Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation

Karima Mersad · Thierry Nazzi

Published online: 11 February 2011
© Psychonomic Society, Inc. 2011

Abstract The present study explored the influence of a new metrics of phonotactics on adults' use of transitional probabilities to segment artificial languages. We exposed French native adults to continuous streams of trisyllabic nonsense words. High-frequency words had either high or low congruence with French phonotactics, in the sense that their syllables had either high or low positional frequency in French trisyllabic words. At test, participants heard low-frequency words and part-words, which differed in their transitional probabilities (high for words, low for part-words) but were matched for frequency and phonotactic congruency. Participants' preference for words over part-words was found only in the high-congruence languages. These results establish that subtle phonotactic manipulations can influence adults' use of transitional probabilities to segment speech and unambiguously demonstrate that this prior knowledge interferes directly with segmentation processes, in addition to affecting subsequent lexical decisions. Implications for a hierarchical theory of segmentation cues are discussed.

Keywords Word segmentation · Transitional probabilities · Phonotactics · Positional frequency · Adults · French

Natural speech presents no obvious markers of word boundaries, such as silent pauses. How infants, who are mostly exposed to utterances, discover the word forms of their native language and how adults parse native and nonnative speech into word units is known as the speech segmentation problem. Language-specific regularities in the native language can provide information on word boundaries. For example, since most content words are stress initial in English (Cassidy & Kelly, 1991; Kelly & Bock, 1988), detecting stressed syllables can help both adults and infants locate the beginning of a majority of words in English (Jusczyk, Houston & Newsome 1999; Nazzi, Dilley, Jusczyk, Shattuck-Hufnagel & Jusczyk 2005). Phonotactics, which refers to constraints on permissible sound sequences in a language, is another source of information on word boundaries. Experimental evidence has shown that both adults and infants treat frequent and infrequent sequences of phonemes of their native language differently and can use phonotactic information to retrieve word boundaries (Mattys, Jusczyk, Luce & Morgan 1999; McQueen, 1998). One goal of the present study was to explore the use of a new metrics of phonotactics—namely, syllable positional frequency—in speech segmentation.

There is another cue to word segmentation that is not specific to a given language or even to speech. In a sequence of items (visual, auditory, speech sounds), certain items tend to share a high co-occurrence frequency, leading them to be perceived as clusters. In the literature, the most studied type of co-occurrence is the transitional probability (TP), defined as the normalized version of the co-occurrence frequency [i.e., for items a and b , $TP(a/b) =$

K. Mersad · T. Nazzi
Laboratoire Psychologie de la Perception,
Université Paris Descartes,
Paris, France

T. Nazzi
CNRS UMR 8158,
Paris, France

K. Mersad (✉)
Laboratoire Psychologie de la Perception,
Institut Pluridisciplinaire des Saints Pères,
45 rue des Sts Pères,
75006 Paris, France
e-mail: crocnotekary@yahoo.fr

frequency of *ab*/frequency of *a*]. This kind of cue can potentially provide information at several levels of linguistic structure. At the lexical level, adults can use TPs between word classes to acquire the rudiments of syntax (Kaschak & Saffran, 2006). At the sublexical level, adults can use TPs between nonadjacent phonemes (Newport & Aslin, 2004), as well as TPs between adjacent or nonadjacent syllables, to extract words from continuous speech (Gomez, 2002; Saffran, Newport, & Aslin 1996). In the present study, we chose to focus on TPs computed at the syllabic level, since many studies have established the importance of the syllabic unit in speech processing in French (the language of our participants) in both adulthood and early infancy (Bertoncini, Floccia, Nazzi & Mehler 1995; Bijeljac-Babic, Bertoncini & Mehler 1993; Goyet, de Schonen & Nazzi 2010; Mehler, Dommergues, Frauenfelder & Segui 1981; Nazzi, Iakimova, Bertoncini, Fredonie & Alcantara 2006).

A satisfactory account of the way listeners segment the speech stream requires understanding how the different segmentation cues are exploited individually, as well as how they will combine and interact in different contexts. Mattys, White and Melhorn (2005) have proposed a model integrating multiple speech cues to account for adults' segmentation abilities. This model postulates that cues are hierarchically organized depending on the context, with descending weights allocated from *lexical* (word forms and semantic, syntactic, and pragmatic information), to *segmental* (information at the level of sublexical units such as phonemes and syllables), to *metrical prosody* (information related to stress assignment at the syllabic/lexical level) cues. According to this approach, lower level cues are used when higher level cues are absent or unusable.

The Mattys et al. (2005) model does not incorporate TPs and, thus, does not take into account the relative weight of TPs with respect to other cues. This might be due to the fact that there is an important difference between TPs and other segmentation cues. Segmentation based on metrical prosody, segmental, and lexical cues supposes prior knowledge of the properties of the native language. On the contrary, TP-based segmentation is more akin to a learning mechanism that operates on the incoming signal independently of prior knowledge. Crucially, it can, in fact, be thought of as (one of) the mechanism(s) that will allow the acquisition of the prosodic, phonotactic, and lexical patterns of the native language. In this perspective, some phonotactic constraints can be viewed as instances of transitional probabilities at a higher level of representation. Given that adults have acquired higher levels of knowledge, they might exploit them to segment speech, relying less on TP cues. Therefore, we need to better understand the link between segmentation based on TPs versus higher level linguistic cues, and some studies have begun to explore this issue.

In a few studies with adults, the segmentation issue when TP cues are convergent with other cues has been investigated. Saffran, Aslin, and Newport (1996) showed that when TP cues are accompanied by convergent prosodic cues, segmentation performance is significantly enhanced. More important for our topic, in a few other studies, the use of TP cues when they are pitted against other cues has been investigated. Two studies with adults explored the interaction with prosody. Shukla, Nespor and Mehler (2007) demonstrated that listeners fail to use TP information to segment words when they span a prosodic phrase boundary, while Toro, Sebastián-Gallés and Mattys (2009) established that pitch rise in the medial syllables of trisyllabic words decreases the effectiveness of TP segmentation for Spanish and English, but not for French, speakers. Fernandes, Ventura and Kolinsky (2007) investigated the way TPs and coarticulation interact and found that although coarticulation has priority when TPs and coarticulation are in conflict in optimal listening conditions, TPs override coarticulation in degraded speech.

In two studies with adults, the impact of phonetic/phonotactic properties on TP-based segmentation has been investigated. The first one was conducted by Onnis, Monaghan, Richmond and Chater (2005), who found that the use of nonadjacent TPs between syllables to segment an artificial language was critically contingent upon the dependency between the two plosive consonants occurring in word-initial and word-final positions, a regularity present in English, the native language of the adult participants. Moreover, Onnis et al. showed that even when there was no nonadjacent structure in the language, preference for phonemes in particular positions alone was sufficient to drive a preference for words over part-words in the test phase. This suggests that phonotactics can have effects both during the segmentation process and during the decision-making test phase.

The second study was conducted by Finn and Hudson Kam (2008), who showed that illegal phonotactic sequences impair adults' use of TPs between adjacent syllables for word segmentation. In a series of experiments, they exposed native English-speaking adults to continuous streams of syllables containing bisyllabic words that had high phoneme transitional probabilities (PTPs) but that started with consonant clusters not allowed word initially in English. After exposure, participants had to choose between words and part-words (which were bisyllabic items with low PTPs). The results showed that participants did not select the words when they started with illegal consonant clusters, whereas they did so in one control experiment in which the words had legal initial clusters. Both studies cited above suggest that TPs are ranked lower than phonotactic information.

The present study explored the role of positional frequency phonotactics on TP-based segmentation, first

extending Finn and Hudson Kam's (2008) results to a more ecological situation in which the phonotactic manipulation did not involve contrasting *illegal* and *legal* sequences, because illegal sequences have a null probability to be encountered within real words and this might have blocked their extraction as parts of wordlike units.

Positional frequency phonotactics is based on the frequency with which a given syllable occupies a given position (initial, middle, or final in the case of trisyllabic words) in French words. The choice of positional frequency phonotactic regularity (here, at the syllabic level, for the reasons explained above) was motivated by two reasons. On the one hand, sensitivity to this pattern of regularity exists early in development, as was shown by Jusczyk, Luce and Charles-Luce (1994) with phoneme positional frequency. On the other hand, preference for phonemes in particular positions has been found with adults—for example, the “bias against assigning word status to candidate strings that begin with a continuant consonant” found by Onnis et al. (2005) in a study with English participants. Note that positional frequency phonotactics could be envisaged as a generalization of the bias underlined by Onnis et al. beyond the specific instance of continuant consonants.

Second, in Finn and Hudson Kam (2008), phonotactics might have played a role in both the segmentation and the decision test phases. The fact that participants selected the words in a second control experiment in which the stream was presegmented by short pauses suggests that the null effect in the main experiment cannot have been due only to a (phonotactic) legality effect in the test phase. However, participants in the control experiment might have taken advantage of all cues available in the input (PTPs and pauses between words), and it remains possible that participants' preference for words (containing illegal clusters) over part-words (legal sequences) was decided during the test phase.

In the present study, our aim was to restrict the locus of the (potential) phonotactic effect to the segmentation phase. To this purpose, the target words (to be presented at test) were not directly manipulated. The phonotactic manipulation was made on the other words of the language, presented only during the familiarization phase. The rationale was that if native language phonotactics interfere with TP segmentation processes, it is highly likely that it will exert an effect on all the potential word sequences of the language. As a consequence, unlike Finn and Hudson Kam (2008), the present study did not put phonotactics and TP cues in direct conflict but tested the hierarchical position of these two cues in a different way. We asked whether learners can segment words using TP cues when they are embedded in a stream of other words that have a phonotactically infrequent pattern in the native language.

Lastly, while in Onnis et al., (2005), TPs were computed on nonadjacent syllables, in the present study, we considered computations between adjacent syllables. This difference is noteworthy because nonadjacent computations appear to be more difficult than adjacent ones and appear to be carried out only under certain conditions (Gomez, 2002). Therefore, the present study should establish whether the phonotactic bias is strong enough to impact easier statistical segmentation processes.

We constructed different languages, all based on the same template—that is, a 3-min-long continuous stream of four trisyllabic nonsense words in which TPs were reliable cues to word boundaries. After familiarization, adults were presented with pairs of target words and part-words that differed in TP properties (with TPs of target words higher than TPs of part-words) but had occurred with equal frequencies in the familiarization phase and had similar phonotactic properties (in fact, the sequences used as target words in half of the languages were part-words in the other half, and vice versa).

Crucially, languages were of two kinds, depending on the level of phonotactic congruency (high vs. low) with French phonotactics of some of the words used. In high-congruence languages, the two target words (which had low-to-medium phonotactic congruence) were embedded in a stream with two other, twice-as-frequent words that were phonotactically highly congruent with French words. In low-congruent languages, the two target words were embedded in a stream with two other, twice-as-frequent words that were phonotactically weakly congruent with French words. As was mentioned earlier, the phonotactic manipulation consisted of varying the positional frequencies with which the syllables of the words occupy a given position in French words. For example, one of the words of the high-congruence languages (denoted L+), /supivã/, was phonotactically highly congruent with French words, because the syllable /su/ is frequent in first position, the syllable /pi/ is frequent in second position, and the syllable /vã/ is frequent in third position of trisyllabic French words. In contrast, in the low-congruence languages (denoted L–), all syllables of the “word” /pivãfa/ can appear in their respective positions in trisyllabic French words, but infrequently.

The effects of phonotactics, originating from the familiarization phase (since the phonotactic manipulation was made on the words not used at test), should result in better performance in the L+ than in the L– conditions. Moreover, note that the same test items were used on each pair of high- versus low-congruence languages. Thus, if a positional frequency phonotactic effect was to be found, it would unambiguously establish that its impact occurs during exposure to the language, rather than during the test/decision phase. Lastly, since TPs and positional frequency phonotactics were not put in direct conflict, the

present experiment should establish whether even a subtle phonotactic manipulation of the linguistic context can influence the use by adults of TPs to segment speech.

Method

Participants

Sixty-four native French-speaking students from the Université Paris Descartes, with no reported language or hearing impairment were tested. They received 5 € for their participation.

Stimuli

Familiarization stimuli As is shown in Table 1, the same 12 CV syllables were used to make eight different languages (L1A+, L1B+, L1A−, L1B−, L2A+, L2B+, L2A−, and L2B−). Each language was a concatenated stream of 270 tokens, constructed by repeating two trisyllabic words 90 times each (the frequent words) and two trisyllabic words 45 times each (the target words). The concatenation was pseudorandom, with the constraints that the same word should never occur twice in a row and each frequent word should be followed by the other frequent word half of the time. TPs between two frequent words were then equal to .5. TPs within a word were all equal to 1.

For the high-congruence languages (the “+” languages), the two frequent words had a phonotactic pattern frequent in French words—that is, each of their syllables was selected to

be highly frequent in its position in trisyllabic French words. Conversely, in the low-congruence languages (the “−” languages), the two frequent words had rare but legal phonotactic patterns for French—that is, each of their syllables was very infrequent in its position in trisyllabic French words (see Table 2). Positional frequencies were determined in trisyllabic French words using the adult database Lexique 3 (New, Pallier, Brysbaert & Ferrand 2004). Importantly, the level of phonotactic congruency of a language was determined by the frequent words, while target words were the same in L+ and L− language pairs.

Test stimuli For each language, the words used at test were the two target words and two part-words. Each part-word was a three-syllable sequence spanning a boundary between two frequent words. The part-words were of two types: For the languages L1A+, L1B+, L2A−, and L2B−, they corresponded to the last syllable of a frequent word plus the two first syllables of the other frequent word (final syllable, initial syllable, medial syllable; FIM type); for the languages L1A−, L1B−, L2A+, and L2B+, they corresponded to the last two syllables of a frequent word plus the first syllable of the other frequent word (medial syllable, final syllable, initial syllable; MFI type) (see Table 1). These complex manipulations on part-word formation were made necessary by the need to avoid high positional frequency part-words (e.g., for L2B−, /sēlāge/ could not be used, so that /lāgekō/ had to be used). Importantly, though, the part-word structure was counterbalanced between the L+ and L− languages, which was done given studies suggesting that some structures give stronger

Table 1 Words used for the high- (L+) and the low-congruence (L−) languages

	Familiarization		Test	
	Target Words	Frequent Words	Target Words	Part-Words
L1A+		/fatōke/		
	/gesēlā/	/supivā/	/gesēlā/	/kesupi/
L1A−	/røkōdu/	/tōkesu/	/røkōdu/	/vāfato/
		/pivāfa/		
L1B+		/kōduge/		
	/kesupi/	/sēlāro/	/kesupi/	/gesēlā/
L1B−	/vāfato/	/dugesē/	/vāfato/	/røkōdu/
		/lārøkō/		
L2A+		/fatovā/		
	/durosē/	/supike/	/durosē/	/tovāsu/
L2A−	/lāgekō/	/kefato/	/lāgekō/	/pikefa/
		/vāsupi/		
L2B+		/kōduru/		
	/tovāsu/	/sēlāge/	/tovāsu/	/durosē/
L2B−	/pikefa/	/gekōdu/	/pikefa/	/lāgekō/
		/rōsēlā/		

Table 2 Positional frequency of each syllable of the frequent trisyllabic words used during familiarization

L+			L−		
%Initial	%Middle	%Final	%Initial	%Middle	%Final
fa	tɔ	ke	tɔ	ke	su
72.2	69.7	91.7	6.3	7.9	18.6
su	pi	vã	pi	vã	fa
74.0	86.6	88.2	11.5	3.4	.1
kõ	du	ge	du	ge	sẽ
78.0	70.0	86.9	27.4	6.7	11.8
sẽ	lã	rø	lã	rø	kõ
86.9	70.4	97.8	4.2	.1	.1
			ke	fa	tɔ
			.35	27.8	23.9
			vã	su	pi
			8.8	7.4	1.8
			ge	kõ	du
			6.3	21.4	2.4
			rø	sẽ	lã
			2.0	1.2	25.3

segmentation results than do others (Saffran, Newport, & Aslin 1996). Words and part-words appeared equally often in familiarization. They differed in their TPs, which were equal to 1 for words and to .75 (average of 1 + .5) for part-words.

The target words and the part-words were the same for the following pairs of languages: L1A + and L1A−, L1B + and L1B−, L2A + and L2A−, and L2B + and L2B−. This was done to create identical conditions in the test phase for the high- and low-congruence languages and to limit positional frequency phonotactic impact to the familiarization phase. The roles of the words and the part-words were reversed between each pair of “A” and “B” languages. Thus, by construction, target words and part-words were matched in both frequency during familiarization and positional frequency phonotactics, whereas they crucially differed in TPs, words having higher TPs than did part-words. For test words (used both as targets and part-words according to the languages), the positional frequency of the first, the second, and the third syllables, respectively, were the following: /gesêlã/: 6.3, 1.3, 25.3; /røkõdu/: 2.1, 21.4, 2.4; /kesupi/: .4, 7.4, 1.8; /vãfat /: 8.4, 27.8, 23.9; /durøsẽ/: 27.3, 0.1, 11.8; /lãgekõ/: 4.2, 6.8, 0.1; /t vãsu/: 6.3, 3.4, 18.7; /pikefa/: 11.6, 7.9, 0.02. These test words have medium-to-low positional frequency phonotactic congruency, a property we will discuss in the General Discussion section.

Familiarization and test stimuli were all synthesized with MBROLA (Dutoit, Pagel, Pierret, Bataille & Van der

Vrecken 1996), using the French female diphone (fr2). All segments had the same duration (111 msec) and f_0 (200 Hz). There were no pauses or acoustic cues to word boundaries in familiarization sequences. All languages lasted 2.99 min. An increasing and decreasing amplitude ramp was applied to the first and last 5 seconds of the stream, respectively, to ensure that no word corresponded to the fade-in or the fade-out of the familiarization stream.

Procedure

Participants were tested individually in a quiet room with a PC computer. They were instructed to listen to an artificial language through headphones for 3 min and were informed that the listening phase would be followed by questions about their knowledge of the language. After familiarization, participants performed a two-alternative forced choice test. They heard pairs consisting of a word and a part-word, separated by 500-ms silences, and had to indicate, after each pair, which of the two items was more likely to be a word of the language by clicking either 1 or 2 on the computer screen. The participants' responses were separated by 2-sec intervals from the next trial. Every word of a language was contrasted with every part-word, leading to four word/part-word pairs. Each pair was presented twice in two different blocks, the order of the two items of each pair being reversed between the two presentations. The order of the presentation of the pairs was randomized between subjects. Each of the eight languages (L1A+, L1B+, L1A−, L1B−, L2A+, L2B+, L2A−, and L2B−) was presented to an equal number of participants.

Results

Figure 1 shows average scores broken down by congruency condition, together with individual performances. An ANOVA with the between-subjects factors level of congruency (high vs. low congruency), test list (“A” vs. “B” languages), and part-word type (FIM [last syllable + first two syllables] vs. MFI [last two syllables + first syllable of a frequent word]) first revealed a significant main effect of congruency, $F(1, 56) = 5.47, p = .023$.

In the high-congruence language group, participants' average score was 70.8% ($SD = 23.6$) of correct responses, which was significantly different from chance, $t(31) = 4.98, p < .01$. This result shows that native French speakers are able to extract words from these artificial languages, relying on TP information, phonotactic information, or both. In the low-congruent language group, participants' average score was 56.1% ($SD = 27.6$), which was not different from chance level, $t(31) = 1.26, p = .22$. This result reveals that

Fig. 1 Mean performance broken down by language condition (high vs. lowcongruence). Dots represent the mean scores of individual participants, the triangle the population average, and the dotted line the chance level of 50%

in the low-congruence condition, participants could not distinguish words from part-words, despite the fact that the same TP information was available in this condition, as compared with the high-congruence condition.

The comparison of the results between the two conditions thus suggests that phonotactic information played a role in the segmentation of the high-congruence languages. But what about the use of TPs? Several studies have shown that French, as well as Italian and English, adults use TPs between syllables to segment speech (Bonatti, Pena, Nespor & Mehler 2005; Endress & Bonatti, 2007; Saffran, Newport, & Aslin 1996). However, in our study, adults in the high-congruence language condition could have used only phonotactic information to retrieve the high-congruence words, using them as anchors to further segment the speech stream. A similar effect was found in one experiment by Onnis et al., (2005), in which adults showed a preference for words over part-words, as defined by the position of phonemes, even when TPs were removed from the artificial speech stream. However, in Onnis et al.'s study, adults likely used position of phonemes *and* TPs when both were available in the stimuli, as suggested by the better performance in this condition than in the phonotactic-cue-alone condition. Thus, even though we did not directly test this comparison in our study, it appears more likely that adults in the high-congruence language condition were using both TPs and phonotactic information to perform segmentation (see Saffran, Newport, & Aslin, 1996, for a similar argument when TPs and prosody are involved).

While the ANOVA failed to reveal an effect of language, $F(1, 56) < 1$, $p = .9$, there was a marginally significant effect of part-word type, $F(1, 56) = 3.04$, $p = .08$, but no interaction between congruency and part-word type. Interestingly, participants tended to have lower scores when the part-words were of the FIM type ($M = 57.8\%$, $SD = 30.4$) than when the part-words were of the MFI type ($M = 68.9\%$, $SD = 23.2$). At first sight, this seems contradictory to previous lines of research in similar distributional segmentation tasks. Indeed, Giroux (2008) showed that adults learn better the two final syllabic units of trisyllabic words than the two initial ones. Saffran, Newport and Aslin (1996) found that participants confused words with non-words containing the two final syllables of the words more often than with nonwords containing the two initial syllables of the words. However, if we consider that, as compared with initial syllables, not only is the identity of the two final syllables better memorized, but also their position within the words, then the presence of two final syllables at the beginning of MFI part-words might make them easier to distinguish from words than would the presence of two initial syllables at the end of FIM part-words. At any rate, note that the present marginal part-word type effect cannot explain the main congruency effect, since it was counterbalanced across high- and low-congruence languages. Lastly, there was no significant interaction between any other factors (congruency $\times$ list, $F(1, 56) = 1.82$, $p = .18$; congruency $\times$ part-word, list $\times$ part-word, congruency $\times$ list $\times$ part-word, $F < 1$).

Discussion

In the present study, we exposed French-speaking adults to artificial speech streams, made of a continuous concatenation of nonsense words, in which two thirds of the tokens had either high or low congruence with French phonotactics. When asked at test to choose between words and part-words, participants' responses depended on the congruency level of the language they had heard. Only in the high-congruence languages did participants perceive syllable sequences sharing high TPs between their syllables as wordlike units, preferring them to sequences with low TPs. Since TP information was equally available in the high- and the low-congruence languages, it is likely that participants could not benefit from it in the latter case. Because the experiment was constructed so that the same words and part-words presented at test were used in the two kinds (low vs. high congruence) of languages, the present findings cannot be due to an effect of phonotactics on the lexical decision test phase. The present study thus unambiguously establishes that the phonotactic effects are located in the learning/segmentation phase, confirming indirect evidence from Finn and Hudson Kam's (2008) study. Therefore, it appears that the ability of adults to use TPs to segment the same target words depends partially on the characteristics of the words in which these target words are embedded.

The present study establishes the effect of a phonotactic regularity, which we called *positional frequency phonotactics* and refers to the regularity with which a given syllable appears in a given position in words of the native language. This extends Finn and Hudson Kam's (2008) results to a case in which the phonotactic manipulation is implemented in terms of degree of congruency, rather than legality—a more ecological situation, since illegal items should, by definition, not be encountered by listeners in a given language. It also extends positional effects found by Onnis et al. (2005) to the syllabic level and beyond the sole class of continuant consonants.

Although the present study cannot identify the exact processes that dropped scores to chance level in the L− languages, some elements of our stimuli and results shed light on this issue. It is noteworthy that the target words had medium-to-low congruency with French positional frequency phonotactics. Therefore, in the L+ languages, adults were able to segment and judge as words sequences of syllables that were not optimal in French. Hence, the failure in the L− languages might be due to the fact that all the words had low congruence, while half of the words (representing two thirds of the tokens heard, since they were twice as frequent) had high congruence in the L+ languages. Therefore, in the L− languages, because TPs gave reliable cues to finding the words of the languages and phonotactics did not, and because the two kinds of

information were always contradictory in these languages (contrary to the L+ languages), it appears that adults favored high-level stable information, rather than TPs. This raises the question, to explore in the future, about whether it is possible to find experimental conditions (such as a longer familiarization) in which an artificial language can be segmented if all its words have low congruence. However, consistent mismatch of phonotactic and TP information might prevent adults' use of TPs, and it might be that adults can segment only artificial languages with at least one high-congruent word. Such a possibility would suggest that these words behave as perceptual/memory anchors, reminding one of the role of known words in the incremental distributional regularization optimization (INCDROP) model of word segmentation (Brent, 1999; Dahan & Brent, 1999). This possibility is also in line with the *segmentation by lexical subtraction* mechanism proposed by White, Melhorn and Mattys (2010) to account for segmentation based on known words by adults (for native language segmentation, see Mattys et al., 2005; for second language segmentation, see White et al., 2010) and infants (Bortfeld, Morgan, Golinkoff & Rathbun 2005; Mersad & Nazzi, 2011).

Taken together, the present results have implications for a hierarchical model of segmentation cues in adulthood (Mattys et al., 2005). Indeed, we found that when the positional dependency phonotactics of some of the words of the language is low congruent, the use by adults of TP cues for segmentation is impaired. It thus appears that phonotactic cues override TPs when these cues are in either direct (Finn & Hudson Kam 2008) or indirect (present study) conflict and on the basis of either legality status (Finn & Hudson Kam 2008) or degree of congruency (the present study). It further appears that phonotactics overrides TPs when relevant TP information bears on nonadjacent elements (Onnis et al., 2005), but also when it relates to adjacent elements, a condition that should have favored the use of TPs (Gomez, 2002). Together with similar data on TPs versus coarticulation (Fernandes et al., 2007) or prosody (Shukla et al., 2007; Toro et al., 2009), our results suggest that TP cues are located at the bottom of the hierarchy. This possibility is somehow intriguing, given the importance of TP-based segmentation demonstrated in numerous studies. However, one possible explanation is that adults have a bias to use stable native language regularities (lexical, phonological, . . .) rather than the novel linguistic contingencies that can be derived through TPs from exposure to a second language or an artificial language using TPs.

Moreover, the way the segmentation cues are ranked can be accounted for in terms of their distance from the lexicon. Indeed, as can be observed in the Mattys et al., (2005) hierarchical model, the lower the cues in the segmentation

hierarchy, the less they involve representations that have a direct link with the words of the vocabulary. Accordingly, segmental cues should follow lexical cues in the hierarchy, since they correspond to sublexical knowledge built on specific *word exemplars*. In turn, metrical prosody cues should follow, since they are often related to categories of words that need not be defined in terms of phonetic content (e.g., the trochaic pattern in English). In this perspective, TP cues would be lowest in the hierarchy, since they are not linked to the existing lexicon of the listener and can operate on incoming new material, as in artificial language experiments.

However, the weighting of the segmentation cues for adults in the situation of learning a new language and for young infants in the situation of learning their native language should be very different. In young infants, prosodic, phonotactic, and lexical knowledge needs to be acquired and consolidated before it can be used for segmentation. For example, infants acquire the native trochaic pattern around 6 to 9 months (Höhle, Bijeljac-Babic, Herold, Weissenborn & Nazzi (2009); Jusczyk, Cutler & Redanz 1993) and phonotactic patterns around 10 months (Friederici & Wessels 1993; Mattys et al., 1999; Nazzi, Bertoncini & Bijeljac-Babic 2009). Therefore, TPs should be crucial in younger infants, and their relative weight, as compared with other cues, should decrease during the first year of life. Evidence regarding the interplay of TPs and prosody support this prediction (Johnson & Jusczyk 2001; Johnson & Seidl 2009; Thiessen & Saffran 2003), which should be investigated in the future.

Acknowledgments This study was conducted with the support of ANR Grant 07-BLAN-0014-01 to TN. We thank Louise Goyet, Arielle Veenemans, Irène Fasiello, and Simon Barthelmé for their comments and discussion.

References

- Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: Rhythmical basis of speech representations in neonates. *Language and Speech*, 38, 311–329.
- Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology*, 29, 711–721.
- Bonatti, L., Pena, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, 16, 451–459.
- Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me. *Psychological Science*, 16, 298–304.
- Brent, M. (1999). An efficient probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, 34, 71–105.
- Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, 30, 348–369.
- Dahan, D., & Brent, M. R. (1999). On the discovery of novel word-like units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General*, 128, 165–185.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1393–1396). Wilmington: Applied Science and Engineering Laboratories.
- Endress, A. D., & Bonatti, L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, 105, 247–299.
- Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: a matter of signal quality. *Perception & Psychophysics*, 69, 856–864.
- Finn, A. S., & Hudson Kam, H. C. (2008). The curse of knowledge: first language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, 108, 477–499.
- Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, 54, 287–295.
- Giroux, I. (2008). *Structure et dynamique du système de traitement du langage oral: Quel est le statut des unités sous-lexicales?* Unpublished doctoral thesis, Université de Bourgogne, Dijon.
- Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431–436.
- Goyet, L., de Schonen, S., & Nazzi, T. (2010). Syllables in word segmentation by French-learning infants: An ERP study. *Brain Research*, 1332, 75–89.
- Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). The development of language specific prosodic preferences during the first half year of life: evidence from German and French. *Infant Behavior & Development*, 32, 262–274.
- 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, 1–20.
- Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, 12, 131–141.
- Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64, 675–687.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginning of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.
- Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630–645.
- Kaschak, M. P., & Saffran, J. R. (2006). Idiomatic syntactic constructions and language learning. *Cognitive Science*, 30, 43–63.
- Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 389–403.
- Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38, 465–494.
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477–500.
- McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39, 21–46.
- Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20, 298–305.

- Mersad, K. & Nazzi, T. (2011). *When Mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech*. Manuscript in preparation.
- Nazzi, T., Bertoncini, J., & Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial-coronal effect in the first year of life. *The Journal of the Acoustical Society of America*, 126, 1440–1446.
- Nazzi, T., Dilley, L. C., Jusczyk, A. M., Shattuck-Hufnagel, S., & Jusczyk, P. W. (2005). English-learning infants' segmentation of verbs from fluent speech. *Language and Speech*, 48, 279–298.
- Nazzi, T., Iakimova, G., Bertoncini, J., Fredonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54, 283–299.
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36, 516–524.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162.
- Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language*, 53, 225–237.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- 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.
- Shukla, M., Nespors, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54, 1–32.
- Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39, 706–716.
- Toro, J. M., Sebastián-Gallés, N., & Mattys, S. L. (2009). The role of perceptual salience during the segmentation of connected speech. *European Journal of Cognitive Psychology*, 21, 786–800.
- White, L., Melhorn, J. F., & Mattys, S. L. (2010). Segmentation by lexical subtraction in Hungarian L2 speakers of English. *The Quarterly Journal of Experimental Psychology*, 63, 544–554.