language. In E. Siple (Ed.), *Understanding language through sign language research* (pp. 59–90). New York: Academic Press.

Mattys, S. L., Jusczyk, P. W., Luce, P., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*(4), 465–494.
Mattys, S. L., & Samuel, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language, 36*(1), 87–116.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111.
McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in japanese speech segmentation. *Journal of Memory and Language, 45*(1), 103–132.
Mehler, J., Dommergues, J., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*(3), 298–305.
Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*(1), 271–280.
Morgan, J. L., & Demuth, K. (1996). *Signal to syntax. Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum.
Nazzi, T., Iakimova, G., Bertoncini, J., FrTdonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring french: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language, 54*(3), 283–299.
Nazzi, T., Nelson, D., Jusczyk, P., & Jusczyk, A. (2000). Six-month-olds? Detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy, 1*(1), 123–147.
Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Foris: Dordrecht.
Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*(3), 191–243.
Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*, 258.
Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*, 246–263.
Pilon, R. (1981). Segmentation of speech in a foreign language. *Journal of Psycholinguistic Research, 10*(2), 113–122.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 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*(1), 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.
Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*(1), 51–89.
Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science, 9*(6), 565–573.
Seidl, A., & Johnson, E. K. (2008). Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *Journal of Child Language, 35*(1), 1–24.
Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247.
Shatzman, K. B., & McQueen, J. M. (2006a). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science, 17*(5), 327–372.
Shatzman, K. B., & McQueen, J. M. (2006b). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics, 68*(1), 1–16.
Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology, 54*(1), 1–32.
Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language, 4*, 1–22.
Soderstrom, M., Seidl, A., Kemler Nelson, D., & Jusczyk, P. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*(2), 249–267.
Suomi, K., McQueen, J., & Cutler, A. (1997). Vowel harmony and speech segmentation in finnish. *Journal of Memory and Language, 36*(3), 422–444.
Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*(1), 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*(4), 706–716.
Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics, 67*(5), 867–875.
Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53–65). berlin: Hamburg, Germany.
Vaissière, J. (2005). Perception of intonation. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 236–263). Malden, MA: Blackwell.
Van de Weijer, J. (1999). *Language input for word discovery*. *Mpi series in psycholinguistics* (vol. 9). Nijmegen: Max Plank Institute for Psycholinguistics.
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*(2), 133–149.
Wakefield, J. A., Doughtie, E. B., & Yom, B.-H. L. (1974). The identification of structural components of an unknown language. *Journal of Psycholinguistic Research, 3*(3), 261–269.
Wilbur, R. B. (2009). Effects of varying rate of signing on ASL manual signs and nonmanual markers. *Language and Speech, 52*(Pt 2–3), 245–285.
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*(10), 451–456.