LANGUAGE-UNIVERSAL CONSTRAINTS ON THE SEGMENTATION OF ENGLISH

D. Norris,
MRC Cognition and Brain Sciences Unit, Cambridge, U.K.
A. Cutler, J. M. McQueen,
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
S. Butterfield, and R. K. Kearns
MRC Cognition and Brain Sciences Unit, Cambridge, U.K.

ABSTRACT

Two word-spotting experiments are reported that examine whether the Possible-Word Constraint (PWC) [1] is a language-specific or language-universal strategy for the segmentation of continuous speech. The PWC disfavours parses which leave an impossible residue between the end of a candidate word and a known boundary. The experiments examined cases where the residue was either a CV syllable with a lax vowel, or a CVC syllable with a schwa. Although neither syllable context is a possible word in English, word-spotting in both contexts was easier than with a context consisting of a single consonant. The PWC appears to be language-universal rather than language-specific.

1. INTRODUCTION

The segmentation of a written text such as this one into its component words is a trivial task for the reader, because the writer has helpfully left empty spaces between the individual words. Speakers do not help listeners in this way. Spoken utterances are continuous and one of the tasks which the listener has to accomplish, in order to understand what the speaker is trying to say, is segmentation: dividing the continuous signal up into its constituent words.

A considerable body of research on the segmentation of spoken language has shown that in accomplishing this task listeners draw to a considerable extent on their knowledge about the phonological structure of their native language. This produces language-specific effects in segmentation. The rhythmic structure of language helps segmentation in the native language, for instance, but can lead to inappropriate listening strategies when the input is in a non-native language which has a different rhythm (see [2] for a review). Phonotactic sequence constraints can be effectively exploited to segment the native language [3], but again can be misleading when the input is in a non-native language in which the constraints are different [4].

One very powerful weapon in the listener's armoury was discovered by Norris, McQueen, Cutler and Butterfield [1]. This is a constraint - the Possible-Word Constraint (PWC) - which disfavours interpretations which would leave a residue of the input which could not itself be parsed into one or more words. The evidence for the PWC came from an experiment in which listeners were required to spot real words in short nonsense strings. Norris et al. found that words were harder to spot when the residue of the nonsense string was only a single consonant than when the residue was a syllable. Thus *sea* was harder to spot in *seash* than in *seashub*, and *apple* was harder to spot in *fapple* than in *vuffapple*. None of the residues - *sh, shub, f, vuff* - are in fact words of English, but *vuff* and *shub* might have been words. The two single consonants could however never themselves be viable candidate words. Norris et al. proposed that this constraint could provide a powerful method for inhibiting activation of words which are spuriously present in an utterance. Thus even if *they met a fourth time* activated *aim*, *for*, *I'm* and *metaphor*, these words could all be rejected on the grounds that each of them would inevitably leave a single-consonant residue, that is, a residue which could not itself be parsed into words.

The PWC is potentially also a language-specific constraint, because the form which words can take differs across languages. In English, for instance, no word can consist of an open syllable with a short full vowel. Open syllables with long vowels are acceptable (e.g. *sea*) and closed syllables with short vowels (e.g. *sell*) are also fine, but *sE* (with the vowel of *sell*) is not a possible English word. It would, however, be a perfectly fine word of Japanese or French. The PWC might on the other hand also be a constraint which is universal, that is, has the same form across all languages (in which case it might for instance reflect early strategies for acquiring the words of one's language). If the PWC is language-specific, then residues which could not be a word in the native language will make spotting embedded words difficult (even though the same residues might be acceptable words in other languages). If the PWC is language-universal, however, then a residue will only be problematic if it could not be a word in any language. The experiments we report here concern the question of the universality versus language-specificity of the PWC.

2. EXPERIMENT 1

Experiment 1 examined English listeners' ability to detect bisyllabic words with Weak-Strong (WS) or Strong-Weak (SW) stress patterns, in nonsense contexts which could or could not themselves form possible English words. For WS words, *canal* for example, the contexts consisted of a single consonant (*scanal*), a Consonant-Vowel (CV) syllable with a tense vowel (*zeecanal*, with the same vowel as in *peek*), or a CV syllable with a lax vowel (*zEcanal*, with the same vowel as in *peck*). If the PWC is language-specific, *canal* should be harder to spot after *s*- and *zE*- than after *zee*-, since only the latter residue is a possible word of English. If the PWC is language-universal, *canal* should

be hard to spot after *s-* but easier after both *zE-* and *zee-*, which could be words in some language.

For SW words (e.g. *eager*) the contexts were single consonants (*theager*) or CVC syllables, again one with a tense vowel (*zeetheager*) and one with a lax vowel (*zEtheager*). Single-consonant contexts should, again, be difficult. In this case, however, the two syllable contexts did not test whether the PWC is language-specific. Rather, they tested whether the difference between tense and lax vowels influences the location of perceived syllable boundaries. Lax vowels demand a closed syllable (*zEth*), which might lead to the segmentation *zEth-eager*. Detecting *eager* should therefore be easy in this condition, since the word is aligned with the syllable boundary after the /θ/ and the entire first syllable is a possible word. Tense vowels, however, allow an open syllable (*zee*), and, combined with the tendency of English to prefer maximal syllable onsets [5], this might lead to the segmentation *zee-theager*. The target *eager* could therefore be as hard to spot in the tense vowel contexts as in the consonant context, since in both cases there is a single consonant between the beginning of the target and a likely word boundary (cued by the syllable boundary in *zee-theager* and by the silence in *theager*).

2.1. Methods

2.1.1. Subjects

36 native speakers of English were paid for their participation.

2.1.2. Design and Procedure

Forty-eight bisyllabic WS words (*canal*) and 30 bisyllabic SW words (*eager*) were selected; none had other words embedded within them. The first syllables of the WS words consisted of a single consonant followed by schwa; the SW words all began with vowels. Twenty-four of the WS words were placed in three preceding contexts: a single consonant (*scanal*); an open CV syllable with a lax vowel (*zEcanal*); and an open CV syllable with a tense vowel (*zeecanal*). It was not possible to find consonant contexts for the other 24 WS words (no phonotactically legal clusters could be formed with words beginning with voiced consonants, like *giraffe*, or those beginning with /s/, like *cigar*). These words were therefore only paired with tense and lax CV contexts. The SW words were also placed in three preceding contexts: a single consonant (*theager*); a closed CVC syllable with a lax vowel (*zEtheager*); and a closed CVC syllable with a tense vowel (*zeetheager*). In all complete strings, the only embedded real word was the intended target word.

The target-bearing items were divided over three lists, such that all of the SW words, and the 24 WS words which had three contexts, appeared on all three lists, with type of context counterbalanced over lists. The remaining target-bearing items (WS words with only two contexts) were also divided over the three lists; 16 of these words appeared in each list, each word appearing in only one context in a given list, with type of context counterbalanced over lists. Each list therefore contained 70 target-bearing items. A further 140 filler items containing no real English words were constructed. The fillers matched the target-bearing items in length and stress patterns; there were twice as many fillers with a particular number of syllables and stress pattern as there were target-bearing items with that structure. Each list contained all fillers, with target-bearing and filler items in pseudorandom order, such that there was always at least one filler between any two target-bearing items.

The materials were recorded by a native speaker of English in a sound-damped booth. The speaker attempted to minimize syllabification cues in the recording; medial consonants (/θ/ in *zEtheager* and *zeetheager*; /k/ in *zEcanal* and *zeecanal*) were ambisyllabic, that is, were neither clearly syllabified in the first syllable nor in the second syllable. Listeners were tested individually in a quiet room; they heard the lists over headphones. They were told they would hear nonsense words, some of which would contain real English words. They were asked to press a button as fast as possible whenever they spotted a real word, and to say aloud the word that they had spotted. Reaction Times (RTs) were measured from target-bearing item onset, and adjusted by subtracting item durations to give RTs from target-word offsets. Each listener heard a practice list, followed by one of the three experimental lists.

2.2. Results and Discussion

Analyses of Variance (ANOVAs) were performed on the RT and error data. An item was excluded from an analysis if, in any one condition in that analysis, it was missed by more than two thirds of the subjects who heard it.

In the analysis of the data summarized in Table 1, the effect of context was significant in the RT analysis ($F1(2,60) = 19.38$, $p<.001$; $F2(2,86) = 10.89$, $p<.001$) and in the error analysis ($F1(2,60) = 15.31$, $p<.001$; $F2(2,86) = 12.42$, $p<.001$). No other effects were fully reliable in either analysis. Planned comparisons between the three contexts for each type of word were then carried out. Responses to WS words like *canal* were faster ($t1(35) = 2.12$, $p<.05$; $t2(19) = 2.37$, $p<.05$) and more accurate ($t1(35) = 5.36$, $p<.001$; $t2(19) = 4.18$, $p<.005$) in the lax-vowel syllable contexts than in the consonant contexts. This result suggests that the PWC is a language-universal mechanism: CV syllables with lax vowels are not treated as impossible residues in English segmentation, like single consonants are, in spite of the fact that such syllables are not possible English words.

Responses to words like *canal* were also faster ($t1(35) = 3.79$, $p<.005$; $t2(19) = 3.85$, $p<.005$) and more accurate ($t1(35) = 3.51$, $p<.005$; $t2(19) = 2.51$, $p<.05$) in the tense-vowel syllable contexts than in the consonant contexts. This result replicates the finding that words are easier to spot in syllabic contexts than in consonantal contexts, as predicted by the PWC. Listeners were also

Table 1. Mean Reaction Times (RTs, in milliseconds, measured from target offset) and mean percentage error rates, Experiment 1.

	Weak-Strong Target Contexts		
	Tense Vowel CV Syllable	Lax Vowel CV Syllable	Single Consonant
Mean RT:	388	446	501
Mean Error:	13%	10%	27%
Example:	zeecanal	zEcanal	scanal

	Strong-Weak Target Contexts		
	Tense Vowel CVC Syllable	Lax Vowel CVC Syllable	Single Consonant
Mean RT:	511	466	607
Mean Error:	14%	10%	20%
Example:	zeetheager	zEtheager	theager

faster to detect WS words in syllable contexts with tense vowels than in syllable contexts with lax vowels ($t1(35) = 2.61$, $p<.05$; $t2(19) = 2.34$, $p<.05$). Note however that there was a small speed-accuracy trade-off in the data: listeners were slightly more accurate in detecting WS words in syllable contexts with lax vowels than in syllable contexts with tense vowels, though this difference was not significant. Note also that in a second analysis, where all the words which appeared in tense and lax syllable contexts were analysed (i.e., the words in the previous analysis plus those words like *giraffe* which appeared only in syllabic contexts), this difference was not significant (Means: tense vowel contexts, 423 ms, 14% errors; lax vowel contexts, 448 ms, 15% errors). There was also no difference in error rates between these two conditions in this analysis. It therefore appears that there was no robust difference between these conditions, while performance in both was reliably better than that in the consonant condition.

Responses to SW words like *eager* were faster ($t1(35) = 3.56$, $p<.005$; $t2(24) = 3.68$, $p<.005$) and more accurate ($t1(35) = 2.88$, $p<.01$; $t2(24) = 2.33$, $p<.05$) in the lax-vowel syllable contexts than in the consonant contexts. This difference is again as predicted by the PWC, and replicates Norris et al. [1]. No other differences within the SW words were fully reliable. This means that while listeners were not reliably slower or less accurate in detecting SW words in tense-vowel syllable contexts than in lax-vowel syllable contexts, they were also not reliably faster or more accurate in this condition than in the consonant context condition. This suggests that there was some tendency for listeners to segment strings like *zeetheager* as *zee-theager*, thus tending to make detection of *eager* as hard as in *theager*. But, since the tense-vowel condition was also not reliably different from the lax-vowel condition, this tendency was not very strong. Since contexts like *zEth* and *zeeth* are both possible words, there is no clear difference between these two conditions.

The principal result of Experiment 1 is clear. Listeners were able to spot words like *canal* faster in CV syllable contexts with lax vowels than in single consonant contexts. This suggests that the PWC operates according to language-universal principles. Contexts which are possible words in some languages (CVs with lax vowels) should therefore be treated as acceptable residues in on-line speech segmentation in any language.

3. EXPERIMENT 2

If the PWC really is determined by a universal rather than language-specific notion of possible word, then we should also expect to see similar results with weak syllables. Weak syllables should behave just like syllables with full vowels. That is, word-spotting should be much easier when the residue constitutes a weak syllable than when it is a consonant, even though weak syllables cannot be content words in English.

3.1. Methods

3.1.1. Subjects

Twenty four native speakers of English were paid for their participation.

3.1.2. Design and Procedure

The stimuli used in Experiment 2 were derived from the following-context materials in Norris et al. [1] by changing the vowels in their full-vowel syllabic contexts to schwa. So, for example, the target word *sea* could appear with either a following consonant context (*seash*) or a following weak syllable context (*seashəb*). In the case of the following syllable contexts only 11 of the 48 items retained exactly the same consonants (C*C) as in [1]. The remaining items were altered to avoid creating phonotactically illegal strings or strings that could be misheard as words, and to increase the variety of contexts. There were 110 filler items many of which had weak final syllables so that a final weak syllable was not a cue to the presence of an embedded word. There were also 8 filler target words with following full syllables. As in [1], half the target words were monosyllabic and half were bisyllabic. Target words only appeared with following contexts. The procedure was otherwise identical to that of Experiment 1.

3.2 Results and Discussion

ANOVAs were performed on the latency and accuracy data. Four words were excluded from the analysis because they were missed by more than two thirds of the subjects who heard them in either consonant or syllable contexts. In the analysis of the data summarized in Table 2, the effect of context was significant by subjects in the RT analysis ($F1(1,22) = 5.71$, $p< .03$; $F2(1,40) = 2.07$, $p = .16$) and by both subjects and items in the error analysis ($F1(1,22) = 16.21$, $p<.001$; $F2(1,40) = 17.20$, $p<.001$). The effect of number of syllables was significant in the RT analysis ($F1(1,22) = 9.35$, $p < .01$;

Table 2. Mean Reaction Times (RTs, in milliseconds, measured from target offset) and mean percentage error rates, Experiment 2.

Target:	Monosyllabic		Bisyllabic	
Context:	CəC	C	CəC	C
Mean RT:	890	1001	789	866
Mean Error:	19%	22%	9%	32%
Example:	seashəb	seash	sugarməl	sugarm

F2(1,40) = 5.72 p<.05) but not in the error analysis (Fs < 1). There was also a significant interaction between context and number of syllables in the error analysis (F1(1,22) = 18.24, p<.001; F2(1,40) = 11.27, p<.002) but not in the RT analysis (Fs < 1). Although the error rates are lower than in Norris et al. [1], the overall pattern, including the fact that the context effect in errors was larger in bisyllables, is very similar to the corresponding consonant and full-syllable conditions in their Experiment 1. In that experiment the overall context effect was 45ms in RTs and 15% in errors, compared with 94ms and 13% here.

The results of Experiment 2 are very straightforward: word-spotting is easier in weak syllable contexts than in consonant contexts. Furthermore, this difference is, if anything, marginally greater than the difference between the corresponding full-syllable and consonant contexts in Norris et al. [1]. There is therefore no suggestion that weak syllables violate the PWC.

4. DISCUSSION

From the perspective of the PWC, CV syllables with lax vowels or syllables with schwa appear to be treated in exactly the same way as syllables with full vowels, despite the fact that the former are not well-formed content words in English. In other words, what drives the PWC is not an abstract phonological constraint on the form of words acceptable in a particular language. It appears that what constitutes a viable residue in determining an acceptable parse of continuous speech is a syllable, and any syllable will do.

One might be concerned that we are trying to make a case for a language-universal strategy based only on data from a single language. However, further evidence that the PWC is indeed a language-universal strategy comes from a word-spotting experiment in Sesotho, a Bantu language spoken in Southern Africa. In Sesotho, any surface realization of a content word must have at least two syllables. Cutler, Demuth and McQueen [6] asked Sesotho listeners to spot words like *alafa* (to prescribe) in *halafa* (where the single consonant context /h/ is an impossible word) and *roalafa* (where the monosyllabic context *ro* is not a well-formed Sesotho word). Listeners spotted words slower and less accurately in the consonantal contexts than in the monosyllabic contexts. Even though *ro* is not a possible word in Sesotho, this does not make it an unacceptable residue in Sesotho speech segmentation. McQueen and Cutler [7] have also found that Dutch listeners find it harder to spot words in preceding consonantal contexts (e.g. *lepel*, spoon, in *blepel*) than in preceding CV contexts with schwa (*səlepel*). As in English, weak syllables are not possible content words in Dutch.

We began by asking whether the PWC is a language-specific or language-universal constraint on speech segmentation. The original study by Norris et al. [1] compared consonant residues with syllable residues. This left open the possibility that the critical unit determining the viability of a parse might be either the minimal phonological word in the language or the syllable. The experiments reported here and in [6] and [7] provide a clear answer to this question. Segmentation is impaired when the residue between the end of a candidate word and the nearest known boundary is a consonant, but not when it is a syllable, regardless of whether the syllable is a possible word in the language. The simulations reported in [1] used a modified version of the Shortlist model [8]. The algorithm used by the model was to penalise any candidate where there was not a vowel between the end of that candidate and the nearest known boundary. This language-universal algorithm appears to be the correct characterisation of the PWC.

5. ACKNOWLEDGEMENTS

Corresponding author: D. Norris, MRC Cognition and Brain Science Unit, 15 Chaucer Road, Cambridge, CB2 2EF, U.K. Email: dennis.norris@mrc-cbu.cam.ac.uk

6. REFERENCES

[1] Norris, D, McQueen, J, Cutler, A & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, **34**, 193-243.

[2] Cutler, A., Dahan, D. & Donselaar, W. van (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech,* **40**, 141-201.

[3] McQueen, J.M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, **39**, 21-46.

[4] Weber, A. (2000). The role of phonotactics in the segmentation of native and non-native continuous speech. *Proceedings of SWAP*, Nijmegen, The Netherlands (pp. 143-146). Nijmegen: MPI for Psycholinguistics.

[5] Selkirk, E.O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

[6] Cutler, A., Demuth, K. & McQueen, J.M. (in preparation). Universality versus language-specificity in listening to speech.

[7] McQueen, J.M. & Cutler, A. (1998). Spotting (different types of) words in (different types of) context. *Proceedings of ICSLP 98*, Sydney, Australia (pp. 2791-2794). Rundle Mall: Causal Productions.

[8] Norris, D. (1994). Shortlist: A hybrid connectionist model of continuous speech recognition. *Cognition*, **52**, 189-234.