Journal of Memory and Language 64 (2011) 171–180

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The contribution of language-specific knowledge in the selection of statistically-coherent word candidates

Juan M. Toro [a,*], Ferran Pons [b], Ricardo A.H. Bion [c], Núria Sebastián-Gallés [a]

[a] Brain and Cognition Unit, Universitat Pompeu Fabra, Barcelona, Spain
[b] Departament de Psicologia Bàsica, Universitat de Barcelona, Barcelona, Spain
[c] Department of Psychology, Stanford University, Stanford, United States

ARTICLE INFO

Article history:
Received 11 January 2010
revision received 8 November 2010
Available online 22 December 2010

Keywords:
Language
Statistical learning
Speech perception
Linguistic constraints

ABSTRACT

Much research has explored the extent to which statistical computations account for the extraction of linguistic information. However, it remains to be studied how language-specific constraints are imposed over these computations. In the present study we investigated if the violation of a word-forming rule in Catalan (the presence of more than one mid vowel within a word) may interfere with word extraction by statistical computations. Catalan native adult participants were presented with a continuous speech stream composed of trisyllabic nonsense words that violated this linguistic constraint. In a subsequent test, participants did not recognize the words from matched foils. Nevertheless, the same words were recognized if the test comprised foils that never appeared during familiarization, or if both words and foils were presented visually. Results suggest that background linguistic knowledge modulates the recognition of statistically-coherent words, but not the on-line computations leading to their extraction.

Introduction

Tracking distributional regularities from the environment is a fundamental process for organizing stimuli at all levels. This idea was already an important aspect in the work of earlier comparative psychologists (e.g. Rescorla & Wagner, 1972), and the extent to which it provides the basis for higher mental processes is at the core of many current debates in Cognitive Sciences. The field of language processing has also been the ground for such debates. Influential studies have demonstrated that both adults and infants can compute simple statistics over syllables to extract groupings with high statistical coherence from an acoustically continuous speech signal (Saffran, Aslin & Newport, 1996; Saffran, Newport & Aslin, 1996). In natural languages, syllables forming words have a higher probability of appearing together than syllables spanning word boundaries. This ability to track statistical regularities would putatively help to segment words from speech in the absence of any lexical knowledge (e.g. Saffran, 2003). It has been demonstrated that participants prefer statistically-coherent nonsense words than similar foils after being familiarized with an artificial speech stream containing only statistical cues to signal the beginning and ending of words (Saffran, Aslin & Newport, 1996). Thus, it has been proposed that a mechanism tracking statistical regularities among syllables can provide a foothold on the problem of speech segmentation.

Importantly, the computations involved in the extraction of statistical regularities are both domain and modality general (see Aslin & Newport, 2008 for a review). That is, they are not only restricted to language, but seem to be the product of a general-purpose mechanism that can be applied to almost any kind of stimuli. However, when humans use such mechanism to segment speech, it seems to be influenced by an array of linguistic features including

* Corresponding author. Address: Brain and Cognition Unit, Universitat Pompeu Fabra, C. Roc Boronat, 138, CP. 08018, Barcelona, Spain.
E-mail address: juanmanuel.toro@upf.edu (J.M. Toro).

0749-596X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.11.005

prosody (Shukla, Nespor, & Mehler, 2007), coarticulation (Fernandes, Ventura, & Kolinsky, 2007), information processing capacities (e.g. attention, Toro, Sinnett, & Soto-Faraco, 2005), and by previous linguistic knowledge in the form of phonotactic constraints (Finn & Hudson Kam, 2008). To begin with, knowledge about statistical coherence within a speech stream seems to be easily overridden by simple prosodic cues. Shukla and collaborators (2007) showed that adult participants in an artificial language learning experiment could not easily recognize statistically-coherent nonsense words spanning intonational contours. Similarly, adult participants seem to prefer coarticulation cues over statistical regularities to mark word boundaries (Fernandes et al., 2007); and statistical information is not readily extracted from all elements with the same ease, as certain phonological representations (consonants) are preferred over others (vowels) as the target of statistical computations (Bonatti, Peña, Nespor, & Mehler, 2005, but see Newport & Aslin, 2004). So, even though it has been demonstrated that statistical information present in speech can be readily tracked, it is still an open issue the extent to which it is actually used as a reliable cue for word segmentation when it is not combined with other sources of linguistic information (Casillas, 2008; Gervain & Mehler, 2010; Yang, 2004).

A considerable amount of research has been devoted to understanding the integration of different cues in the early stages of language learning. Infant studies on word segmentation demonstrate that, in English, prosodic cues are very useful for segmentation purposes (likely due to the trochaic bias of the language). These cues can be reliably used as early as 7.5 months of age to extract disyllabic words from connected speech, (Johnson & Jusczyk, 2001; Jusczyk, Houston, & Newsome, 1999). Evidence of this ability in French-learning infants at the same age has not been found, suggesting that early rhythmic biases could be guiding later segmentation abilities (Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006). When the use of stress cues for speech segmentation is pitted against the use of statistical cues, results show that by 7 months of age infants rely more heavily on statistics (Thiessen & Saffran, 2003). But, by 11 months, infants use prosody as the preferred marker of word boundaries (Johnson & Seidl, 2009), suggesting a gradual integration of different segmentation cues. In fact, even in laboratory settings, infants can learn to use stressed syllables as relevant features of words and their boundaries (Curtin, Mintz, & Christiansen, 2005; Thiessen & Saffran, 2007). Thus, from an early age, information about statistical dependencies between syllables seems to be integrated with other sources of relevant information regarding the parsing of the speech input.

To have a more complete picture of how statistical processes operate over linguistic stimuli, it is also necessary to establish how previous knowledge influences the extraction of new statistical regularities by adults, once development has been completed and all processing capacities are operating. A way to approach this issue is to assess how successful participants are at tracking dependencies in an artificial speech stream after they have learned similar dependencies in a previous stream. Results of studies taking this approach are mixed. While Weiss, Gerfen, and Mitchel (2009) reported that adult participants successfully segment two streams only when indexical cues help to discriminate them, Gebhart, Aslin, and Newport (2009) found that participants readily recognize the words forming the first speech stream, but not those forming the second (unless explicit instructions regarding the existence of two different languages, and 30 s of silence were introduced between languages).

A second way to approach the effects of previous knowledge on the extraction of statistical regularities is to explore how the listener's native language constrains the extraction of novel distributional regularities. Finn and Hudson Kam (2008) advanced in this direction by showing that English speakers did not prefer nonsense words over foils when these words began with a consonantal cluster that violated phonotactic rules in English (e.g. /tfobu/). In their study the authors presented participants with a stream of nonsense words containing either valid (Control group) or invalid (experimental group) English onsets. While participants in the control group correctly segmented the stream, participants in the experimental group did not. Results thus showed that the presence of word onsets violating phonotactic rules prevented participants from correctly segmenting the stream using statistical computations, as linguistic knowledge interfered with the extraction of distributional information.

However, aspects in the stimuli used by Finn and Hudson Kam (2008) warrant further explorations of the interaction of linguistic knowledge and the extraction of statistical regularities. Across languages, onset clusters such as /tf/, /bt/, and /ps/ tend to be universally dispreferred in comparison to clusters such as /pl/, /kr/, /tw/ (Greenberg, 1978). Some proposals claim that this pattern is due to domain-general restrictions on auditory perception (Ohala, 1990), while others emphasize universal constraints in the sound structure of languages (Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009). The difference in sonority (partly correlated to the difference in physical energy) between the speech sounds forming the cluster can explain differences in preferences (Berent, Steriade, Lennertz, & Vaknin, 2007). Crucially, the clusters in the words rejected by the participants in the experimental group in the study by Finn and Hudson Kam (2008) had different sonorities from the clusters in the words segmented by the participants in the control group. Therefore, the clusters in the experimental group could be dispreferred as compared to the clusters present in their control group because of factors not related to the extraction of statistical regularities during the experiment. This distinction is central to identify how either language-specific, or more universal language knowledge, constrains statistical computations. Closely related to this difference in preferences, the probability of an English word boundary occurring within the clusters forming the words used in their experimental group was very high (on average, 0.8) making it difficult to recognize these sequences as words, while the opposite was true for the words in the control group (on average, 0.3), making it more likely that they could be recognized as valid word candidates. While some phonotactic knowledge can be acquired through the extraction of statistical dependencies (Onishi, Chambers, & Fisher,

2002), it is not clear if participants were using the transitional probabilities between phonemes in their language to segment the stream. That is, results reported so far have clearly demonstrated an effect of invalid onset clusters over word recognition. But it is still unknown how previous language-specific knowledge interacts with statistical computations, as there is the possibility that universal preferences are playing a substantial role. Even more, it is important to dissociate the effects that linguistic background might have over the on-line computation of statistical dependencies and over the recognition of already-segmented words.

Thus, the studies reviewed so far have shown that statistical information interacts with several different sources of information. Both statistical regularities extracted within a laboratory setting, and more universal linguistic preferences seem to affect subsequent learning of distributional dependencies among syllables. Linguistic knowledge may interfere with how syllables are grouped at three different levels: (1) by imposing constraints on which combinations are considered as valid word onsets (Finn & Hudson Kam, 2008); (2) by filtering valid word candidates once distributionally coherent sequences are extracted by statistical computations (as in Shukla et al., 2007), or by (3) imposing reliable competing segmentation strategies which are favored over statistical computations (see Norris, McQueen, Cutler, & Butterfield, 1997, for an empirical demonstration of the possible-word constraint). In the present study we try to move this issue forward and disentangle the level at which specific linguistic constraints are implemented when a listener is extracting distributional information from a new speech stream. We approach this issue focusing in two relevant aspects. First by trying to dissociate the effects of universal and language-specific constraints. And second, by determining whether these constraints are operating at the "speech segmentation" or the "word recognition" level. Importantly, segment combinations in the present study only violated word-forming constraints specific to the participants' native language, and involved phoneme combinations that did not violate universal phonemic restrictions and preferences (see below). The current experimental series also departs from previous works in two important aspects. First, most studies on how other sources of information constrain statistical processes have been performed with English speakers, which makes it necessary to explore the performance of speakers of other languages to asses both language-specific and language-general constraints on statistical computations (see Toro, Sebastián-Gallés, & Mattys, 2009; Tyler & Cutler, 2009). In the experiments reported in the present work, participants' native language was Catalan (a Romance language). Second, and more importantly, in the present study we implemented linguistic constraints not at the segmental, but at the word level. More specifically, we explored if segmental combinatorial rules on word formation that are specific to the participants' native language may "filter out" statistically-coherent groups of syllables to be considered valid lexical candidates.

For this, we took advantage of vowel reduction in Catalan. In languages that reduce vowels (such as English), unstressed vowels are reduced, usually to schwa. In Catalan, this constraint only applies to mid and low vowels: they can only appear in stressed positions, otherwise, they are reduced. In the word "dirigent" [diɾiʒ'en] (leader), the vowel /e/ is not reduced because it is stressed. But in the word "escoltar" [əskul'ta] (to listen), all the vowels but the stressed vowel /a/ are reduced. Similar examples include words like "dependent" [dəpən'den] (salesman) where stress is on the last syllable, and all other vowels are reduced, and "pastanaga" [pəstə'nagə] (carrot) where only the penultimate /a/ is not reduced, as is located in the stressed position. More generally, in unstressed syllables, the vowels /o/ and /ɔ/ are reduced to [u], and the vowels /a/, /e/ and /ɛ/ are reduced to [ə]. The vowels /i/ and /u/ are not reduced. For instance, the stressed vowel [e] in the word "verd" [bɛr] (green) is reduced to [ə] in the morphological derivation "verdós" [bər'ðos] (greenish), and the vowel [ɔ] in the word "ploure" [plɔ́urə] (to rain) is reduced to [u] in "plovent" [plu'βen] (raining) (Wheeler, 2005). Because of this, no more than one unreduced mid or low vowel can appear in a morphologically simple word. Creating nonsense words that either comply with or violate such word-forming constraints can result in two possible outcomes. First, statistical computations may produce word candidates that are preferred to foils because of their strong distributional coherence, with small influences from other sources of information besides universal restrictions in phoneme combinations. On the other hand, language-specific knowledge may filter out unlikely word candidates, making it more difficult for participants to prefer them to foils. In Experiment 1a we presented participants with an artificial speech stream composed of words violating a Catalan suprasegmental rule on word formation. With this experiment we wanted to test if participants could recognize statistically-coherent items, even though they violate a given constraint in their native language. In Experiments 1b and 1c we explored the level at which previous linguistic knowledge might be influencing the extraction of statistical dependencies between syllables by presenting participants with test foils that never appeared during familiarization, or with written test items. Finally in Experiment 2a, to ensure participants could extract coherent words when they did not violate any linguistic constraint, all words composing the stream were well formed according to the rules of the participants' native language. In order to explore if results were not due to any *a priori* preference for a given set of test items, we also ran Experiment 2b, in which participants were presented during familiarization with the same set of syllables as in Experiment 2a, but arranged in a completely random manner.

Experiment 1a. Words violating language-specific constraints

Participants

Eighteen Psychology students from the Universitat de Barcelona participated in this study for course credit. They were native Catalan speakers. None of them reported hearing deficits.

Stimuli

We created an artificial speech stream by concatenating four nonsense CVCVCV trisyllabic words. Each of the words was composed of three mid-vowels, including open-mid [ɛ] and [ɔ], and close-mid [e] and [o] vowels (see Table 1). As mentioned, mid-vowels in Catalan are always stressed, so no more than one can occur in a given morphologically simple word. Therefore all four nonsense words composing the stream violated word-forming constraints in the participants' native language. A female native Catalan speaker produced in isolation a series of four tokens of each syllable. After one of these tokens was selected, the syllables were normalized for mean duration (280 ms), mean pitch (175 Hz), and intensity (56 dB) using Praat software (Boersma & Weenink, 2009). They were then concatenated at their zero crossings into a continuous speech stream that contained no acoustic cues regarding the beginning or ending of the words. Each word was presented 214 times, with the only restriction that no immediate repetitions were allowed. Thus, the probability that any given syllable was followed by another was 1.0 within words, and 0.33 between words. The resulting stream lasted a total of 12 min. Eight trisyllabic test items were also created. These consisted in the four "words" that composed the stream (/bedɔpɛ/, /tɛmose/, /mɔbɛno/ and /dotekɔ/), and four "part-words" created by concatenating the last syllable of a word and the two first syllables of another, resulting in the items /pɛmɔbɛ/, /sedote/, /nobedɔ/ and /kɔtɛmo/. Importantly, both words and part-words were formed by three different mid-vowels and three different consonants. Therefore, only information about transition probabilities could be used as a cue to segment this stream. Each test item lasted 840 ms.

Procedure

Participants were tested individually in a sound-attenuated room. They were told they would listen to a Martian language, and their task was simply to pay attention to it. Participants were presented with the continuous speech stream through headphones for 12 min. Immediately after the presentation of the stream they were given a 2-alternative forced choice (2AFC) test. They were presented with two test items (a word and a part-word) separated by 500 ms. They were asked to press a key indicating whether the first ("1") or the second ("2") word they heard was more likely to belong to the Martian language they heard before. Each of the four words in the stream were matched with four part-words for a total of 16 test trials. Order of presentation was counterbalanced.

Table 1
Nonsense words used to create artificial speech streams in the present study.

Experiment 1	Experiment 2
bedɔpɛ	bedupə
tɛmose	timuse
mɔbɛno	mobinu
dotekɔ	duteki

Results and discussion

Responses were converted to a percent of correct responses and averaged across participants. Performance in the 2AFC test was not better than chance (46.5%, SD = 15.3; $t(17) = 0.96$, $p = .35$; see Fig. 1). Thus, participants did not prefer words to part-words, suggesting that the violation of the word-forming constraint in their native language prevented the use of statistical information alone as cue to creating new word candidates. Nevertheless, Finn and Hudson Kam (2008) showed that participants in their study did recognize the words violating English phonotactic constraints when they were paired with words that did not occur at all during familiarization. There is thus the possibility that participants in our experiment may also recognize the words if they are pitted against syllable combinations that did not appear during familiarization. If so, it would suggest that violation of language-specific word-forming rules do not completely prevent the formation of statistically-coherent syllable groupings, but rather makes it more difficult to later recognize them. We thus presented new participants with exactly the same familiarization speech stream as in Experiment 1a, but this time the words during the test were paired with non-words, that is, with syllable sequences that had not appeared during familiarization.

Experiment 1b. Words versus non-words

Participants

Eighteen Psychology students from the Universitat de Barcelona participated in this study for course credit. They were native Catalan speakers. None of them reported hearing deficits, and none of them participated in Experiment 1a.

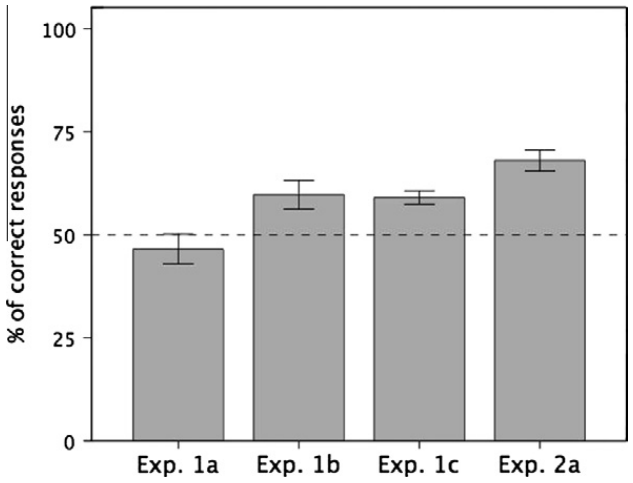

Fig. 1. Mean percentage of correct responses and standard error in Experiments 1a, 1b, 1c and 2a. Chance level is at 50%. Participants in Experiment 1a did not prefer words forming the speech stream to foils. In Experiment 1b, performance improved when words were tested against non-words. Participants in Experiment 1c also correctly identified the words from the part-words when test items were presented visually. Finally, in Experiment 2a participants segmented a speech stream made of well-formed words.

Stimuli

The familiarization stream was identical to the one used in Experiment 1. The only difference is that during test, the words (/bedɔpɛ/, /tɛmose/, /mɔbɛno/ and /dotekɔ/) were paired with non-words. These were syllable sequences that did not appear in the familiarization stream, and were created by concatenating syllables composing the words, organized in sequences that never were presented to the participants during familiarization (/pɛmodo/, /notemɔ/, /sedɔbe/ and /kɔbɛtɛ/).

Procedure

The procedure was the same as in Experiment 1a.

Results and discussion

Participants preferred words over non-words significantly over chance (59.7%, $SD = 14.7$; $t(17) = 2.8$, $p < .05$). This result was different from that observed in Experiment 1 ($t(34) = -2.63$, $p < .05$). That is, even though the words forming the artificial speech stream violated word-forming constraints in the participants' native language, they could recognize them above chance levels when compared with words composed by the same syllables, but arranged in sequences that were not encountered during familiarization. Importantly, non-words also violated the same word-forming constraints as words, so any preference for the latter reflects that statistical information was extracted during the familiarization phase as to form coherent groupings of syllables. This result suggests that the computation of statistical dependencies among syllables is not suppressed by background linguistic knowledge. Statistics seem to be nevertheless computed among syllable sequences (as suggested by the preference of words over non-words). Linguistic knowledge seems to affect the process of recognizing those sequences among likely candidates.

But what is the level at which background linguistic knowledge is affecting the recognition of the statistically-coherent words forming the stream? Shukla et al. (2007) showed participants could not extract statistically-coherent words from an artificial speech stream if they straddled prosodic contours. If the words were internal to the prosodic contour, participants could extract them. This suggested prosody affected the extraction of word-like units using statistical information. Nevertheless, Shukla and collaborators presented participants in a further experiment with written test items, and found that participants correctly recognized words that straddled prosodic contours during familiarization. The authors concluded prosody did not override the computation of statistical dependencies between syllables, but rather filtered its results. In a similar line, it is important to determine if the violation of a word-forming rule in Catalan is affecting participants during the on-line computation of dependencies, or if it is, like prosody, filtering the recognition of what may be considered valid syllabic sequences. In Experiment 1c we presented participants with exactly the same familiarization stream as in Experiment 1a, but this time test items were presented visually, written on the computer screen.

Experiment 1c. Written test items

Participants

Eighteen Psychology students from the Universitat de Barcelona participated in this study for course credit. They were native Catalan speakers. None of them reported hearing deficits, and none of them had participated in Experiments 1a or 1b.

Stimuli

The familiarization stream and test items were identical to the ones used in Experiment 1a. The only difference is that during test, the words (/bedɔpɛ/, /tɛmose/, /mɔbɛno/ and /dotekɔ/) and the part-words used during test (/pɛmɔbɛ/, /sedote/, /nobedɔ/ and /kɔtɛmo/) were presented visually instead of acoustically. Participants were thus presented with a phonetic transcription of the test items. In the Catalan schooling system children are taught the phonetic transcription of vowels used in the language. All participants reported to be familiar with these transcriptions before starting the experiment. In each test trial the participant was presented with the two test items written on the computer screen, one on the left side of the screen and the other on the right side of the screen. The position of the items in the screen was counterbalanced.

Procedure

The procedure was the same as in Experiment 1a.

Results and discussion

Participants preferred the words over the part-words when they were presented visually (59.1%, $SD = 6.8$; $t(17) = 5.58$, $p < .005$). Performance with the visual presentation of test items significantly differed from performance when the same items were presented acoustically, as in Experiment 1a ($t(34) = -3.15$, $p < .005$). This suggests that the on-line computation of statistical dependencies between syllables was not affected by the fact that statistically-coherent sequences violated word-forming constraints in the participants' native language. If this was the case, words could not have been preferred over part-words independent of the mode of presentation (acoustic or visual). Instead, the effect of background linguistic knowledge seems to be by-passed by the presentation of test items in the visual modality. A visual presentation of the test items seems to have allowed the participants to use the dependencies learned during familiarization to distinguish words from part-words. By presenting visual test items, participants had to read them, likely reducing the mid-vowels and producing well-formed words in Catalan that are statistically-coherent as a result of the familiarization phase. If this interpretation is true further studies

could explore the possibility that acoustically presenting test items with reduced vowels should also allow participants to recognize words over foils. A different interpretation of the present results is that linguistic constraints are only imposed by an auditory test phase. That is, the effects of linguistic knowledge might not be inherent to the process, but only imposed by a test phase on the auditory modality. The fact that participants preferred ill-formed words over items that did not appear at all during familiarization (Experiment 1b) seems to provide support to this later interpretation, as constraints were not apparent when contrasting items in the test phase were very different. Thus, statistical computations over syllables during familiarization would go unhindered by linguistic knowledge. Only later, on the word recognition test phase, background knowledge of what can be consider as a valid syllable combination in the participants' native language would guide the preference for sequences with higher statistical dependencies among their syllables.

Finally, to confirm that segmentation of a speech stream using distributional regularities could be improved by using words abiding to word-forming constraints in the participants' native language, we presented new participants with a speech stream composed only of words complying with all compositional rules across segments in Catalan, so no more than one mid vowel was present within each word.

Experiment 2a. Items abiding to word-forming rules in Catalan

Participants

Eighteen Psychology students from the Universitat de Barcelona participated in this study for course credit. They were native Catalan speakers. None of them reported hearing deficits, and none of them had participated in previous experiments from this study.

Stimuli

We created a continuous speech stream and eight new test items in exactly the same manner as in Experiment 1. The words composing the stream were composed of close-mid-vowels [e] and [o], high vowels [i] and [u], and the schwa [ə] (see Table 1). The test items comprised words (/bedupə/, /timuse/, /mobinu/ and /duteki/) and part-words used as test foils (/pəmobi/, /sedute/, /nubedu/ and /kitimu/). In order to present test items that were matched for their statistical dependencies, made by putting together the last syllable of a word with the two first syllables of another one, we had to include one item containing two unstressed mid-vowels as a part-word (/sedute/). Item analyses were included to explore differences in response to this item.

Procedure

The procedure was exactly the same as in Experiment 1a to 1c.

Results and discussion

Participants preferred words to part-words above chance levels (68.1%, $SD = 10.6$; $t(17) = 7.1$, $p < .005$), suggesting they correctly segmented the speech stream. Results from Experiment 2a were significantly different from those from Experiment 1a ($t(34) = 4.88$, $p < .005$). Thus, when the stream is composed of legally-formed words, Catalan participants correctly segment them as to prefer the words from the part-word foils. An item analysis yielded no differences in responses for any of the test items used during the test ($F(3, 68) = 2.09$, $p = .11$). Results from Experiment 2a were also significantly different from those observed with written test items (Experiment 1c; $t(34) = 3.05$, $p < .005$). This suggests a general improvement over the recognition of the words once they abide with all constraints in the participants' native language. Nevertheless, one should consider the possibility that this difference might also reflect the cost of modality changes from familiarization (auditory) to test (visual). Conway and Christiansen (2006) showed that it is difficult for participants to generalize statistical regularities learned in one modality into another modality. Participants in our Experiment 1c performed above chance at visually recognizing words that were originally presented acoustically. Nevertheless, their performance was significantly below that of participants in Experiment 2a where items were presented in the same modality in the familiarization and test phases. More conclusive is thus the fact that there was a significant improvement in the performance between experiments in which words did not abide to word-forming rules in Catalan (Experiment 1a) and that in which words were well-formed in the participants' native language (Experiment 2a).

As in many other artificial language experiments, there is the possibility that the preferences found in the test phase for a given set of items were not motivated by the exposure phase but rather were an artifact of the test phase. In this case, it would be important to investigate whether the same preferences would be observed when the exposure phase was composed of randomly organized syllables or clusters. We ran a further experiment to rule out the possibility these results were driven by an *a priori* preference for the nonsense words used during the test phase of this experiment. We thus presented new participants (drawn from the same pool of participants from all previous experiments) with a speech stream composed of the same syllables as in Experiment 2a, but arranged completely at random.

Experiment 2b. Random distribution of syllables

Participants

Eighteen Psychology students from the Universitat de Barcelona participated in this study for course credit. They were native Catalan speakers. None of them reported hearing deficits, and none of them had participated in previous experiments from this study.

Stimuli

We composed a familiarization stream with the same syllables as those used in Experiment 2a. Nevertheless, instead of being organized around four trisyllabic words, they were concatenated randomly. Transitional probabilities between syllables ranged from 0.055 to 0.106. After listening to this stream for 12 min, they were given the exact same test as participants from Experiment 2.

Procedure

The procedure was exactly the same as in Experiment 2a.

Results and discussion

Results showed participants did not prefer words to part-words (52.4%, $SD = 9.8$; $t(17) = 1.1$, $p = .31$). Performance in this experiment was different from that in Experiment 2a ($t(34) = 4.55$, $p < .005$). Nevertheless, an item analyses showed differences among participants' preferences for different words ($F(3, 68) = 3.02$, $p < .05$). In fact, participants tended to prefer the words when these were paired with the ill-formed part-word /sedute/ (68%). On the contrary, participants did not prefer words when they were paired with any of the other part-words. As could be expected, in the absence of strong statistical cues that group together the syllables, participants tend to dis-prefer words that are not well formed in their language. To further explore if results observed in Experiment 2a were not somehow driven by any familiarization with the syllables, we ran another experiment in which participants were presented with the same test phase as in Experiments 2a and 2b, but with no exposure phase. This experiment yielded almost identical results to those observed in Experiment 2b (50.8%, $SD = 7.2$). Thus, preference for nonsense words in Experiment 2a could not be attributed to an *a priori* preference for the specific test items used during testing. Instead, results suggest that participants can correctly segment the stream composed of words that do not violate any word-forming rule in their native language. When nonsense words violate such rules, statistically-coherent groupings of syllables are not recognized above chance levels (Experiment 1a).

General discussion

In this study we addressed the extent to which the violation of language-specific rules in the formation of nonsense words would affect their segmentation from a speech stream using statistical information or their subsequent recognition. In the present study we disambiguated the contribution of language-specific knowledge by using phoneme combinations that are illegal in the participants' native language but that do not violate general restrictions on phoneme combinations. We took advantage of word-forming constraints in Catalan, the native language of the participants. In Catalan, mid-vowels occur in stressed positions, otherwise, they are reduced. In Experiment 1a, results showed that when words composing the stream violate this rule, that is, when they contain up to three mid-vowels, participants do not prefer them from matched foils in a subsequent test. When words are contrasted with test items that did not occur at all during familiarization, participants recognize them (Experiment 1b), suggesting they have in fact learned some regularities from the stream. Even more, participants prefer words to part-words if both are presented visually (Experiment 1c). So any effect the participants' background linguistic knowledge may have over the segmentation of the stream seems to be modulated by the modality of presentation of items. Finally, in experiment 2a, participants correctly segment the artificial speech stream once all four words comply with formation rules in Catalan. The results thus suggest that language-specific knowledge modulates the recognition of the word candidates, but very likely does not compromise the on-line computation of statistical dependencies between syllables during familiarization.

The present set of results extends previous studies exploring the interaction of the participants' knowledge of certain linguistic regularities, and the extraction of novel statistical dependencies. It has been demonstrated that speakers know universal restrictions of what can be considered valid phoneme combinations (Berent et al., 2007, 2009). In the current study, we have demonstrated that knowledge of word-forming constraints that are specific to a given language (in our case, Catalan) affects the recognition of statistically-defined words. Importantly, all well- and ill-formed words across all experiments in the present study have the same statistical distribution, so differences in performance cannot be accounted for by any putative difference across distributions in the exposure material. Finally, preference for certain test items (as in Experiment 2a) cannot be accounted for by *a priori* preferences, as was shown by the control experiment reported above. Such preference can only be explained in terms of correct segmentation of the stream using statistical information, without interference from linguistic constraints that barred possible words.

At what level are word-forming constraints acting in the present study? One possibility is that knowledge about word-forming constraints in one's language is already helping to segment an artificial speech stream (e.g. Norris et al., 1997). That is, given that two mid-vowels cannot occur in a single word in Catalan, participants could be assigning a word boundary when two mid-vowels occurred adjacently. In Experiment 1a, this would result in segmenting the stream into monosyllabic words, and therefore showing no segmentation at all. In fact, in the study by Finn and Hudson Kam (2008), experimental words began with consonantal clusters that were not valid word onsets in English (or that are universally dispreferred, as an alternative explanation). This prevented participants from correctly segmenting the words from the stream, suggesting that phonotactic rules were in fact modulating the way syllables were grouped, and interfering with statistical computations. If the rules used in our study are acting at the same level, results would show that word-forming information would completely override statistical cues as a marker of word boundaries. On the

contrary, it seems to be the case that violations of a word-forming rule do not interfere with the "on-line" computation of statistical dependencies between syllables during familiarization, but with their later selection as valid word candidates. Remember that rules used in our study are implemented at the suprasegmental level, and affect how well formed a word is considered. It thus seems unlikely they are acting over the very process that groups syllables into what would be word candidates. The rules we used in the present study are only meaningful once statistically-coherent groups of syllables have been extracted from the stream. Even more, participants in Experiment 1b in the current study correctly recognize words when they are pitted against non-words (see also Finn & Hudson Kam, 2008). It thus is possible that constraints are "filtering out" ill-formed groupings of syllables, in a similar way as prosodic contours have been proposed to act during a similar task (Shukla et al., 2007). That is, not by preventing the extraction of statistical information across syllables in a speech stream, but rather by selecting later which groupings would be coherent with prosodic boundaries. If this is the case in the present study, as results from Experiments 1b and 1c suggest, statistical computations would nevertheless be performed over the stream of syllables, extracting sequences with high distributional coherence. However, once these sequences are extracted, language-specific information would reject sequences that violate a given constraint on word formation.

The extraction of distributional regularities among syllables within a continuous stream has been shown to be a powerful source of information for the formation of coherent units (e.g. Aslin & Newport, 2008). The work by Graf Estes and collaborators (e.g. Graf Estes, Evans, Alibali, & Saffran, 2007) suggests that young and adult learners can use syllable groupings extracted via statistical computations as words. Moreover, infants' abilities to track frequency information in speech are involved in language processing at several levels, including the organization of phonetic categories (Maye, Werker, & Gerken, 2002), learning non-adjacent morphological dependencies (van Heugten & Johnson, 2010), and identifying function words (Gervain, Nespor, Mazuka, Horie, & Mehler, 2008; Hochmann, Endress, & Mehler, 2010). The present results showing linguistic filtering during the recognition of valid syllable sequences contributes to a growing literature that suggests multi-facet interactions across different levels during word segmentation. Linguistic rules would help to select sequences of elements extracted via statistical computations that fit within the appropriate language context and discard those that do not (e.g. Peperkamp, Le Calvez, Nadal, & Dupoux, 2006). Thus, these rules would reduce the elements fed into the system, controlling for part of the noisy input and helping to solve the problem posed by the combinatorial explosion of almost-infinitive possible regularities extracted via statistical computations among a given set of items.

The extent to which the units resulting from statistical computations are used as actual words is a concern that has been raised by both adult (Endress & Mehler, 2009; Turk-Browne & Scholl, 2009) and infant studies (Johnson & Tyler, 2010). Our current work departs from this issue and focus on the extent to which the recognition of these units is modulated by linguistic knowledge. As such, it adds to a number of studies suggesting that accessing dependencies among different elements is constrained when processing linguistic regularities. These limits may be completely language-independent (in the sense that they are not restricted to a given language) as seems to be the case with prosodic contours (Shukla et al., 2007), coarticulation (Fernandes et al., 2007), final-word lengthening (Saffran, Newport, & Aslin, 1996b), and phonological representations (Bonatti et al., 2005); or they may depend on learned, specific features of each individual language, as in the case of suprasegmental rules reported in the present work. Together, this suggests that during adult speech perception, accessing statistically-coherent sequences of syllables seems to be subordinated to several sources of linguistic information. At the same time, the exact nature of the statistical computations our participants performed over the streams (whether transitional probabilities, or frequency of co-occurrence) cannot be determined with the data we have. Aslin, Saffran, and Newport (1998) showed that infants tend to compute transitional probabilities over syllable streams. The stimuli used in the present study allow us to infer that the participants are computing conditional statistics in order to segment the stream. But we remain neutral to whether these are conditional probabilities or more generally, frequency of co-occurrence. The fact that such probabilities seem to be nevertheless computed by participants in the present study opens the door to the possibility that different levels of representation are used at different moments of the process. It might be that, during segmentation of the stream, participants are encoding the probabilities at the phonetic level on short-term memory. However, when they incorporate these words into their lexicon, these sequences are encoded at the phonological level on long-term memory, being filtered by language knowledge. Differences at the level of encoding could make it more difficult for participants to recognize the words when presented acoustically (Experiment 1a) than when presented visually (Experiment 1c). It would also suggests participants are converting phonetic representations extracted via statistical computations into phonological representations to add new words into their lexicon. If so, further studies could tackle the specific representations over which statistics are being computed when applied over linguistic materials.

The present results bear on an on-going debate in the field of language processing that reflects more general discussions in Cognitive Sciences. The debate focuses on the extent to which different sources of information account for much of stimuli processing. Three of these sources are especially salient: (i) universal biases (e.g. the iambic–trochaic law, Bion, Varela, & Nespor, 2011; differences in sonority among consonant clusters, Berent, Lennertz, Jun, Moreno, & Smolensky, 2008), (ii) universal learning algorithms (TPs or bayesian-type computations), and (iii) previous knowledge (language-specific word-forming constraints and valid phonotactics; although there is debate around the extent to which these are the results of universal learning algorithms). The finding that, from an early age, listeners can track distributional regularities from speech

in order to segment words from it (Saffran et al., 1996a) has fueled much interest in the idea that powerful learning mechanisms may account for much of language processing (e.g. Bates & Elman, 1996). And recent work suggests their implementation over certain types of stimuli can be restricted by general bias (see Endress, Nespor, & Mehler, 2009). The present study aims to further understand how previous knowledge interacts with universal learning algorithms. Indeed, results suggest that computation of TPs is very pervasive, and in our case seems to provide a first analysis of the input. Once this analysis suggests possible word candidates, language-specific knowledge seems to reject words that violate Catalan word-forming rules. However, memory traces for these rejected words are still present, and can be assessed when written stimuli are presented or when these words are compared against non-words.

In sum, there is very consistent evidence that statistical regularities can be extracted from a wide array of stimuli, including syllables within a speech stream. However, the recognition of the statistically-coherent sequences extracted by this general mechanism is constrained when applied to linguistic signals. In this study, we have provided data showing that violation of language-specific rules impedes correct identification of word-candidates resulting from statistical computations. More specifically, background linguistic knowledge restricts what constitutes a valid word candidate once statistically-coherent groupings of syllables are extracted from the speech stream.

Acknowledgments

This research was supported by a *Ramón y Cajal* fellowship (RYC-2008-02909) to JMT, and grants by the Spanish Ministerio de Ciencia e Innovación Consolider Ingenio 2010 (CSD2007-00012), PSI2009-09066 (NSG), PSI2010-20294 (FP), and PSI2010-20029 (JMT), Fondos Feder and ICREA Academia. We would like to thank Marta Ramón for lending her voice to create the stimuli, and Daniela Martinez for helping running the experiments.

References

Aslin, R., & Newport, E. (2008). What statistical learning can and can't tell us about language acquisition. In J. Colombo, P. McCardle, & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research directions*. Mahwah, NJ: Erlbaum.

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

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

Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences, 105*, 5321–5325.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology, 26*, 75–108.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition, 104*, 591–630.

Bion, R. A. H., Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence adults' and infants' memory of speech sequences. *Language & Speech, 54*, 1–18.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1) [Computer program] <http://www.praat.org/> (retrieved 11.05.09).

Bonatti, L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*, 451–459.

Casillas, G. (2008). The insufficiency of three types of learning to explain language acquisition. *Lingua, 118*, 636–641.

Conway, C., & Christiansen, M. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science, 17*, 905–912.

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

Endress, A., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*, 351–367.

Endress, A., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints in language acquisition. *Trends in Cognitive Sciences, 13*, 348–353.

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception & Psychophysics, 69*, 856–864.

Finn, A., & Hudson Kam, C. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108*, 477–499.

Gebhart, A., Aslin, R., & Newport, E. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science, 33*, 1087–1116.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology, 61*, 191–218.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese–Italian cross-linguistic study. *Cognitive Psychology, 57*, 56–74.

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

Greenberg, J. (1978). Some generalizations concerning initial and final consonant clusters. In J. Greenberg, C. Ferguson, & E. Moravcsik (Eds.). *Universals of human language* (Vol. 2). Stanford, CA: Stanford University Press.

Hochmann, J. R., Endress, A., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition, 115*, 444–457.

Johnson, E., & Jusczyk, P. (2001). Word segmentation by eight-month-olds: When speech counts more than statistics. *Journal of Memory and Language, 44*, 548–567.

Johnson, E., & Seidl, A. (2009). At 11 months, prosody still outranks statistics. *Developmental Science, 12*, 131–141.

Johnson, E., & Tyler, M. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*, 339–345.

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

Maye, J., Werker, J., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–B111.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language, 54*, 283–299.

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

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

Ohala, John. J. (1990). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers on the Regional Meetings of the Chicago Linguistic Society, 2*, 319–338.

Onishi, K., Chambers, K., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*, B13–B23.

Peperkamp, S., Le Calvez, R., Nadal, J. P., & Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition, 101*, B31–B41.

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In H. Black & W. Prokasy (Eds.), *Classical conditioning II*. NY: Appleton-Century-Crofts.

Saffran, J. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110–114.

Saffran, J., Aslin, R., & Newport, E. (1996a). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J., Newport, E., & Aslin, R. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

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

Toro, J. M., Sebastián-Gallés, N., & Mattys, S. (2009). The role of perceptual salience during the segmentation of connected speech. *European Journal of Cognitive Psychology, 21*, 786–800.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*, B25–B34.

Thiessen, E., & Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries in 7- and 9-month-old infants. *Developmental Psychology, 39*, 706–716.

Thiessen, E., & Saffran, J. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development, 3*, 73–100.

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

Tyler, M., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America, 126*, 367–376.

van Heugten, M., & Johnson, E. (2010). Linking infants' distributional learning abilities to natural language acquisition. *Journal of Memory and Language, 63*, 197–209.

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

Wheeler, Max. W. (2005). *The phonology of Catalan*. Oxford: Oxford University Press.

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