Psychological Research (2017) 81:990–1003
DOI 10.1007/s00426-016-0793-y

ORIGINAL ARTICLE

Familiar units prevail over statistical cues in word segmentation

Bénédicte Poulin-Charronnat[1] · Pierre Perruchet[1] · Barbara Tillmann[2] · Ronald Peereman[3]

Received: 13 October 2015 / Accepted: 10 August 2016 / Published online: 31 August 2016

Abstract In language acquisition research, the prevailing position is that listeners exploit statistical cues, in particular transitional probabilities between syllables, to discover words of a language. However, other cues are also involved in word discovery. Assessing the weight learners give to these different cues leads to a better understanding of the processes underlying speech segmentation. The present study evaluated whether adult learners preferentially used known units or statistical cues for segmenting continuous speech. Before the exposure phase, participants were familiarized with part-words of a three-word artificial language. This design allowed the dissociation of the influence of statistical cues and familiar units, with statistical cues favoring word segmentation and familiar units favoring (nonoptimal) part-word segmentation. In Experiment 1, performance in a two-alternative forced choice (2AFC) task between words and part-words revealed part-word segmentation (even though part-words were less cohesive in terms of transitional probabilities and less frequent than words). By contrast, an unfamiliarized group exhibited word segmentation, as usually observed in standard conditions. Experiment 2 used a syllable-detection task to remove the likely contamination of performance by memory and strategy effects in the 2AFC task. Overall, the results suggest that familiar units overrode statistical cues, ultimately questioning the need for computation mechanisms of transitional probabilities (TPs) in natural language speech segmentation.

✉ Bénédicte Poulin-Charronnat
benedicte.poulin@u-bourgogne.fr

[1] Université Bourgogne Franche-Comté, LEAD-CNRS UMR5022, F-21000, Dijon, France

[2] CNRS UMR5292, INSERM U1028, Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, Université of Lyon I, Lyon, France

[3] Univ. Grenoble Alpes, CNRS UMR5105, LPNC, F-38000, Grenoble, France

Introduction

A decisive stage in language acquisition is the discovery of words from continuous speech streams. The major difficulty to deal with speech segmentation is that, contrary to written language, there are no reliable cues for boundaries between words. However, several sources of information present in speech can help word discovery (see Gómez, 2007; Johnson, 2012; Jusczyk, 1999, for reviews). From the end of their first year, children become sensitive to the properties of their native language, and prosodic, allophonic, phonotactic, and distributional constraints belonging to their own language can help them to segment words from speech streams.

Cues in word segmentation

A relevant cue for word segmentation is the role of stressed syllables in stress languages. In English, for instance, there are strong and weak syllables. Strong syllables are pronounced slightly louder, longer, and higher in pitch than weak syllables. The number of words beginning with a strong syllable is three times larger than the number of words beginning with a weak syllable. Strong syllables could thus serve as a probabilistic cue for word beginning. Cutler and Norris (1988) embedded real words (e.g., *mint*) in two types of nonsense strings (e.g., *mintayve* versus *mintesh*). The task of the participants was to detect the word *mint*. In one case, the second syllable was a strong syllable (e.g., *mintayve*). If strong syllables are used as cues for word segmentation, *tayve* should be perceived as the beginning of a word and the nonsense string *mintayve* should be segmented as *min* + *tayve*, making difficult the detection of the word *mint*. In the other case, the second syllable was a weak syllable (e.g., *mintesh*), which should not result in segmentation and should not hinder the perception of the word *mint*. The results showed that the word *mint* was indeed easier to detect in *mintesh* than in *mintayve*, suggesting that strongly stressed syllables triggered word segmentation in adults. This finding was extended to infants by Jusczyk, Houston, and Newsome (1999b), who showed that 7.5-month-old English infants were able to detect strong/weak (but not weak/strong) target words from sentential contexts.

Other relevant cues for word segmentation are allophonic cues, which correspond to variants of the same phoneme. Jusczyk, Hohne, and Bauman (1999a) used pairs of items, like *nitrates* and *night rates*. In *nitrates* and *night rates*, the first *t* and the *r* have different acoustical properties. Once familiarized with one of the two items, 10.5-month-old infants were capable to detect this item in sentential contexts based on the difference in allophonic information. Other cues posited as pertinent to word segmentation are phonotactic cues, which refer to the frequency with which phonemes occur next to each other. In 2001, Mattys and Jusczyk (2001) embedded a target word (i.e., *gaffe*) in sentential contexts with good or poor phonotactic cues. A sentential context with good phonotactic cues was for instance *gean gaffe hold*, in which the averaged frequency of occurrence of [ŋg] and [fh] in natural language was 21.0 between words and 0 within words, leading to segment between *gean* and *gaffe* and between *gaffe* and *hold*. A sentential context with poor phonotactic cues was *fang gaffe tine*, in which the averaged frequency in natural language of Mg and *tu* was 0.3 between words and 32.0 within words, leading to segment within the target word *gaffe*. Infants listened longer to the target word when it was previously embedded in a sentential context with good phonotactic cues rather than poor phonotactic cues, demonstrating that 9-month-old infants were able to use phonotactic cues (see McQueen 1998, for similar results with adults).

Using an artificial language, Saffran and colleagues (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b) investigated the role of syllable distributional cues in word segmentation. Saffran et al. (1996b) used a synthesized artificial speech stream of six trisyllabic nonsense words (e.g., *babupu*, *bupada*, *dutaba*, *patubi*, *pidabu*, and *tutibu*). These six nonsense words were presented continuously without immediate repetition, for instance *babupubupadadutabapatubibabupututibu*… The only available cue for word segmentation was the transitional probabilities (TPs) linking syllables. Given the pair of syllables XY, TP refers to the probability of Y given X. In this artificial speech stream, TPs within words (ranging from 0.31 to 1) were higher than TPs between words (ranging from 0.1 to 0.2). After listening to 21 min of such an artificial speech stream, adult learners successfully distinguished words from nonwords (a test item with no syllable pairs appearing in the speech stream) and also from nonwords sharing one syllable pair with a word. Saffran et al. (1996a) extended these findings to 8-month-old infants and proposed that from an early age on, learners compute TPs and perceive word boundaries when TPs are low.

Other potential sources of information that help word segmentation are the words already known by the learners. In 1999, Dahan and Brent proposed a lexically driven model of speech segmentation, the INCDROP model, for INCremental Distributional Regularity OPtimization, in which "segmentation and word discovery during native-language acquisition may be driven by recognition of familiar units from the start, with no need for transient bootstrapping mechanisms" (p. 165). They illustrated their claim with the following example: "if *look* is recognized as a familiar unit in the utterance *Lookhere!* then *look* will tend to be segmented out and the remaining contiguous stretch, *here*, will be inferred as a new unit" (p. 165). Bortfeld, Morgan, Golinkoff, and Rathbun (2005) provided an experimental confirmation of this hypothesis in infants. The infants, as young as 6 months, successfully segmented a novel word when it followed the own name of the infant or the moniker used for his/her mother. However, they failed to do it when this novel word followed an unfamiliar word, showing clear evidence for the role of known words in the discovery of new words.

Several cues (e.g., prosodic, allophonic, phonotactic, syllable distribution, known words) are present in natural language. Importantly, each of these sources of information is probabilistic and not deterministic, and this raises the question of how these cues combine to allow for word segmentation.

Converging word-segmentation cues

Studies using artificial stimuli demonstrated that word segmentation improved when several converging cues were present in the speech flow. Mattys, Jusczyk, Luce, and Morgan (1999) showed that 9-month-old infants preferred listening to bisyllabic nonsense words including converging phonotactic and prosodic cues (rather than conflicting ones). In their Experiment 3, Saffran et al. (1996b) added a consistent prosodic cue (i.e., final vowel lengthening) to distributional cues, and observed segmentation improvement. Combining statistical cues and known words also helps word segmentation. In Cunillera, Càmara, Laine, and

Rodríguez-Fornells (2010), participants were presented with an artificial language composed of eight nonsense words. One group (the anchor group) learned two of the eight words before being familiarized with the artificial speech stream, while a nonanchor group learned two non-words. The performance on a two-alternative forced choice (2AFC) task indicated better word segmentation for the anchor group than for the nonanchor group (even when the number of the to-be-learned words was equated between the two groups). More recently, Lew-Williams, Pelucchi, and Saffran (2011) extended this finding to natural language, and demonstrated that isolated words enhanced statistical learning in 8–10-month-old infants. The infants listened significantly longer to high TP words than to low-TP words only when the speech was composed of both fluent speech and single-word utterances, but not when the speech was constituted of fluent speech only.

Therefore, there is a general consensus that combining different coherent cues improves word segmentation (Cairns, Shillcock, Chater, & Levy 1997; Christiansen, Allen, & Seidenberg, 1998; Dahan & Brent, 1999; Jusczyk 1999; Norris, McQueen, Cutler, & Butterfield, 1997). However, although segmentation cues are most often consistent in natural language, there are interesting exceptions where different cues collide. The use of artificial languages is well suited to investigate such exceptions to highlight the relative weight of different segmentation cues, which is at issue in the present study.

Conflicting word-segmentation cues

Mattys et al. (1999) presented 9-month-old infants with nonsense words in which phonotactic and prosodic cues were in conflict. They showed that prosodic cues overrode phonotactic cues. In a subsequent study, Johnson and Jusczyk (2001) observed that at 8 months of age, infants relied more on prosodic and coarticulation cues than on statistical cues. Similar results were obtained by Johnson and Seidl (2009) by opposing prosodic cues and statistical cues in 11-month-old infants. However, a developmental change exists and infants at different ages favor different word-segmentation cues. Thiessen and Saffran (2003) observed that, when prosodic cues and statistical cues were conflicting, 9-month-old infants preferentially used prosodic cues as cues to segmentation, while 7-month-old infants preferentially used statistical cues as cues to segmentation. Developmental changes have also been demonstrated with natural language. In the study by Jusczyk et al. (1999b), 7.5-month-old infants detected target words they were previously familiarized with only when the targets were strong/weak syllable words (as were the majority of English words), and not weak/strong syllable words. In contrast, 10.5-month-old infants detected both types of familiar target words. This finding suggests that before 10.5 months, infants used prosodic cues but afterwards they relied more on familiar words as cues for segmentation. Furthermore, at 10.5 months, even when a weak/strong syllable word was always followed by the same syllable (resulting in high TP between the strong syllable of the target word and the following syllable), the infants detected the weak/strong familiar target word. This finding suggests that familiar words were favored over prosodic and statistical cues.

More recently, Perruchet, Poulin-Charronnat, Tillmann, and Peereman (2014), using an artificial language, investigated the relative influence of familiarized words and statistical cues on speech segmentation. Given that the present study is in the continuation of Perruchet et al.'s study, a detailed description of the material is warranted. The artificial language consisted of three nonsense trisyllabic words (*ABC*, *DEF*, and *GHI*, in which each letter stands for a syllable). An interesting property of a language comprising only three artificial words is that it can be alternatively segmented into six trisyllabic part-words (e.g., *CDE*, *CGH*, *FAB*, *FGH*, *IAB*, *IDE*). Although exhaustive and consistent, such an alternative part-word segmentation (PW-segmentation) does not result in the smallest number of constituent units, and hence is not optimal as defined in the Minimum Description Length framework (e.g., Robinet, Lemaire, & Gordon, 2011). However, learning six units remains manageable within one experimental session (as shown by Saffran et al., 1996b).

Before listening to the speech stream, adult learners were first exposed to isolated bisyllabic words that reoccurred at word-external transitions in the following speech stream (e.g., *CD*, which occurred in the speech stream when *ABC* was followed by *DEF*). Relying on the familiar bisyllabic words should lead to PW-segmentation, whereas favoring statistical cues should result in word segmentation (W-segmentation). In the test phase, performance on a 2AFC task, in which participants had to choose between trisyllabic words and part-words, was significantly below-chance level (indicating that part-words were preferentially chosen). This finding suggested that the influence of the familiar bisyllabic words was stronger than the influence of statistics on speech segmentation. Indeed, participants preferred part-words to words, even though part-words were less statistically cohesive than words in the speech stream. The authors proposed that the probability of creating a new unit depends on the units already present in the lexicon, whether they are relevant or not. Whereas a TP-based approach predicted word preference, PARSER, a chunk-based model (Perruchet & Vinter, 1998), correctly predicted that participants perceived the speech stream as a sequence of part-words.

The mentioned studies show that when cues collide in artificial languages, some cues overtake others, and this predominance may lead to better understand the mechanisms involved in word segmentation in natural language acquisition.

The present study

The present study first aimed at replicating the influence of familiar units observed in Perruchet et al. (2014). This previous study reported only one experiment and although the observed effect was significant, its magnitude remained small. More importantly, the present study was designed to generalize the Perruchet et al.'s results along several dimensions, by introducing three major changes. Perruchet et al. presented bisyllabic words in the initial-familiarization phase, whereas the words composing the language were trisyllabic words. This discrepancy in unit length between the initial-familiarization phase and the exposure phase could have interfered with the way participants segmented the speech stream by misleading them to search for bisyllabic words. In the present study, trisyllabic units were presented to the participants during the initial-familiarization phase. A second point is that in Perruchet et al., participants had to learn three or six new units according to whether they segmented the speech stream in words or part-words, respectively. This should have given an advantage for the W-segmentation. The design of the present study ensures that whatever the perceived segmentation (words or part-words), participants had the same number of units to learn. Finally, Perruchet et al. did not have a control condition without the initial-familiarization phase to ensure that participants preferred words to part-words after exposure to a three-word language, as typically observed with languages that involve more than three words.[1] In the first experiment of the present study, this control condition has been added.

The artificial language of Perruchet et al. (2014), composed of three trisyllabic nonsense words (*ABC*, *DEF*, and *GHI*), was used. The crucial manipulation of the present study is to initially familiarize a group of participants with three part-words (*CDE*, *FGH*, and *IAB*). Assuming that segmentation is exhaustive, when three part-words are already familiar and guide segmentation, only three other part-words remain to be learned (*CGH*, *FAB*, and *IDE*). These remaining part-words were compared to words in test phase.

The artificial speech stream was created by randomly playing the three nonsense words in succession without immediate repetition. This resulted in higher word-internal TPs (1.00) compared with word-external TPs (0.50). Conversely, regarding the six part-words, TPs within part-words (0.75; the mean of 0.50 and 1.00) were lower than TPs between part-words (1.00). In addition, it is important to note that in the continuous speech stream, words were more frequent than part-words. This design allows the dissociation of the influence of statistical cues and familiar units in word segmentation, with statistical cues favoring W-segmentation and familiar units favoring PW-segmentation.

Experiment 1

Method

Participants

Fifty-three psychology undergraduate students from the Université de Bourgogne took part in the experiment in partial fulfillment of a course requirement. They were randomly assigned to an unfamiliarized group ($N = 17$), a nonword-familiarized group ($N = 18$), and a part-word familiarized group ($N = 18$).

Material

The artificial speech stream was composed of three trisyllabic nonsense words, *ABC*, *DEF*, and *GHI* (each letter representing one syllable). In an initial-familiarization phase, three part-words or three nonwords were used, according to the group (part-word familiarized versus nonword familiarized, respectively). The three part-words were created from the last syllable of a word with the first two syllables of another word (*CDE*, *IAB*, and *FGH*). The three nonwords (*FBG*, *CHD*, *IEA*) recombined the syllables of the three words composing the speech stream in such a way that the pairs of syllables constituting the nonwords never occurred in the speech stream (Table 1).

For the exposure phase, the three words composing the language were randomly concatenated without immediate repetition. To avoid acoustical biases, the syllables *pa*, *pɔ̃*, *gy*, *bi*, *do*, *te*, *dɑ̃*, *kɛ̃*, and *tu* were randomly assigned to the three syllabic positions of the three words for pairs of yoked participants in the three groups (Table 1; procedure from Perruchet et al., 2014). The word-internal TPs were 1.00 and the word-external TPs were 0.50. Each word was

[1] While decreasing the number of words to learn seems to make learning easier (think for instance of a list of to-be-memorized items or of the decreased number of to-be learned words used for infants in the work of Saffran et al., 1996a compared to adults), in fact, decreasing the number of words composing the artificial language reduces the differences between word-internal and word-external TPs. With three words, the TPs within words were 1 and the TPs between words were .5 (compared respectively to 1. and .33 for a four-word language), and the difference in frequency of occurrence decreases between words and part-words.

Table 1 Material used in Experiments 1 and 2

Initial-familiarization phase	Exposure phase	Test phase (2AFC)		Test phase (syllable detection)	
Part-word familiarized group		Words	Part-words	Words	Part-words
FGH (papɔ̃gy)		*ABC (kɛ̃tubi)*	*FAB (pakɛ̃tu)*	*C (bi)*	*H (gy)*
CDE (bidote)		*ABC (kɛ̃tubi)*	*CGH (bipɔ̃gy)*	*F (pa)*	*E (te)*
IAB (dãkɛ̃tu)		*ABC (kɛ̃tubi)*	*IDE (dãdote)*	*I (dã)*	*B (tu)*
	ABC (kɛ̃tubi)	*DEF (dotepa)*	*FAB (pakɛ̃tu)*		
	DEF (dotepa)	*DEF (dotepa)*	*CGH (bipɔ̃gy)*		
Nonword-familiarized group	*GHI (pɔ̃gydã)*	*DEF (dotepa)*	*IDE (dãdote)*		
FBG (patupɔ̃)		*GHI (pɔ̃gydã)*	*FAB (pakɛ̃tu)*		
CHD (bigydo)		*GHI (pɔ̃gydã)*	*CGH (bipɔ̃gy)*		
IEA (dãtekɛ̃)		*GHI (pɔ̃gydã)*	*IDE (dãdote)*		

The language used in the exposure phase can be segmented into six part-words composed of the last syllable of a word and the first two syllables of another word (*FGH*, *CDE*, *IAB*, *FAB*, *CGH*, *IDE*). Three of them were used in the initial-familiarization phase for the part-word familiarized group, while nonwords were used for the nonword-familiarized group. The items in brackets are phonetic examples

repeated 180 times, leading to speech stream duration of approximately 6 min. To avoid word boundary cues due to the start and end of the stream, the resulting WAV file was modified with Peak Pro 5.2 by applying progressive fade in and fade out to the first and last 5 s. In addition, the speech stream could begin and end with the first, the second, or the third syllable of a word.

For the 2AFC task used in the test phase, the three part-words (also composed of the last syllable of a word and the first two syllables of another word) that were not played during the initial-familiarization phase were created (*CGH*, *FAB*, and *IDE*). Using three part-words in the test phase allowed every word and every part-word to be equally repeated during the test.

All materials were synthesized using the MBROLA (Multiband Resynthesis Overlap Add) speech synthesizer (http://tcts.fpms.ac.be/synthesis/; Dutoit, Pagel, Pierret, Bataille, & Van Der Vrecken, 1996) with the fr2 diphone database. The duration of each syllable was 232 ms.

Procedure

During the initial-familiarization phase, the participants were presented with three isolated part-words in the part-word familiarized condition, and with three isolated non-words in the nonword-familiarized condition. On each trial, the three part-words/nonwords were played in succession in random order separated by a 2-s interval. After each group of three items, participants heard a syllable, and they were asked to indicate, by pressing on the appropriate key, which one of the three items comprised the target syllable (the 1st, the 2nd, or the 3rd). This short-term memory task was implemented to ensure that participants paid attention to the to-be-familiarized items. There were 15 trials and, therefore, each of the three part-words/nonwords was repeated 15 times. The unfamiliarized group did not perform the initial-familiarization phase.

In the exposure phase, the participants of the three groups were required to listen to an imaginary language as if they were listening to music, with the explicit instruction to not analyze the speech stream.

Finally, in the following test phase, participants of the three groups performed a 2AFC task, in which they were presented with pairs of items, and had to judge, for each pair, which item seemed more like a word of the imaginary language they were exposed to before. Each word (*ABC*, *DEF*, and *GHI*) was associated with each of the three remaining part-words (novel part-words; *CGH*, *FAB*, and *IDE*). This resulted in nine different pairs of items, and each pair was repeated twice. The members of each pair were separated by a 3-s interval. The order of the items within a pair, as well as the order of the pairs in the test sequence, was randomized with a different randomization for triplets of yoked participants (each from one group). When a pair was repeated, the order of the two items was systematically reversed.

Results and discussion

The percentages of correct responses (as defined by correctly choosing the word from the two alternatives) are displayed in Fig. 1. A one-way ANOVA with Group (part-word familiarized, nonword familiarized, unfamiliarized) as a between-subjects variable revealed a significant effect, $F(1, 50) = 11.45$, $p < 0.001$, $\eta_p^2 = 0.31$.

The performance of the unfamiliarized group ($M = 63.07$, $SD = 19.44$) was significantly higher than the performance of the part-word familiarized group ($M = 33.95$, $SD = 27.87$), $t(33) = 3.565$, $p < 0.001$, $d = 1.24$ and significantly above chance, $t(16) = 2.78$,

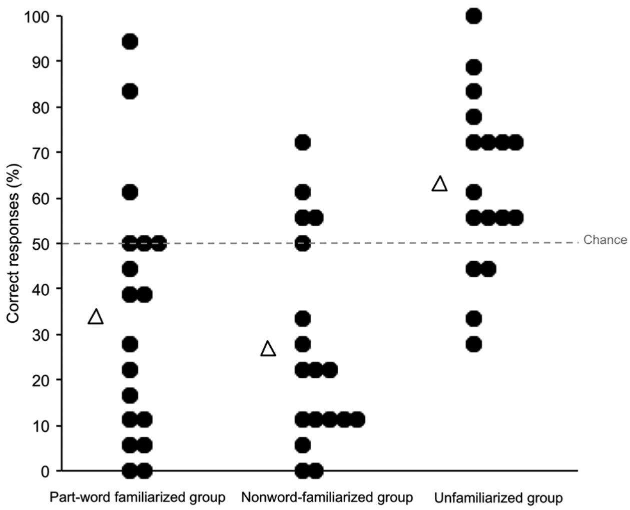

Fig. 1 Accuracy in Experiment 1 presented as a function of group (part-word familiarized, nonword familiarized, and unfamiliarized) with individual data points presented as *black circles* and mean performances as *white triangles*. Note that chance level is at 50 % and that correct responses were defined as the selection of the word from the 2AFC test pair

$p = 0.014$, $d = 0.67$. Without the initial-familiarization phase, the learners preferred W-segmentation to PW-segmentation, showing that they were able to segment a three-word language as well as previously used languages with more than three words (e.g., Saffran et al., 1996b). The performance of the part-word familiarized group ($M = 33.95$, $SD = 27.87$) was significantly below-chance, $t(17) = 2.443$, $p = 0.026$, $d = 0.58$, suggesting that learners gave greater weight to familiar items (part-words) than to statistical cues when segmenting a continuous speech stream.

However, the performance of the unfamiliarized group was also significantly higher than the performance of the nonword-familiarized group ($M = 26.85$, $SD = 22.57$), $t(33) = 5.074$, $p = 0.001$, $d = 1.77$, who obtained significantly below-chance score, $t(17) = 4.352$, $p < 0.001$, $d = 1.03$, which did not differ from the score of the part-word familiarized group, $t(34) < 1$, $p = 0.41$.[2]

It is essential to understand why the nonword-familiarized group got below-chance performance, as did the part-word familiarized group. Indeed, by construction, the nonword-familiarized group could not rely on familiarized units to segment the speech flow and, hence, below-chance performance in this group must be due to another, yet unidentified cause. A challenge arises from the possibility that this other cause also explains the below-chance performance observed in the part-word familiarized group. If this were the case, then our prior interpretation for this latter group, which relies on the role of familiarized units, would become unwarranted. A careful examination of the materials shown in Table 1 indicates that, for the two groups, the part-word test items have the same initial syllables (*F*, *C*, or *I*) as the items used in the initial-familiarization phase. A possibility is that after learning the words played during the familiarization phase, participants would tend to transfer some acquired knowledge about these words onto the speech stream played during the subsequent exposure phase. If participants used the first syllable of the words played during the initial phase as a marker of word onset for the subsequent speech stream, then, by construction, this would mislead participants to segment this speech stream into part-words. The 2AFC test would reflect this mode of segmentation. If this interpretation is correct, then statistical cues would be still overcome by other cues during segmentation, but these other cues would be unrelated to the chunking processes based on familiar units.

However, there is a second possibility. Up to now, we have assumed that the effect of the initial-familiarization phase on the performance during the test is mediated by the influence of this phase on the segmentation process during the exposure phase. In other words, performance during the test phase is assumed to be entirely dependent on the way the language played during the exposure phase has been segmented. It cannot be excluded, however, that the initial-familiarization phase exerts a direct influence on the 2AFC test. In this perspective, participants presented with test-item pairs could simply prefer the item that began by the same syllable as the items that were played in isolation during the initial phase, independently of what happens during the intermediate exposure phase. This would leave open the possibility that participants familiarized with the part-words segment the speech stream into part-words (due to the influence of existing chunks), while participants familiarized with the nonwords segment the speech stream into words (due to the statistical structure of the language).

Testing these hypotheses is not a trivial matter, because it turns out to be impossible to avoid the bias raised above, namely the fact that the part-words used in the 2AFC test began with the same syllables as the items used in the initial-familiarization phase (any other arrangement would even lead to two shared syllables). It thus appears that a 2AFC test or any other post-experimental test comparing words and part-words (such as a yes/no recognition test) are not appropriate, because these tests are potentially

[2] For completion, we performed Linear Mixed Model (LMM) on the data with participants and items as random effects and Group as fixed effect. The LMM showed a significant effect of Group, $F(2, 474) = 40.21$, $p < .001$, a significant difference between the unfamiliarized group and both the part-word familiarized group, $F(1, 313) = 44.71$, $p < .001$, and the non-word familiarized group, $F(1, 313) = 77.01$, $p < .001$, while the difference between the part-word and nonword familiarized groups failed to reach the conventional significance threshold, $F(1, 322) = 2,80$, $p = .095$. These results thus lead to the same conclusions as the results obtained with the ANOVA.

sensitive to the direct influence of the initial-familiarization phase on test performance, with the possible involvement of memory strategies, but without informing us about speech segmentation. In Experiment 2, an online measure of segmentation, which prevented from the influence of memory strategies, was used (see Franco, Eberlen, Destrebecqz, Cleeremans, & Bertels, 2015a).

The rationale of the online syllable-detection task is that if the learners have segmented units from the speech stream, they should process the last syllable of the segmented units faster (thanks to in-word syllable expectations). Franco et al. (2015a) observed that after an exposure phase to an artificial language, the last syllable of the trisyllabic nonsense words composing the language was detected faster than their first and second syllables (for similar result in visual statistical learning, see Bertels, Franco, & Destrebecqz, 2012; Kim, Seitz, Feenstra, & Shams, 2009; Turk-Browne, Jungé, & Scholl, 2005).

In the online syllable-detection task, contrary to the 2AFC task, no isolated test items were presented and there was no reference to the words composing the artificial language. Instead, small blocks of continuous speech stream were presented, and participants had to detect online a given syllable. In Experiment 2, the syllables to detect were *C*, *F*, and *I*, which were the last syllables of the words, and *E*, *H*, and *B*, which were the last syllables of the part-words (see Table 1). Importantly, *E*, *H*, and *B* were also the second syllables of the words. As a consequence, if participants segment the speech stream into words, they should detect *C*, *F*, and *I* more quickly than *E*, *H*, and *B* (as observed by Franco et al., 2015a). Likewise, *C*, *F*, and *I* were also the first syllables of the part-words. If participants segmented the speech stream into part-words, they should detect *E*, *H*, and *B* more quickly than *C*, *F*, and *I*.

In keeping with the rationale underlying Experiment 1, statistical cues should still favor W-segmentation and familiar units should still favor PW-segmentation. Our main hypothesis is that participants initially familiarized with part-words, who are given both familiar units and statistical information, should be more sensitive to familiar units than to statistical information and, hence, should segment the speech stream into part-words (i.e., faster detection for the syllables *E*, *H,* and B, which ended the part-words than for the syllables *C*, *F*, and *I*, which started the part-words). By contrast, participants initially familiarized with nonwords may rely only on statistical information and, hence, should segment the speech stream into words (i.e., faster detection for the syllables *C*, *F*, and *I*, which ended the words than for the syllables *E*, *H*, and B). In all cases, the TP of the syllable to detect with reference to the preceding syllable was always 1.00.[3]

Finally, the online syllable-detection task also allows for the comparison between the processing of the familiar part-words, learned under very favorable circumstances (seen repeatedly and in an isolated way), and the processing of the remaining part-words (referred below as *unfamiliar* part-words) potentially segmented during the exposure phase. It would be possible to make such a comparison with a 2AFC task, but memory effects would very likely lead to favor the familiar part-words. Assuming that participants segmented the speech stream into part-words, the familiar part-words should show a processing advantage with faster response times than the remaining part-words.

Experiment 2

Method

Participants

Thirty psychology undergraduate students from the Université de Bourgogne took part in the experiment for course credit. They were randomly assigned to either a part-word familiarized group or a nonword-familiarized group, with 15 participants per group.

Material

For the familiarization phase and the exposure phase, the material was the same as in Experiment 1. For the syllable-detection task (test phase), nine new excerpts of the same artificial language were created, in which each word occurred 26 times (approximately 54 s). To avoid any cue on word boundaries, the excerpts began either by the first syllable, the second syllable or the third syllable of one of the three words. The nine excerpts were synthesized and played as described in Experiment 1 for the speech stream of the exposure phase.

Procedure

The familiarization phase and the exposure phase were identical to those in Experiment 1.

In the final syllable-detection test phase, nine new blocks of short excerpts of an artificial language composed of the same three words as used in the exposure phase were presented. Before each block, a syllable was played in isolation. Participants had then to detect this syllable in the excerpt by pressing a key on a numeric pad. The six blocks,

[3] Taking the last two syllables of a word and the first syllable of another word would be another possible segmentation requiring six part-words. However, in that case, the TP of the last syllable of the part-words would be .50 and not 1.00 (as for the words). For the sake of equality, only the part-words composed of the last syllable of a word and the first two syllable of another word were used.

in which the syllables to detect were *C*, *F*, *I*, *E*, *H*, or *B*, were presented in random order. The six experimental blocks were preceded by three training blocks, in which the to-be-detected syllables were the remaining syllables (A, D, and G). In each block, there were 26 occurrences of the syllable to detect.

Results and discussion

Responses included in the analyses reported below were in a time range from 100 ms before the onset of a target syllable (for anticipatory responses) to 800 ms after the onset of this syllable. The mean rate of key presses falling outside the time windows (false alarms) was 8.38 % ($SD = 4.33$) for the part-word familiarized group and 8.42 % ($SD = 6.09$) for the nonword-familiarized group. The mean rates of absence of no response on a target syllable (misses) were 16.84 % ($SD = 12.56$) and 22.56 % ($SD = 9.55$), respectively. For the part-word familiarized group, there were more correct detections for the final syllables of the part-words (80.77 %, $SD = 19.43$) than the final syllables of the words (85.47 %, $SD = 11.66$). For the nonword-familiarized group, the percent of correct syllable detection was very similar for both words (77.52 %, $SD = 12.54$) and part-words (77.26 %, $SD = 11.81$). There were no significant differences between conditions for any of the three measures (false alarms, misses, and correct detections), ps > 0.10. The task of detecting a previously heard syllable within a continuous artificial speech stream is not an easy task, and errors may stem from both perceptual confusions and memory failures for the target syllable. If for instance, a participant, for a given block, consistently responds to a wrong syllable, this would increase both misses (no response to the true target syllable) and false alarms. These errors may have introduced some noise in the results, but, importantly, they should not generate systematic biases, given that (1) the syllables were randomly assigned to the words for each participant within a group and (2) participants from the two groups were yoked in such a way that they were exposed to the very same familiarization and test languages.

Figure 2 displays the response times for syllable detection in words, and part-words, with the part-words being divided into familiar and unfamiliar, as a function of Group. The response times were examined with a 2 × 2 ANOVA with Group (part-word familiarized, nonword familiarized) as a between-subjects variable and Target (words, part-words) as a within-subjects variable. There was no effect of Group, $F(1, 28) = 0.01$, $p = 0.946$, but a significant main effect of Target, $F(1, 28) = 12.08$, $p = 0.002$, $\eta_p^2 = 0.30$, which was qualified by a significant Group × Target interaction, $F(1, 28) = 5.93$, $p = 0.022$, $\eta_p^2 = 0.17$.

Subsequent analyses taking into account the division between familiar and unfamiliar part-words were performed using *t* tests. For the part-word familiarized group, the response times were faster for the last syllables of both the familiar part-words, $t(14) = 3.81$, $p = 0.002$, $d = 1.23$, and the unfamiliar part-words, $t(14) = 3.98$, $p = 0.001$, $d = 1.28$, than for the last syllables of the words. There was no significant difference between familiar part-words and unfamiliar part-words, $t(14) = 0.137$, $p = 0.893$.[4]

No significant differences were observed for the nonword-familiarized group (familiar part-words versus words, $t(14) = 1.14$, $p = 0.274$; unfamiliar part-words versus words, $t(14) = 0.477$, $p = 0.641$; and familiar part-words versus unfamiliar part-words, $t(14) = -0.978, p = 0.345$).[5]

The syllable-detection task was introduced to remove the influence linked to the presentation of isolated items in the test phase (such as the similarity between the first syllable of the familiar and test items). Contrary to what was observed in Experiment 1, the performance of the nonword-familiarized group no longer reflected PW-segmentation. This finding suggests that in Experiment 1, the first syllable of the familiar items was probably not used as

[4] To address the possibility that the preselected range may have been too narrow, analyses were run again with a [−300, 1000 ms] range, ensuring a very broad coverage. This new range was in fact the largest possible one, because a still larger range would have generated some overlaps between the response windows surrounding two successive target syllables. The resulting changes were quite minor. Means differed only by a few milliseconds, and the *p* values of the statistical tests reported in the main text differed only on their third or fourth decimals, never affecting their interpretation in terms of significance. Unsurprisingly, the rates of false alarms and misses decreased, but remained substantial. The mean rate of false alarms was 5.34 % for the part-word familiarized group and 5.30 % for the nonword-familiarized group. The mean rates of misses were 14.10% and 19.91 %, respectively. These analyses suggest that the relatively high rates of false alarms and misses were not due to ill-fitted exclusion criteria, but to genuine detection errors.

[5] We additionally performed LMM with participants as random effect and Group and Target as fixed effects. There was no effect of Group, $F(1, 3723) = 0.295$, $p = 0.587$, but a significant main effect of Target, $F(1, 3723) = 59.37$, $p < 0.001$, which was qualified by a significant Group × Target interaction, $F(1, 3723) = 28.70$, $p < 0.001$. Subsequent analyses taking into account the division between familiar and unfamiliar part-words were performed. For the part-word familiarized group, the response times were faster for the last syllables of both the familiar part-words, $F(1, 1461) = 69.84$, $p < 0.001$, and the unfamiliar part-words, $F(1, 1425) = 64.15$, $p < 0.001$, than for the last syllables of the words. There was no significant difference between familiar part-words and unfamiliar part-words, $F(1, 998) = 0.02$, $p = 0.874$. For the nonword-familiarized group, a significant difference was observed between familiar part-words and words, $F(1, 1366) = 4.91$, $p = 0.027$, while no significant difference was observed for unfamiliar part-words versus words, $F(1, 1348) = 2.29$, $p = 0.131$, and familiar part-words versus unfamiliar part-words, $F(1, 902) = 0.40$, $p = 0.528$. These results thus lead to very similar conclusions as the results obtained with the ANOVA.

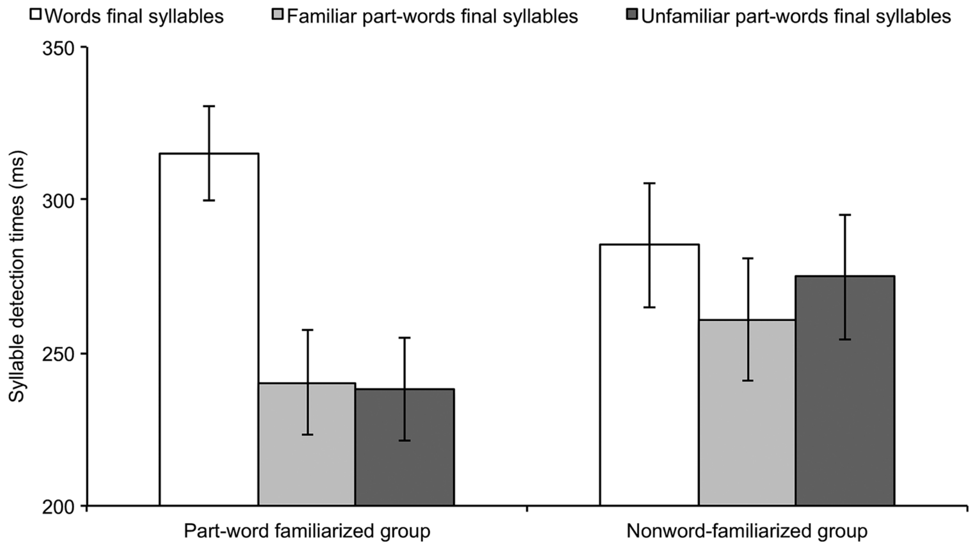

Fig. 2 Detection times presented as a function of Target (words, familiar part-words, unfamiliar part-words) and Group (part-word familiarized versus nonword familiarized). *Error bars* represent standard errors

a cue for PW-segmentation (if it were, a PW-segmentation should have been replicated). Rather, the below-chance performance in Experiment 1 was independent of what happened during the exposure phase, and only relied on the similarity between the first syllable of the familiar items and the first syllable of the test part-words. In Experiment 2, the final syllables of the part-words were detected faster than the final syllables of the words for the part-word familiarized group, but not for the nonword-familiarized group. This finding clearly demonstrated with an implicit speeded measure, that adult learners gave more weight to familiarized units than to statistical cues in word segmentation.

PARSER simulations

To account for the extraction of words from a continuous speech stream, Perruchet and Vinter (1998) proposed a chunk-based model, PARSER, which is designed around plausible psychological principles. Word segmentation emerges naturally through basic processes of memory and associative learning.

To explain PARSER's functioning, let us return to the previous example of the speech stream taken from Saffran et al. (1996b) presented in the introduction. The initial chunks are formed on a random breaking down of the speech flow and a weight is assigned to each chunk. PARSER could initially randomly segment the stream as follows:

babu pubupa dadu taba patubi babu putu tibu

Some of the chunks are relevant for the segmentation, such as *patubi*, which is a word (and as such a *good chunk*) or *babu* and *taba*, which are parts of words (and as such *good parts of chunks*) and, others, such as *pubupa* and *dadu* are irrelevant (*bad chunks*).

At the beginning of the segmentation, some few good chunks are created. When an initially randomly created chunk is again encountered, its weight is incremented, while decay and interference decrease the weight of initially randomly created chunks that did not reoccur. Words that reoccur frequently in the language are thus extracted. Then, the chunks created during learning may guide perception by becoming primitives of the system. In the example above, *babu* or *taba* will serve as new primitives to create larger chunks such as *babupu* and *dutaba*.

Over the past few years, PARSER has been used to simulate several results obtained in word segmentation studies (e.g., Frank et al., 2010; Giroux, & Rey, 2009; Perruchet, & Poulin-Charronnat, 2012; Perruchet et al., 2014; Perruchet & Tillmann, 2010), lending strong support for the chunk-based approach in accounting for speech segmentation.

To evaluate, whether chunk-based models can account for the findings of the present study, simulation was performed with PARSER, using the initial parameters of the model (Perruchet & Vinter, 1998) and entering exactly the same stimuli the participants have heard during the initial-familiarization phase and the exposure phase of Experiments 1 and 2 (see Fig. 3). PARSER partially simulated the results of the present study. For the part-word familiarized group, PARSER nicely predicts PW-segmentation, with the three words being created less frequently than the three unfamiliar part-words in the *mental lexicon* of PARSER. However, for the nonword-familiarized group, the three words were more frequently abstracted than were the three unfamiliar part-words in the *mental lexicon* of PARSER,

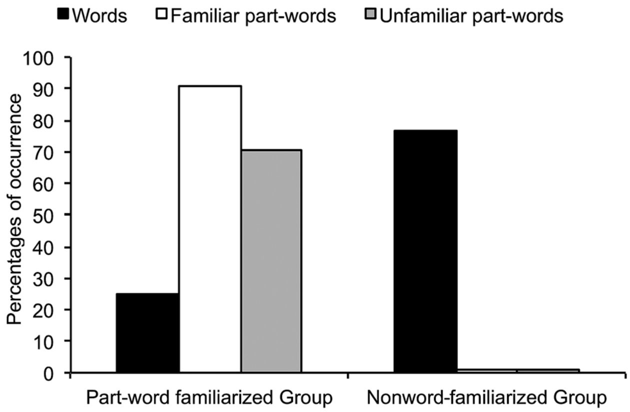

Fig. 3 PARSER's simulations

differing from the results observed in Experiment 2, in which there was no difference between words and part-words.

General discussion

The present study evaluated whether adult learners preferentially used known (familiar) units or statistical cues for segmenting continuous speech. To familiarize participants with some of the potential units (here part-words) of the artificial speech stream, an initial-familiarization phase was added before the exposure phase. If familiar units guide speech segmentation, learners should segment the speech stream into part-words. However, if learners give more weight to statistical cues than to familiar units, they should favor W-segmentation.

In Experiment 1, under standard conditions (i.e., without an initial-familiarization phase) and using a 2AFC task, the statistical cues present in the three-word artificial language were strong enough to clearly induce the usually observed W-segmentation. The introduction of an initial-familiarization phase, in which part-words were presented, modified the speech segmentation. As a result, learners preferred novel unfamiliar test part-words over words, suggesting a PW-segmentation. However, a control group learning nonwords during the initial-familiarization phase exhibited, like the part-word familiarized group, a preference for part-words. This suggested that the preference for part-words observed in Experiment 1 could be due to the direct influence of the items used in the initial-familiarization phase on the processing of test items, both of them beginning with the same syllables (that was not the case for the words). To prevent this influence, Experiment 2 used a syllable-detection task in the test phase. The results of the part-word familiarized group demonstrate that familiar units overrode statistical cues in word segmentation. These findings replicate and extend those obtained by Perruchet et al. (2014).

In addition, in Experiment 2 no difference was observed between the processing of the last syllable of familiar part-words (learned before the exposure phase) and the last syllable of unfamiliar part-words (being part of the continuous speech stream during the exposure phase). This finding has several implications: First, it indicates that the initial-familiarization phase had no direct influence on the syllable-detection task used in the test phase. If part-words used in the initial-familiarization phase had influenced the syllable-detection task, processing the last syllable of these part-words should have been speeded with regard to the unfamiliar part-words. Second, it shows that learning new units based on familiar words is fast and efficient. This is in complete agreement with the results of Perruchet and Tillmann (2010). In that study, participants were presented with an artificial language composed of six trisyllabic nonsense words. The ease with which a unit of three syllables was spontaneously perceived as a word (referred to as the initial word-likeness, IWL) was manipulated for three of these six words (based on pre-tests). For one participant group (IWL+ group), the three words, when heard in a continuous speech stream, were spontaneously perceived as words more often than trisyllabic part-words, and in the other group (IWL− group), the three words were perceived less often as words than trisyllabic part-words (see Perruchet, Tyler, Galland, & Peereman, 2004 for more details on the IWL bias). IWL+ can be construed as analog to the familiarity created by an initial-familiarization phase. The biased words were learned faster in the IWL+ group than in the IWL- group. More crucially, the unbiased words (the three remaining words of the artificial language) were also learned better by the IWL+ group than by the IWL− group, and there was no difference between the biased and unbiased words at a final test for the IWL+ group. This finding suggested that the IWL+ biased words facilitated the extraction of the three unbiased remaining words, resulting (as in the present study) in similar performance for biased and unbiased words at final test. Relying on familiar units appears to be a very fast and efficient process to learn new units.

2AFC versus syllable-detection task

The 2AFC task, which is widely used in studies on word segmentation, presents some limitations. Several caveats have already been raised concerning the use of a 2AFC task (Dahan & Brent, 1999, Footnote 5; Valian & Coulson, 1988), notably the possible role of memory strategies and the opportunity of learning during test. In Experiment 1 of the present study, one might have predicted that the nonword-familiarized control group would show either a

W-segmentation (if the familiar nonwords had no influence on segmentation) or no preference between words and part-words (if the familiar nonwords prevented segmentation). However, the participants in the nonword-familiarized group segmented the speech stream into part-words, as did the part-word familiarized group. An analysis of the material revealed that both the familiar part-words and the nonwords began with the same syllables as the test part-words. This raised the possibilities that either participants segment the speech stream into part-words using the first syllable of the familiar unit as an onset-cue or, even simpler, irrespective of what they did during speech stream exposure, they chose the part-word test items because these items shared their first syllable with the familiar units of the initial-familiarization phase (whether part-words or nonwords). The syllable-detection task of Experiment 2 supported this latter hypothesis: the detection performance of the nonword-familiarized group was not faster for part-words than for words. This finding, thus, suggests that in Experiment 1, the participants of the nonword-familiarized group did not use the first syllable of the familiar units as an onset-cue for segmenting the continuous speech flow, but rather chose the part-word test items because they have the same first syllable than the familiar units. Given that in Experiment 1, participants of the familiarized part-word group received the same information on the first syllable as the nonword-familiarized group, it is possible that they also relied on this information in the 2AFC task. It, thus, appears that the 2AFC task does not ensure that learning actually occurs and online task, such as the syllable-detection task, should be favored.

The instruction given for the detection tasks of specific events in the speech stream (Franco et al., 2015a; Gómez, Bion, & Mehler, 2011; Turk-Browne et al., 2005) makes no reference to the language heard during the exposure phase. The test items are never presented in isolation and are never compared during a forced choice task, which is crucial to minimize memory influences, explicit strategies or learning during test. In addition, the fact that detection times are quite fast makes unlikely the development of strategies. In word segmentation studies, online tasks may thus be an interesting alternative to 2AFC tasks at test. The syllable-detection task looks more promising than the click-detection task, another online task previously used (Gómez et al., 2011; Morgan & Saffran 1995). The results obtained with the click-detection tasks are inconsistent. In Gómez et al., the detection times were faster to clicks between words than to clicks within words. However, in an early study, Abrams and Bever (1969) failed to observe faster detection times for clicks occurring between clauses compared to clicks occurring before or after clauses. In line with these data, a recent study by Franco, Gaillard, Cleeremans, and Destrebecqz (2015b) also failed to observe faster detection times for clicks between words compared to within words. However, with a syllable-detection task, Franco et al. (2015a) observed faster detection times for the third element of the trisyllabic items than for the first and the second elements, revealing learning, a result very similar to that of Turk-Browne et al. (2005), who used a detection task of visual elements. In any case, in future research, the use of online tasks or online measures such as event-related brain potentials (e.g., Abla, Katahira, & Okanoya, 2008; Sanders, Newport, & Neville, 2002) should be favored to alleviate problems encountered in 2AFC tasks.

How to account for the preferred use of familiarized units over statistical cues?

The main approach to account for word segmentation of a speech stream is the computation of TPs (e.g., Saffran et al., 1996b). In the present case, a TP-based approach could not account for the learners' performance. The TPs within words (1.00) were greater than the TPs (0.50) between words. Moreover, the TPs within part-words (0.75) were smaller than the TPs between part-words (1.00). If participants based their segmentation on TP computations, the three words composing the language should have been extracted (even more so because the words were more frequent than the part-words in the speech stream). The present data were, thus, more in line with studies that minimize the role of TP computations in word segmentation (Johnson & Tyler, 2010; Perruchet & Poulin-Charronnat, 2012; Perruchet et al., 2014; Yang, 2004).

An alternative to the TP-based approach relies on chunking mechanisms (Dahan & Brent, 1999; French, Addyman, & Mareschal, 2011; Perruchet & Vinter 1998; Robinet et al., 2011). To evaluate, whether chunk-based models can account for the findings of the present study, simulations were performed with PARSER.

PARSER partially simulated the results of the present study. For the part-word familiarized group, PARSER nicely predicts PW-segmentation, with the three words being created less frequently than the three unfamiliar part-words in the *mental lexicon* of PARSER. However, for the nonword-familiarized group, the three words were more frequently abstracted than were the three unfamiliar part-words in the *mental lexicon* of PARSER, differing from the results observed in Experiment 2, in which there was no difference between words and part-words. Concerning the segmentation failure of the nonword-familiarized group in Experiment 2, several interpretations are possible. The first one is similar to the one proposed by Perruchet et al. (2014): Learning the language in the exposition phase could be made difficult by the prior presentation of items that do not belong to the language but that share the same syllables (Gebhart, Aslin, & Newport, 2009). However, this explanation appears unlikely because when the shift between two languages is clear, previous studies have shown that there was no deficit in the learning of the second language (Gebhart et al., 2009; Weiss, Gerfen, & Mitchel, 2009). A second possibility is that the participants of the nonword-familiarized group successfully segmented the speech stream into words, but the syllable-detection task failed to reveal this learning. Indeed, the participants had to detect either the second or the third syllable of the words. Though some studies demonstrated that the third element was processed significantly faster than the second element (Franco et al., 2015a; Turk-Browne et al., 2005, see also Minier, Fagot, & Rey, 2015), others failed to observe a difference between the second and third elements (Hunt & Aslin, 2001). Finally, another explanation could be linked to what is called the "Uniqueness Point" (UP) in linguistic domain. The UP is "the moment at which the acoustic–phonetic information already presented remains compatible with a single lexical entry" (Radeau & Morais, 1990, p. 1). For a PW-segmentation, the UP is located after the second syllable of the part-words, whereas for a W-segmentation, the UP is located after the first syllable of the word. In the case of very limited vocabulary as in the artificial language used in the present study, assuming that performance improves until the UP, and no longer improves once the UP is reached, could account for the results. Indeed, this could explain the facilitation observed in the part-word familiarized group for the syllables, E, H, and B (third syllable of the part-words) compared to the syllables C, F, and I (first syllable of the part-words). By contrast, for the nonword-familiarized group, detecting the syllables E, H, and B (second syllables of the words) or the syllables C, F, and I, the last syllable of the words) could lead to similar time processing. By assuming that the participants in the nonword-familiarized group segment the speech into words, an interpretation based on UP count accounts for the absence of difference between words and part-words.[6]

[6] According to the UP interpretation, the mean response times for the last syllables of the part-words should not differ between the nonword familiarized group (i.e., 2nd syllable of the segmented unit if this group segment the speech stream into words) and the part-word familiarized group (i.e., last syllable of the segmented unit if this group segment the speech stream into part-words). Contradicting this prediction, the mean RTs were numerically slower for the nonword-familiarized group than for the part-word familiarized group. However, this difference did not reach significance ($p = 0.25$).

The role of familiarized units in word segmentation revisited

Although the role of statistical cues and other cues such as those presented in the introduction is not called into question, the present study stresses the underestimated role of familiarized units in word segmentation. It appears that as soon as some units are known, they could take precedence over other segmentation cues.

One way to become familiarized with units in the natural environment is the occurrence of isolated words in infant-directed speech. Brent and Siskind (2001) found that on average, 9 % of the maternal utterances consisted of isolated words, and that 27 % are repeated in close temporal proximity (see Fernald & Morikawa, 1993, for similar results). In addition, Lew-Williams et al. (2011) estimated that "a child may hear up to 4 millions isolated word tokens by their fourth birthday", p. 1324. Previous data have already highlighted the potentially important role of familiarized units at the beginning of language acquisition. As early as 4.5 months, infants recognize their own name (Mandel, Jusczyk, & Pisoni, 1995) and at 6 months, infants exploit these highly known words (e.g., own name, monikers like Mommy's) as anchors to infer new words (Bortfeld et al., 2005). The present study, along with other data (e.g., Perruchet & Tillmann, 2010), supports the idea that familiar units facilitate the discovery of new units, and goes one step further by demonstrating that learners give more weight to familiar units than to statistical cues when these cues collide (see also Perruchet et al., 2014). Ultimately, it might, thus, be argue against the necessity to assume TP-computation mechanisms in natural language speech segmentation.

To summarize, the present study confirms a previously advanced hypothesis (although not dominant up to now) about the important role familiar units can play in initial word segmentation (Dahan & Brent, 1999). Future research should collide familiar units with other segmentation cues (prosodic, allophonic…) to assess the weight learners (adults and infants) attribute to familiar and new units in language acquisition.

Acknowledgments The authors are grateful to Pascal Morgan and Cédric Foucault for help with collecting the data.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*(6), 952–964. doi:10.1162/jocn.2008.20058.

Abrams, K., & Bever, T. G. (1969). Syntactic structure modifies attention during speech perception and recognition. *The Quarterly Journal of Experimental Psychology, 21*, 280–290. doi:10.1080/14640746908400223.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324. doi:10.1111/1467-9280.00063.

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology. Learning, Memory, and Cognition, 38*(5), 1425–1431. doi:10.1037/a0027210.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*, 298–304. doi:10.1111/j.0956-7976.2005.01531.x.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*, B33–B44. doi:10.1016/S0010-0277(01)00122-6.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: a bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33*, 111–153. doi:10.1006/cogp.1997.0649.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes, 13*, 221–268. doi:10.1080/016909698386528.

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. (2010). Words as anchors: known words facilitate statistical learning. *Experimental Psychology, 57*, 134–141. doi:10.1027/1618-3169/a000017.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113–121. doi:10.1037/0096-1523.14.1.113.

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: an artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General, 128*, 165–185. doi:10.1037/0096-3445.128.2.165.

Dutoit, T., Pagel, N., Pierret, F., Bataille, O., & Van Der Vrecken, O. (1996). The MBROLA Project: towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes. Proc. ICSLP'96. *Philadelphia, 3*, 1393–1396.

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development, 64*, 637–656. doi:10.1111/1467-8624.ep9308115002.

Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015a). *Rapid serial auditory presentation*. A new measure of statistical learning in speech segmentation: Experimental Psychology. doi:10.1027/1618-3169/a000295.

Franco, A., Gaillard, V., Cleeremans, A., & Destrebecqz, A. (2015b). Assessing segmentation processes by click detection: online measure of statistical learning, or simple interference? *Behavior Research Methods,*. doi:10.3758/s13428-014-0548-x.

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition, 117*, 107–125. doi:10.1016/j.cognition.2010.07.005.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: a recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review, 118*, 614–636. doi:10.1037/a0025255.

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: learning along the statistical garden path. *Cognitive Science, 33*(6), 1087–1116. doi:10.1111/j.1551-6709.2009.01041.x.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science, 33*, 260–272. doi:10.1111/j.1551-6709.2009.01012.x.

Gómez, R. (2007). Statistical learning in infant language development. In M. G., Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 601-616). New York: Oxford University Press.

Gómez, D. M., Bion, R. A. H., & Mehler, J. (2011). The word segmentation process as revealed by click detection. *Language and Cognitive Processes, 26*, 212–223. doi:10.1080/01690965.2010.482451.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General, 130*, 658–680. doi:10.1037/0096-3445.130.4.658.

Johnson, E. K. (2012). Bootstrapping language: Are infant statisticians up to the job? In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 55–89). Berlin: De Gruyter Mouton.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language, 44*, 548–567. doi:10.1006/jmla.2000.2755.

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science, 12*, 131–141. doi:10.1111/j.1467-7687.2008.00740.x.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning of word segmentation. *Developmental Science, 13*(2), 339–345. doi:10.1111/j.1467-7687.2009.00886.x.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3*, 323–328. doi:10.1016/S1364-6613(99)01363-7.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999a). Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics, 61*, 1465–1476. doi:10.3758/BF03213111.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999b). The beginnings of word segmentation in english-learning infants. *Cognitive Psychology, 39*, 159–207. doi:10.1006/cogp.1999.0716.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing the assumptions of statistical learning: is it long-term and implicit? *Neuroscience Letters, 461*, 145–149. doi:10.1016/j.neulet.2009.06.030.

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science, 14*, 1323–1329. doi:10.1111/j.1467-7687.2011.01079.x.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science, 6*, 314–316. doi:10.1111/j.1467-9280.1995.tb000517.x.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91–121. doi:10.1016/S001002770000109-8.

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. doi:10.1006/cogp.1999.0721.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21–46. doi:10.1006/jmla.1998.2568.

Minier, L., Fagot, J., & Rey, A. (2015). The temporal dynamics of regularity extraction in non-human primates. *Cognitive Science,*. doi:10.1111/cogs.12279.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66*, 911–936. doi:10.1111/1467-8624.ep9509180265.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*, 191–243. doi:10.1006/cogp.1997.0671.

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: extracting word-like units when only statistical information is available. *Journal of Memory and Language, 66*, 807–818. doi:10.1016/j.jml.2012.02.010.

Perruchet, P., Poulin-Charronnat, B., Tillmann, B., & Peereman, R. (2014). New evidence for chunk-based models in word segmentation. *Acta Psychologica, 149*, 1–8. doi:10.1016/j.actpsy.2014.01.015.

Perruchet, P., & Tillmann, B. (2010). Exploiting multiple sources of information in learning an artificial language: human data and modeling. *Cognitive Science, 34*, 255–285. doi:10.1111/j.1551-6709.2009.01074.x.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: no need for algebraic-like computations. *Journal of Experimental Psychology: General, 133*(4), 573–583.

Perruchet, P., & Vinter, A. (1998). PARSER: a model for word segmentation. *Journal of Memory and Language, 39*, 246–263. doi:10.1016/jmla.1998.2576.

Radeau, M., & Morais, J. (1990). The uniqueness point effect in the shadowing of spoken words. *Speech Communication, 9*(2), 155–164. doi:10.1016/0167-6393(90)900068-K.

Robinet, V., Lemaire, B., & Gordon, M. B. (2011). MDLChunker: a MDL-based cognitive model of inductive learning. *Cognitive Science, 35*, 1352–1389. doi:10.1111/j.1551-6709.2011.01188.x.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928. doi:10.1126/science.274.5294.1926.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: the role of distributional cues. *Journal of Memory and Language, 35*, 606–621. doi:10.1006/jmla.1996.0032.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience, 5*(7), 700–703. doi:10.1018/nn873.

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. doi:10.1037/0012-1649.39.4.706.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134*(4), 552–564. doi:10.1037/0096-3445.134.4.552.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: the role of marker frequency. *Journal of Memory and Language, 27*, 71–86. doi:10.1016/0749-596X(88)90049-6.

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech segmentation in a simulated bilingual environment: a challenge for statistical learning? *Language Learning and Development, 5*(1), 30–49. doi:10.1080/15475440802340101.

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences, 8*(10), 451–456. doi:10.1016/j.tics.2004.08.006.