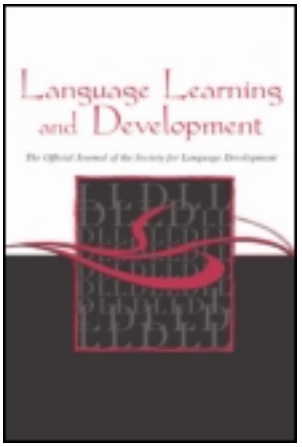

Language Learning and Development
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hlld20

When Mommy Comes to the Rescue of Statistics: Infants Combine Top-Down and Bottom-Up Cues to Segment Speech

Karima Mersad [a] & Thierry Nazzi [a]
[a] Laboratoire Psychologie de la Perception, Université Paris Descartes

Available online: 18 May 2012

To cite this article: Karima Mersad & Thierry Nazzi (2012): When Mommy Comes to the Rescue of Statistics: Infants Combine Top-Down and Bottom-Up Cues to Segment Speech, Language Learning and Development, 8:3, 303-315

To link to this article: http://dx.doi.org/10.1080/15475441.2011.609106

Language Learning and Development, 8: 303–315, 2012

ISSN: 1547-5441 print / 1547-3341 online
DOI: 10.1080/15475441.2011.609106

Psychology Press
Taylor & Francis Group

When Mommy Comes to the Rescue of Statistics: Infants Combine Top-Down and Bottom-Up Cues to Segment Speech

Karima Mersad and Thierry Nazzi

Laboratoire Psychologie de la Perception, Université Paris Descartes

Transitional Probability (TP) computations are regarded as a powerful learning mechanism that is functional early in development and has been proposed as an initial bootstrapping device for speech segmentation. However, a recent study casts doubt on the robustness of early statistical word-learning. Johnson and Tyler (2010) showed that when 8-month-olds are presented with artificial languages where TPs between syllables are reliable cues to word boundaries but that contain words of varying length, infants fail to show word segmentation. Given previous evidence that familiar words facilitate segmentation (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005), we investigated the conditions under which 8-month-old French-learning infants can succeed in segmenting an artificial language. We found that infants can use TPs to segment a language of uniform length words (Experiment 1) and a language of nonuniform length words containing the familiar word "maman" (/mamã/, mommy in French; Experiment 2), but not a similar language of nonuniform length words containing the pseudo-word /mãma/ (Experiment 3). We interpret these findings as evidence that 8-month-olds can use familiar words and TPs in combination to segment fluent speech, providing initial evidence for 8-month-olds' ability to combine top-down and bottom-up speech segmentation procedures.

INTRODUCTION

Infants learn the words of their native language mostly from utterances that lack reliable separation between consecutive words. One cue suggested to be crucial at the onset of infants' segmentation abilities is transitional probabilities (TPs). This cue, which can be formalized in different ways, refers basically to the notion that in a sequence of items (visual, auditory, e.g., music -, speech sounds) some items tend to co-occur frequently leading them to be perceived as clusters (Curtin, Mintz, & Christiansen, 2005; Swingley, 2005). In the present study, TPs will refer to the normalized version of the co-occurrence frequency (i.e., for items a and b, TP (b/a) = frequency of ab/ frequency of a).

Saffran, Aslin, and Newport (1996a) were the first to demonstrate that 8-month-old infants can rely on TPs to find words in continuous speech. They familiarized 8-month-olds with a two-minute artificial stream made up of four trisyllabic nonsense words where the only cue to

Correspondence should be addressed to Karima Mersad, Laboratoire Psychologie de la Perception, Institut Pluridisciplinaire des Saints Pères, 45 rue des Sts Pères, 75006 Paris, France. E-mail: karima.mersad@gmail.com

word boundaries was higher TPs for consecutive syllables within words than between words. During the test phase, infants discriminated the words from the part-words (which was replicated with stimuli better controlled for frequency by Aslin, Saffran, & Newport, 1998). Further studies showed that this kind of regularity can be computed by adults on adjacent and nonadjacent syllables or phonemes (Endress & Bonatti, 2007) and on the consonantal tier of words (Bonatti, Pena, Nespor, & Mehler, 2005). Moreover, recent work demonstrated that statistical word-form segmentation facilitates subsequent word-object mapping in infants (Graf-Estes, Evans, Alibali, & Saffran, 2007), supporting the hypothesis that segmentation mechanisms based on TPs play a role in lexical acquisition. Furthermore, such statistical learning may be domain- and species-general as it has also been observed with musical tones and visual stimuli (Fiser & Aslin, 2001; Saffran, Johnson, Aslin, & Newport, 1999) and in other mammals (Hauser, Newport, & Aslin, 2001). Hence, TPs are regarded as a powerful learning mechanism that is functional early and has been proposed as an initial bootstrapping device for segmentation (Saffran, Newport, & Aslin, 1996b).

However, a recent study casts doubt on the robustness of early statistical word-learning (Johnson & Tyler, 2010). These authors presented 5- and 8-month-old Dutch-learning infants with two kinds of artificial languages in which TPs were a reliable cue to word boundaries but that differed on their word length. One language was made of words of uniform length (four bisyllabic words), and the other was made of words of varying length (two bisyllabic and two trisyllabic words). While the infants segmented the language made of words of uniform length, they failed to show segmentation of the language made of words of different length. These data suggest that TPs might not be robust enough to scale up to the challenge of natural languages (Johnson & Tyler). Hence, TP use in artificial speech segmentation might be the artifactual product of the word-length uniformity characterizing the stimuli used in the early studies in this domain, which results in a drop of TPs every three syllables obviously not present in natural languages (Yang, 2004).

Nevertheless, other studies with English-learning infants bring data suggesting that infants are indeed able to use TPs in natural language situations. First of all, Pelucchi, Hay, and Saffran (2009a, 2009b) showed that English-learning 8-month-olds are able to use TPs to segment stimuli in a natural unfamiliar language (Italian), at least when using infant-directed speech and bisyllabic words with the predominant trochaic stress pattern of English (see also recent findings by Lew-Williams, Pelucchi, & Saffran, in press, showing infants' combined use of TPs and isolated presentations of unfamiliar words in similar conditions). Next, Thiessen, Hill, and Saffran (2005) established that English-learning 8-month-olds can use TPs to segment a language with two disyllabic and two trisyllabic words when the stimuli are pronounced in infant-directed (though not adult-directed) speech. In the above studies, the stimuli were recorded by a speaker rather than being synthesized and were broken down in sentences of about 10 syllables, rather than presented as a continuous stream of two to three minutes. Both elements make these stimuli more ecological than those used in previous studies. However, this raises the question of the relative weight of TPs and other segmentation cues. Indeed, a myriad of bottom-up segmentation cues other than TPs are provided by natural speech, and, as stated in Pelucchi et al. (2009b), natural language stimuli do not tell us "the degree to which infants actually use statistical cues for word segmentation" (p. 682). Infants could have used prosodic boundary information (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Gout, Christophe, & Morgan, 2004; Nazzi, Kemler-Nelson, Jusczyk, & Jusczyk, 2000), phonotactic constraints (Mattys & Jusczyk, 2001), allophonic information (Hohne & Jusczyk, 1994), or coarticulation (Johnson & Jusczyk, 2001).

In the Pelucchi et al. (2009b) studies, infants' ability to use TPs could also have been favored by the fact that they had to segment trochaic rather than iambic units from the speech stream, even though this feature could not explain the observed results on its own (Jusczyk, Houston, & Newsome, 1999; Nazzi, Dilley, Jusczyk, Shattuck-Hufnagel, & Jusczyk, 2005). In Thiessen et al. (2005), adults could not distinguish between words and part-words when presented with isolated sentences (extracted from the infant-directed stimuli) providing no relevant TP information, suggesting that these cues were not sufficient, without TPs, to segment words. However, this additional data do not preclude the possibility that there might have been subtle cues in the signal (such as prosodic cues) that infants might have been more sensitive to than adults.

The present study explores in a controlled experimental setting the conditions under which 8-month-olds can use TPs to segment a language. The infants were learning French, a syllable-based language in which TP use has not been investigated. French is an interesting language to investigate, given differences in early segmentation abilities reported between syllable- and stress-based languages (Gout, 2001; Goyet, de Schonen, & Nazzi, 2010; Jusczyk, Houston, et al., 1999; Kooijman, Hagoort, & Cutler, 2005, 2009; Nazzi et al., 2005; Nazzi, Iakimova, Bertoncini, Fredonie, & Alcantara, 2006; Polka & Sundara, in press), suggesting that differences in TP use between the two types of languages might exist. Accordingly, following Aslin et al. (1998), Experiment 1 was conducted to first determine if French-learning 8-month-olds can use TPs to segment artificial languages of uniform word length.

Experiments 2 and 3 were conducted to test whether French-learning 8-month-olds can segment artificial languages of nonuniform length on the basis of TPs and, if so, under which conditions. Given the results with Dutch-learning infants (Johnson & Tyler, 2010), it was hypothesized that infants at that age might fail to do so unless the language of nonuniform word length contains both TP cues and an additional cue. In the present study, the additional segmentation cue selected was the presence of several occurrences of a familiar word in the fluent speech stream. This was motivated by the fact that results in such conditions would also provide information about 8-month-olds' ability to combine bottom-up cues (here TP information) and top-down cues (word recognition) to find words in continuous speech streams. The word /mamã/ (Mommy in French) was chosen given evidence that by 6 months of age, English-learning infants are able to link this word appropriately to their own mothers (Tincoff & Jusczyk, 1999) and to use it to segment adjoining words from fluent speech (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). Accordingly, two experiments were run, exploring whether TP-based segmentation of artificial languages of nonuniform length is modulated by the presence of the familiar word /mamã/ (Experiment 2) versus its absence, /mamã/ being then replaced by the pseudoword /mãma/ (Experiment 3).

EXPERIMENT 1: METHOD

Participants

Twenty infants (eight females) were tested (mean age = 8 months, 19 days; range: 8 months, 3 days to 9 months 1 day). Data from three additional infants were excluded due to fussiness. All infants were born full-term and were exposed to French at least 80% of the time in their environment. None of the infants was known to suffer from hearing impairment.

Stimuli

Two language variants (A and B) were constructed in which the role of words and part-words was counterbalanced (see below and Table 1) in order to control for possible differences in stimuli salience.

The familiarization streams of each language were constructed by concatenating four trisyllabic nonce-words (hereafter words), repeating two words 90 times each (the frequent words), and the two other words 45 times each (the target words). Words are listed in Table 1. TPs within words were all equal to 1. The concatenation was pseudo-random, with no pauses or other cues to word boundaries, with the following constraints: the same word never occurred twice in a row and each frequent word was followed by the other frequent word half of the time (resulting in TPs between the two frequent words equal to .5). The resulting languages were identical to those used in Aslin et al. (1998).

The words used at test were the two target words and two part-words. The part-words were constructed by concatenating the last syllable of a frequent word and the first two syllables of the other frequent word. Words and part-words were then all trisyllabic (see Table 1). TPs within words were equal to 1 whereas TPs within part-words were equal to .75 (average of .5 for the two syllables crossing the frequent word boundary and 1 for the two syllables taken from the second frequent word). Hence, TPs within words were higher than TPs within part-words. However, words and part-words had appeared equally frequently in the familiarization strings due to the way part-words were made from the frequent words. Test stimuli were presented in lists each consisting of 15 occurrences of one of the words or part-words separated by 500 ms silences. All lists were 13.65 s long.

In order to control the acoustic properties of the stimuli, familiarization and test stimuli were all synthesized with MBROLA (Dutoit, Pagel, Pierret, Bataille, & Van der Vrecken, 1996) using the French female diphone database (fr2). All phonemes had the same duration (111 msec) and F0 (200 Hz). There were no pauses or acoustic cues to word boundaries in the familiarization sequences. The familiarization streams lasted 3.03 minutes with a 4.45 syllable/s speech rate. An increasing and decreasing amplitude ramp was applied, respectively, to the first and last 5 seconds of the streams to ensure that words corresponding to the fade in or the fade-out of the familiarization were not clearly audible.

TABLE 1
Stimuli used in Experiment 1

Language A		*Language B*	
words	*part-words*	*words*	*part-words*
pabiku	tudaro	tudaro	pabiku
tibudo	pigola	pigola	tibudo
golatu[f]		budopa[f]	
daropi[f]		bikuti[f]	

[f]frequent words.

Procedure and design

The experiment was conducted inside a sound-attenuated room, in a three-sided test booth. The test booth had a red light and a loudspeaker (SONY xs-F1722) mounted at eye level on each of the side panels and a green light mounted on the centre panel. A video camera used to monitor infants' behavior was placed below the center light. A Dell Optiplex computer, a TV screen connected to the camera, and a response box connected to the computer were located outside the sound-proof room. The box was controlled by the observer, who was looking at the video of the infant on the TV screen. The response box allowed the observer to send to the computer the information about the infant's head direction and hence to control the flashing of the lights and the presentation of the sounds. The observer, and also the infant's caregiver, wore earplugs and listened to masking music over tight-fitting headphones, which prevented them from hearing the stimuli presented. Information about the direction and duration of the head-turn and the total trial duration were stored in a data file on the computer.

The procedure used in the present study was the version of the Headturn Preference Procedure (HPP) adapted by Saffran et al. (1996a). Each infant was held on a caregiver's lap seated in a chair in the centre of the test booth. Each session began with a familiarization phase during which infants heard the language. The stimuli were delivered by the loudspeakers via an audio amplifier (Marantz PM4000). During the 3.03-minute familiarization, the central and lateral lights were alternatively made to blink to keep infant's attention, but orientation times were not measured. Note that in this phase of the experiment, the blinking of the lights was not contingent on the infant's head-turns.

The test phase, immediately following the end of the familiarization phase, consisted of three blocks, each corresponding to the presentation of the four lists (one for each word and part-word of the language) in a pseudo-random order within each block. Each test trial began with the green light on the center panel blinking until the infant had oriented in that direction. Then, the center light was extinguished and the red light above the loudspeaker on one of the side panels began to flash. When the infant made a turn of at least 30° in the direction of the loudspeaker, the stimulus of the trial began to play. Stimulus was played to completion or stopped immediately after the infant failed to maintain the 30° head-turn for 2 consecutive seconds (200 ms fade-out). If the infant turned away from the target by 30° in any direction for less than 2 s and then turned back again, the trial continued, but the time spent looking away was not included in the orientation time. Thus, the maximum orientation time for a given trial was the duration of the entire speech sample. If a trial was less than 1.5 s, the trial was repeated and the original orientation time was discarded. The flashing red light remained on for the entire duration of the trial.

Half of the infants were assigned to language A, in which the words were /pabiku/ and /tibudo/ and the part-words /tudaro/ and /pigola/, and the other half to language B, in which the words were /tudaro/ and /pigola/ and the part-words /pabiku/ and /tibudo/ (see Table 1).

RESULTS AND DISCUSSION

Infants showed significantly longer orientation times to part-words (8.48 s, SD = 2.65) than to words (6.68 s, SD = 2.34), $t(19) = 3.25$, $p = .004$. Fourteen of the 20 infants showed this pattern. Mean orientation time differences for test items are presented in Figure 1, left panel.

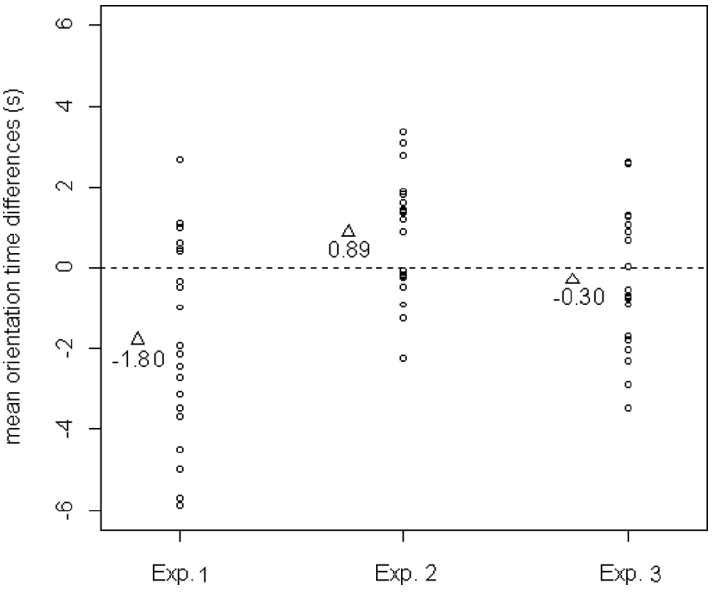

FIGURE 1 Mean orientation time differences (words minus part words) for Experiments 1 (words of uniform length), Experiment 2 (words of nonuniform length + familiar word /mamã/), and Experiment 3 (words of nonuniform length + pseudo-word /mãma/). Each dot represents the mean orientation time differences of a given infant, and each triangle the group average.

These results establish that French-learning 8-month-olds can use TP cues to segment an artificial language of uniform word length, exhibiting a novelty pattern of preference. Therefore, they appear to behave like English- and Dutch-learning infants of the same age when it comes to the use of TPs, even though crosslinguistic differences have been previously found between these infant populations when it comes to segmenting speech using prosodic cues. With this first result in hand, Experiments 2 and 3 were conducted to explore how French-learning 8-month-olds segment an artificial language of nonuniform word length containing TP cues to word boundaries (a kind of language that Dutch-learning infants failed to segment at that age). As explained earlier, we started with Experiment 2 in which infants were tested in a condition in which the language is potentially made easier by the insertion of an additional cue, namely the presence of repeated occurrences of a known word (/mamã/, Mommy in French).

EXPERIMENT 2: METHOD

Participants

Twenty infants (eight females) were tested (mean age = 8 months, 13 days; range: 7 months, 29 days to 9 months 1 day). Data from two additional infants were excluded due to fussiness.

All infants were born full-term and were exposed to at least 80% of French in their environment. None of the infants was known to suffer from hearing impairment.

Stimuli

As in Experiment 1, two counterbalanced languages were constructed. For each language, the familiarization streams were constructed in two steps. The first step consisted of concatenating four nonce-words (hereafter words), two disyllabic, and two trisyllabic words by repeating the two trisyllabic words 90 times each (the frequent words) and the two disyllabic words 45 times each (the target words). Words are listed in Table 2. TPs within words were all equal to 1. The concatenation was pseudo-random, with no pauses or other cues to word boundaries with the following constraints: the same word never occurred twice in a row and each frequent word was followed by the other frequent word half of the time (resulting in TPs between the two frequent words equal to .5). The resulting artificial languages had the same structure as the one used in the mixed word condition (MWL) in Johnson and Tyler (2010).

In the second step, we inserted in the language stream a fifth word: the word /mamã/ which is very frequent in the input, as confirmed by the parents of the tested infants all reporting that /mamã/ was the name used to refer to the infant's mother. Specifically, we added the word /mamã/ before 1/5 of the occurrences of each word of the language. However, in order to maintain TPs for part-words identical to those of the Johnson and Tyler (2010) experiment (.5), the word /mamã/ was never introduced between two frequent words; hence, it only occurred before a frequent word when that frequent word was not preceded by the other frequent word. Overall, the word /mamã/ appeared 54 times in the streams and had a number of occurrences comparable to the other words constituting the language.

The words used at test were the two target words and two part-words. The part-words were constructed by concatenating the last syllable of a frequent word and the first syllable of the other frequent word. Words and part-words were then all disyllabic. TPs within words were equal to 1, TPs within part-words were equal to .5, and words and part-words were equally frequent in the language. Test stimuli were presented as described in Experiment 1. Familiarization and test stimuli were all synthesized as described in Experiment 1.

TABLE 2
Stimuli used in Experiment 2 (similar stimuli were used in Experiment 3, except that the known word /mamã/ was replaced by the pseudo-word /mãma/)

Language A		*Language B*	
words	*part-words*	*words*	*part-words*
pabi	tuda	tuda	pabi
tibu	pigo	pigo	tibu
golatu[f]		budopa[f]	
daropi[f]		bikuti[f]	
mamã		mamã	

[f]frequent words.

Procedure and design

The procedure and design were identical to the one used in Experiment 1, except that in language A, the words were /pabi/ and /tibu/ and the part-words /tuda/ and /pigo/ and in language B the words were /tuda/ and /pigo/ and the part-words /pabi/ and /tibu/ (see Table 2).

RESULTS AND DISCUSSION

Infants oriented significantly longer to words (M = 6.71 s, SD = 1.59) than to part-words (M = 5.87 s, SD = 1.25), t(19) = 2.53, p = .02. Twelve of the 20 8-month-old infants showed this pattern. Mean orientation time differences for test items are presented in Figure 1, middle panel.

The significant preference for words over part-words establishes that infants discriminated the items with high TPs from the items with low TPs despite their equal frequency in the stream, and despite the varying length of the words composing the language.[1] Importantly, the languages used in the present experiment had a structure similar to the language used in Johnson and Tyler (2010), except that the present languages contained the familiar word /mamã/ inserted 1/5 of the times before each word in the stream. Hence, if Johnson and Tyler (2010) failed to show that 8-month-old infants can segment artificial speech when the words are not of uniform length, the present results suggest that 8-month-olds can overcome the mixed word length difficulty, provided that the language contains other segmentation cues. Here, the additional cue was the presence of a familiar word in the stream that has been found to facilitate segmentation in a different experimental situation (Bortfeld et al., 2005).

However, an alternative interpretation is that French-learning infants are better at processing TP information in a context of mixed length words than Dutch-learning infants. Another possibility is that infants were successful in segmenting the words in the present experiment not because they used the familiar word /mamã/ to anchor speech segmentation but rather because any word appearing repeatedly before the target words would have provided an additional segmentation cue. Experiment 3 was run to clarify these issues, in which the stream was exactly the same as in the present experiment except that the word /mamã/ was replaced by the pseudo-word /mãma/, obtained by reversing the order of the two syllables. Note that because the two syllables share the same consonant, /mamã/ and /mãma/ only differ in the order of their two vowels that minimally differ by the nasality feature.

[1]The observed direction of preference, a familiarity preference, differs from most prior statistical learning studies, which report novelty effects (Johnson & Tyler, 2010; Aslin et al., 1998; Saffran et al., 1996b). Nevertheless, statistical learning sometimes results in familiarity effects (Thiessen et al., 2005; Thiessen & Saffran, 2003, Johnson & Seidl, 2009) which have been interpreted according to Hunter and Ames (1998)'s model as effects associated to situations of relatively difficult segmentation. Accordingly, the familiarity effect in Experiment 2 can be due to the fact that segmentation of languages made of words of different lengths and containing a familiar word, although possible at 8 months, is more difficult than segmentation of languages made of words of uniform length. Such an interpretation is contingent upon the fact that French-learning 8-month-olds show a novelty effect when presented with a language made of words of uniform length. This is precisely the pattern found in Experiment 1 of the present study with the design used in Aslin et al. (1998).

EXPERIMENT 3: METHOD

Participants

Twenty infants (seven females) were tested (mean age = 8 months, 21 days; range: 8 months, 13 days to 8 months 30 day). Data from three additional infants were excluded due to fussiness. All infants were born full-term and were exposed to at least 80% of French in their environment. None of the infants was known to suffer from hearing impairment.

Stimuli

The familiarization stimuli corresponded to the language stream of Experiment 2, in which the word /mamã/ was replaced by the word /mãma/. The test stimuli were the same as in Experiment 2.

Procedure and design

The procedure and design were identical to the ones used in Experiment 2.

RESULTS AND DISCUSSION

Infants did not show a significant difference in orientation times to words (7.02 s, SD = 1.61) and part-words (7.32 s SD = 1.99), $t(19) = .79$, $p = .44$. Nine of the 20 infants listened longer to words than to part-words. Mean orientation time differences for test items are presented in Figure 1, right panel.

To compare the outcome of the two experiments with the languages of nonuniform word length, mean orientation times were subjected to a two experiments (Experiment 2 vs. Experiment 3) x two test-item (word vs. part-word) ANOVA, with repeated measure on the second factor. There was a marginally significant main effect (of experiment), $F(1, 38) = 3.82$, $p = .06$ corresponding to slightly longer orientation times for Experiment 3 (M = 7.16, SD = .60) than for Experiment 2 (M = 6.29, SD = .22). There was no significant effect of test-item $F(1,38) = 1.09$, $p = .3$. Importantly, there was a significant interaction between the two factors, $F(1,38) = 5.05$, $p = .03$. This shows that when infants are presented with an artificial stream in which words are defined by their high TPs, if the words are of different lengths (two and three syllables), the addition in the stream of the known word /mamã/ helps the infants to retrieve the words from the speech stream while the addition of the pseudo-word /mãma/ does not. Note that an inspection of Figure 1 shows that this effect is not due to a few outliers, but rather a genuine group effect.

The present results allow us to discard the two alternative explanations raised in the discussion of the segmentation results of Experiment 2. First, French-learning infants are not better than Dutch-learning infants (Johnson & Tyler, 2010) at segmenting languages made of words of varied length when no familiar word is present. Second, segmentation in Experiment 2 was not due to

the addition, in the familiarization streams, of any word, but to the addition of a specific word, the familiar word /mamã/.

GENERAL DISCUSSION

Infants' ability to use TPs as a speech segmentation cue appears to break down when the language is made of words of different lengths. This idea was proposed by Johnson and Tyler (2010) in a study suggesting that 5- and 8-month-old Dutch-learning infants fail to segment languages made of words of two and three syllables. In Experiment 1, we first presented French-learning 8-month-olds with languages made of four trisyllabic words using the same design as in Aslin et al. (1998). Results established that French-learning 8-month-olds, like English- and Dutch-learning infants of the same age, are able to extract the words from languages made of words of uniform length. Therefore, our study establishes that even though differences have been found regarding the use of rhythmic cues by infants learning syllable-based languages on the one hand (Gout, 2001; Goyet et al., 2010; Nazzi et al., 2006; Polka & Sundara, in press) and stress-based languages (English: Jusczyk, Houston, et al., 1999; Dutch: Kooijman et al., 2005, 2009) on the other hand, TPs are used to segment fluent speech by infants learning these two types of languages by 8 months of age.

Second, we evaluated TP-based segmentation of artificial languages in which the words have different lengths. To do so, we constructed languages with similar structures to Johnson and Tyler (2010) and then inserted in the streams, 1/5 times before each word, either the known word /mamã/ (Mommy in French, Experiment 2) or the pseudo-word /mãma/ (Experiment 3). Tested on these languages, French-learning 8-month-olds showed a significant preference for words over part-words when the known word /mamã/ was present in the familiarization stream but no preference when the pseudo-word /mãma/ was present. These results suggest that the infants have extracted the words from the languages made of words of varying length when the speech stream contained several occurrences of the word /mamã/ (hence providing experimental evidence that French-learning 8-month-olds know that word), but not when it contained the pseudo-word /mãma/. Therefore, although TP computations have been proposed as an important mechanism that enables infants to bootstrap the other segmentation cues, our results concur with those of Johnson and Tyler (2010) on Dutch-learning infants in showing that French-learning 8-month-olds also fail to segment languages made of words of varied length using TPs alone.[2]

Our study further establishes that this does not necessarily make TPs an irrelevant cue to segmentation in more complex languages such as natural languages. Indeed, we demonstrated that the addition of a known word is enough for infants to segment similar languages made up of words of varying lengths. Thus, our study confirms that known words can help to segment speech streams, extending Bortfeld et al. (2005)'s findings to a different language and a different experimental paradigm. The present experimental design allows us to establish that known

[2]Given our interest in exploring the interaction between TP-based and known word segmentation, infants had to segment five different new words in Experiment 3 compared to four new words in Experiment 1 and four new words plus one familiar word in Experiment 2. However, this minimal difference in number of new words is unlikely to account for the difference in performance between the uniform versus mixed word length conditions in particular given that Johnson and Tyler (2010) found the same pattern of results with equal number of words in both conditions.

words facilitate segmentation even when the familiar words do not systematically precede the target words.

How did the presence of known words affect segmentation mechanisms? In the following, we discuss three different alternatives. The first one is that the familiar words did not directly help segmentation by marking some word boundaries in the familiarization streams, but only increased infants' attention to the streams containing the familiar word /mamã/, with the effect of boosting their use of TPs in that condition. Unfortunately, this possibility cannot be evaluated with our data since orientation times during familiarization were not monitored. However, if it seems highly likely that the familiar word /mamã/ has captured infants' attention, it also seems reasonable to think that this recognition process would have led infants to isolate it from the preceding and following words.

A second alternative is that infants in Experiment 2 did not use TPs but rather relied only on top-down cues (the recognition and segmentation of the occurrences of the familiar word). Indeed, since all words in the stream were preceded 1/5 of the time by the familiar word /mamã/ (which was not the case for the part-words), it is possible that the segmentation of the familiar word would have given enough word-initial and word-final boundaries in the signal to allow infants to determine the words constituting the streams without recourse to TP information.

At this point, it is not possible to fully discard the above interpretations on the basis of our experimental results. However, both hypotheses make the assumption that infants in Experiment 2 only used one type of information (TPs according to the first interpretation, familiar words according to the second one). However, previous studies suggest that 8-month-old infants are already able to combine the use of different cues (see below). Accordingly, since we have found that French-learning 8-month-old infants can use both TPs (Experiment 1) and familiar names (comparison of Experiments 2 and 3) to segment continuous streams, we offer a third, more parsimonious account of early speech segmentation according to which infants in Experiment 2 would have used TPs and familiar words in combination to segment the streams. More specifically, we propose that the recognition of the known word could have firstly allowed an initial segmentation of the fluent speech into shorter sequences of speech, a chunking mechanism similar to the one implemented in the IncDROP model (Brent & Cartwright, 1996). Second, these shorter sequences would have been further segmented into words using TPs. From this perspective, our study offers the first piece of evidence that infants are able to combine bottom-up (TPs) and top-down (familiar words) segmentation mechanisms. The present results go beyond previous demonstrations that infants can combine bottom-up cues to segment speech (distributional and allophonic cues: Jusczyk, Hohne, & Bauman, 1999; distributional and rhythmic cues: Curtin et al., 2005; Johnson & Jusczyk, 2001) or use the presence of a few isolated occurrences of unfamiliar target words to boost TP-based segmentation (Lew-Williams et al., in press) by establishing that they can also combine cues of different types (bottom-up versus top-down). To strengthen the present interpretation, future studies should vary the way the familiar words appear in the familiarization streams (for example by inserting the familiar word /mamã/ only before the frequent words, thus keeping words and part-words in a symmetrical design with respect to the top-down cue position), or analyze infants' attention during the presentation of the familiarization strings when varying the number of occurrences presented.

As a final point, it is worth noting that while the presence of the familiar word /mamã/ constituted a helpful segmentation cue, this was not the case for the pseudo-word /mãma/. This suggests that 8-month-olds are not processing those two sequences similarly and that the

inversion of the two vowels between /mamã/ and /mãma/ blocks the recognition of the familiar word /mamã/. This result complements previous studies showing that by 6–8 months of age, word form representations extracted by segmentation processes are specified at the consonantal level (Bortfeld et al., 2005; Jusczyk & Aslin, 1995) and suggests that the specification extends to the vocalic level as well. This result, especially if confirmed beyond the phonetic feature of vowel nasality that distinguishes /a/ from /ã/, and whose processing has not been studied in infancy), suggests that infants' better processing of consonantal over vocalic information, which has been found in word-learning tasks in the second year of life (Havy & Nazzi, 2009; Nazzi & Bertoncini, 2009; Nazzi & New, 2007; Nazzi, 2005) and is supposed to reflect a bias in processing consonantal versus vocalic information at the lexical level (Nespor, Mehler, & Peña 2003), might not operate at the word segmentation level at 8 months of age. This issue should be explored in future research.

REFERENCES

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Bonatti, L. 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*(6), 451–459.

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.

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.

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, *96*(3), 233–262.

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 noncommercial purposes. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1393–1396). Wilmington, DE: Applied Science and Engineering Laboratories.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247–299.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

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

Gout, A. (2001). Etapes précoces de l'acquisition du lexique. (Unpublished doctoral dissertation). Ecole des Hautes Etudes en Sciences Sociales, Paris.

Goyet, L., de Schonen, S., & Nazzi, T. (2010). Syllables in word segmentation by French-learning infants: An ERP study. *Brain Research*, *1332*, 75–89.

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.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–64.

Havy, M., & Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy*, *14*(4), 439–456.

Hohne, E. A., & Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception and Psychophysics*, *56*(6), 613–623.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, *5*, 69–95.

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. & Seidl, A. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*, 131–141.
Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345.
Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23.
Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*(8), 1465–1476.
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.
Kooijman, V. M., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, *24*(1), 109–116.
Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds'. *Infancy*, *14*(6), 591–612.
Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (in press). Isolated words enhance statistical language learning in infancy.
Mattys, S., & Jusczyk, P.W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*, 91–121.
Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*(1), 13–30.
Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: New evidence from consonants in coda positions. *Language and Speech*, *52*, 463–480.
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*(3), 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*(3), 283–299.
Nazzi, T., Kemler-Nelson., D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, *1*, 123–147.
Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical acquisition. *Cognitive Development*, *22*, 271–279.
Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *ii*, 221–247.
Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247.
Pelucchi, B., Hay, J. F. & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685.
Polka, L., & Sundara, M. (in press). Word segmentation in monolingual infants acquiring Canadian-English and Canadian-French: Native language, cross-language and cross-dialect comparisons. *Infancy*.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.
Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621.
Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132.
Thiessen, E. D., Hill, E., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*, 53–71.
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.
Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172–175.
Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Science*, *8*(10), 451–456.