THE ROLE OF PHONOTACTICS IN THE SEGMENTATION OF NATIVE AND NON-NATIVE CONTINUOUS SPEECH

Andrea Weber
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ABSTRACT

Previous research has shown that listeners make use of their knowledge of phonotactic constraints to segment speech into individual words. The present study investigates the influence of phonotactics when segmenting a non-native language. German and English listeners detected embedded English words in nonsense sequences. German listeners also had knowledge of English, but English listeners had no knowledge of German. Word onsets were either aligned with a syllable boundary or not, according to the phonotactics of the two languages. Words aligned with either German or English phonotactic boundaries were easier for German listeners to detect than words without such alignment. Responses of English listeners were influenced primarily by English phonotactic alignment. The results suggest that both native and non-native phonotactic constraints influence lexical segmentation of a non-native, but familiar, language.

1. INTRODUCTION

Understanding spoken language requires the isolation of individual words from a continuous speech stream. This lexical segmentation problem can be solved by competition between a set of candidate words. Only those candidates which provide an optimal parse of the input win the competition [1]. But competition alone does not account for the evidence that listeners can use acoustic and phonological cues, when available, to help solve the segmentation problem.

Quené has shown that Dutch listeners can use durational cues to word boundaries when asked to choose between two alternative readings of an ambiguous two-word utterance [2]. Language-specific metrical information also appears to provide listeners with important segmentation cues. In a stress-timed language like English, the majority of content words begin with strong syllables [3]. Cutler and Norris [4] showed that English listeners use this metrical information for segmentation by inferring a word boundary at the onset of strong but not weak syllables. English listeners found it easier to spot a word like *mint* in a nonsense sequence consisting of a strong syllable followed by a weak syllable, like /mɪntəv/, than in two strong syllables, like /mɪnteɪv/. The use of language-specific metrical information for segmentation does not contradict the importance of competition. The marking of any likely word boundary by multiple cues, such as metrical cues, instead modulates the competition.

Phonotactic constraints (restrictions on permissible phoneme sequences within syllables) are apparently another source of information used for segmentation. The knowledge that a language does not allow certain consonant sequences as word onsets, for example, could be used to infer a potential word boundary between such consonants. Dutch listeners indeed find it easier to detect words in nonsense sequences when the word onsets are aligned with a phonotactic boundary (e.g., *rok*, ‘skirt,’ in /fim.rɒk/) than when they were misaligned (e.g., *rok* in /fi.drɒk/; [5]). /fim.rɒk/ requires a syllable boundary between /m/ and /r/, since /mr/ is not permissible within syllables in Dutch. This leaves the onset of *rok* aligned. /fi.drɒk/, however, requires a syllable boundary between /i/ and /d/ according to Dutch phonotactics, leaving the onset of *rok* misaligned in /drɒk/, and hence harder to detect.

Adult listeners are known to be language-specific listeners. Previous research has shown that there is a profound influence of the native language on both native and non-native speech perception (as discussed in [6], among many other works). In the case of metrical segmentation, for instance, it has been shown that French listeners show sensitivity to syllabic structure not only in their native language (French is a syllable-timed language), but also when listening to English. English listeners do not show that sensitivity, either when listening to their native language or when listening to French (English is a stress-timed language; [7]).

Altenberg and Cairns [8] tested the influence of phonotactic constraints on the processing of visually presented words in English monolinguals and English-German bilinguals. When asked to decide whether a visually presented item was an English word or not (lexical decision), the bilinguals did not show the same pattern of response times (RTs) as the monolinguals did. Bilinguals were affected in their decisions by the phonotactic legality of the stimuli in both German and English. The present study concentrates on the segmentation problem in spoken non-native language. It addresses the question of whether both native and non-native phonotactic constraints influence the segmentation of a non-native language.

German and English listeners were presented with English speech stimuli. The German listeners were rather proficient in English, while the English listeners had no experience with German. For both groups, the task was to detect embedded English words in nonsense sequences (word spotting). The onset of the embedded words was either aligned with a clear syllable boundary or not. There were four different ways in which the phonotactics of the two languages could mark a syllable boundary. In one condition, both German and English require a syllable boundary at the onset of the word. In /punlʌk/ for example, the onset of the embedded word *luck* is aligned since both languages force a syllable boundary between /n/ and /l/. Words cannot begin with /nl-/ in either language. In another condition, German, but not English, requires a syllable boundary at the onset of the word, as for example in *luck* embedded in /mɔɪslʌk/. Words cannot begin with /sl-/ in German, but can in English. In a third condition, English, but not German, requires a syllable boundary at the onset of the word, as for example in /gɑɹʃlʌk/. Whereas /ʃl/ is not a possible syllable onset in English, it is one in German. In the fourth condition, neither language forces a syllable boundary at the word onset, as for example in /mɑɹflʌk/. /fl/ is a possible syllable onset in both German and English.

Since word spotting in a non-native language requires high proficiency in the non-native language, it was predicted that English as well as German phonotactics would influence the German listeners' segmentation of English, though to a lesser degree. Words that were not aligned with a clear syllable boundary according to either German or English phonotactics were predicted to be the hardest for German listeners to spot (e.g., *luck* in /mɑɹf<u>lʌk</u>/); with words which were aligned according to English but not to German (e.g., /gɑɹʃ<u>lʌk</u>/) somewhat easier, words aligned according to German but not English (e.g., /mɔɪs<u>lʌk</u>/) even easier, and words aligned according to both languages (e.g., /pun<u>lʌk</u>/) the easiest. For the English listeners, who had no knowledge of German, only the English phonotactics should influence the segmentation process. Embedded words were predicted to be easy to detect when the onset was aligned with a clear syllable boundary according to English phonotactics (e.g., /pun<u>lʌk</u>/ and /gɑɹʃ<u>lʌk</u>/) or hard when not (e.g., /mɑɹf<u>lʌk</u>/ and /mɔɪs<u>lʌk</u>/).

The present study contrasted the detection of embedded words when the word onset was aligned versus when it was not clearly aligned. The reader should note that 'not clearly aligned' as used here is different from the condition McQueen called 'misalignment' [5]. The sequence /fl/ for example allows both the syllabifications /mɑɹ.flʌk/ and /mɑɹf.lʌk/. Therefore in the present study the manipulation was 'clear alignment' versus 'no clear alignment' rather than 'alignment' versus 'misalignment.' English words with initial /l/ and initial /w/ were chosen for the experiment. Whereas the lateral /l/ belongs to the phoneme inventories of both languages, the approximant /w/ does not occur in German. This difference might affect processing the non-native language.

2. METHODS

2.1. Subjects

Forty-eight native speakers of German, students of English translation and interpretation at the University of Heidelberg, were paid to take part in the experiment. Another forty-eight native speakers of American English, students at the University of South Florida, were tested. Students of USF could choose to receive either monetary compensation or extra credit points for their participation. They had no knowledge of German.

2.2. Materials and Procedure

68 mono- and bisyllabic English words with initial /l/ (e.g. *luck*) or /w/ (e.g. *weapon*) were selected as target words. Each target word was appended to four different English nonsense monosyllables in order to create the four alignment conditions. The final consonant of the nonsense syllable determined whether the onset of the target word was aligned with a clear syllable boundary or not. Different final consonants were used for the nonsense syllables within an alignment condition where it was possible. Each target-bearing nonsense sequence contained its target word in final position but no other English or German embedded words. In addition there were 55 filler nonsense sequences which contained embedded English words in final position with an initial consonant other than /l/ or /w/. A further 251 bi- and trisyllabic nonsense sequences contained no embedded English or German words. Four lists were constructed. Each list contained all 306 filler sequences and 68 target-bearing sequences, in a pseudo-random order, such that before each target-bearing sequence there was at least one filler that contained no embeddings. The fillers appeared in the same sequential position in all four lists. Each target also appeared in the same sequential position, but in only one of its four possible contexts in any given list. Each list contained all four types of target-bearing sequences. 14 more representative practice items were added to the lists.

All materials were recorded onto DAT tape in a sound-proof booth by a female native speaker of American English. The speaker was instructed to avoid any clear syllable boundaries in her production. The materials were transferred to a computer, and durations were measured using the Xwaves speech editor. Items were presented in the list orders using a portable computer and the NESU experiment control software. Subjects were instructed to listen to the nonsense sequences and press the button in front of them as fast as possible if they detected an embedded English word at the end of one of the nonsense sequences. They then had to say the word aloud. The computer timed and stored manual responses, and oral responses were recorded on tape. Each subject heard the 14 practice stimuli first, followed by one of the four experimental lists. Prior to statistical analyses, RTs were adjusted so as to measure from the offset of the target words.

3. RESULTS

Missed manual responses and manual responses that were accompanied either by no oral response or by a word other than the intended target word, as well as RTs outside the range of -200 to 2000 ms, were treated as errors. Seven target words with particularly high error rates were excluded from the analysis, leaving 61 words for the analysis. Mean RTs and error rates are given below in Figure 1. Analyses of Variance with both subjects (F1) and items (F2) as the repeated measures were performed.

TABLE 1

Mean RTs in ms, Measured from Target Offset, and Mean Percentage Errors. G=German, E=English

Measure	Subject group	Not align. in E/G	Align. in E	Align. in G	Align. in E/G
RT	German	672	626	608	572
	English	555	513	544	468
Errors	German	33%	23%	24%	18%
	English	25%	17%	21%	13%

A five factor mixed ANOVA was used, with language of the listener and experimental list as the between subjects factors, and initial sound (/l/ or /w/) and German and English phonotactics (each with the two levels 'clearly aligned' and 'not clearly aligned') as the repeated measures factors.[1] The four-way interaction between English phonotactics, German phonotactics, language of the listener, and initial sound was significant by subjects and items ($F1(1, 88) = 6.23$, $p < .02$; $F2(1, 53) = 5.95$, $p < .02$).

ANOVAs were then performed separately for target words with initial /w/ and /l/. Mean RTs and error rates appear in Figure 2. For RTs to words with initial /w/ the interaction between German phonotactics, English phonotactics and language was significant ($F1(1, 88) = 6.00$, $p < .02$; $F2(1, 23) = 9.51$, $p = .005$). German phonotactics significantly influenced RTs of German listeners when responding to words with initial /w/ ($F1(1, 44) = 17.92$, $p < .001$; $F2(1, 23) = 4.74$, $p = .04$), as did English phonotactics, though English phonotactics reached significance only by subjects ($F1(1, 44) = 9.30$, $p = .004$; $F2(1, 23) = 2.29$, $p > .1$). There was no interaction between the constraints of the two languages for German listeners. English listeners did not show an effect of German phonotactics when responding to words with initial /w/, but did show an effect of English phonotactics. The interaction between German and English phonotactics reached significance by items for English listeners ($F1(1, 44) = 3.77$, $p > .05$; $F2(1, 23) = 8.52$, $p = .008$). Analyses of errors for words with initial /w/ revealed effects very similar to the RT results. The results show clearly that both native and non-native phonotactic constraints influenced lexical segmentation of the non-native language, whereas the English speakers' segmentation was only influenced by native phonotactic constraints.

Whereas the pattern of RTs to words with initial /w/ matches the predictions for listeners of both languages, the pattern of words with initial /l/ shows weaker effects for German listeners and an unpredicted influence of German phonotactics for English listeners. In consequence no interaction was found between German phonotactics, English phonotactics and language ($F1$ & $F2 < 1$). Both German and English phonotactics were significant (German phonotactics: $F1(1, 88) = 18.80$, $p < .001$; $F2(1, 30) = 14.18$, $p = .001$; English phonotactics: $F1(1, 88) = 10.46$, $p = .002$; $F2(1, 30) = 6.75$, $p < .02$). There was no interaction between German and English phonotactics. Analyses of errors for words with initial /l/ again revealed very similar results.

[1] The factor initial sound is only repeated measures for the subjects analysis.

4. DISCUSSION

Earlier studies have shown that the process of spoken word segmentation is influenced by language-specific phonotactic constraints. Dutch listeners find it easier to detect words in nonsense sequences when the words are aligned with a phonotactic boundary than when they are misaligned [5]. The results of the present study support the claim that the legality of phoneme sequences is used to help solve the segmentation problem. The present data also provide further support for the influence of the native language on the segmentation process of a non-native language. Cutler et al. have shown that the process of segmentation in a non-native language is influenced by metrical cues from the native language [7]. Both English and French listeners show sensitivity to specific native metrical cues in speech not only when listening to their native language but also when listening to a non-native language. But whereas tests of metrical segmentation only show the influence of native or non-native cues, testing phonotactic constraints allows a comparison of the influence of both native and non-native cues. A phoneme sequence can either provide the same phonotactic cue for two languages or it can provide different cues for two languages. Therefore the present study allowed for comparison of the degree to which both native and non-native phonotactic constraints influence the segmentation of a non-native language. The results demonstrate a clear influence of both native and non-native phonotactic constraints on the segmentation of a non-native language.

German subjects, when asked to spot English embedded words in nonsense sequences, found it easier to spot words that were aligned with either a German or English phonotactic boundary than words that lacked such clear alignment. Word spotting in a non-native language requires high proficiency in the non-native language. Accordingly German listeners' responses showed not only an influence of native but also of non-native phonotactics.

TABLE 2

Mean RTs in ms, Measured from Target Offset, and Mean Percentage Errors. G=German, E=English

Measure	Initial sound	Subject group	Not align. in E/G /rukwɛpən/ /maɹflʌk/	Align. in E /jiʃwɛpən/ /gaɹʃlʌk/	Align. in G /mɔɪtwɛpən/ /mɔɪslʌk/	Align. in E/G /jaɹlwɛpən/ /punlʌk/
RT	/□/	German	678	589	563	529
		English	522	468	590	452
Errors	/□/	German	32%	20%	21%	18%
		English	24%	17%	24%	14%
RT	/□/	German	675	656	650	602
		English	590	541	510	480
Errors	/□/	German	33%	24%	25%	17%
		English	26%	17%	17%	11%

However, the influence of native phonotactic constraints on segmentation was stronger than the influence of non-native phonotactic constraints. For German listeners, in the two conditions where the phonotactics of the two languages conflict, it was easier to spot words for which the German phonotactics provided a word boundary cue than those for which only English phonotactics provided a cue.

Responses of English listeners, who were performing in their native language, were influenced by English phonotactic alignment. Since the English listeners had no knowledge of German, German phonotactic boundaries should not influence their segmentation strategies. This is indeed the result found for English listeners responding to words with initial /w/, but these listeners' results for words with initial /l/ are not clear. Additional acoustic factors may have been involved for some of the /l/-initial conditions.

In conclusion, the results of the present study confirm previous findings that listeners use phonotactic constraints to identify likely word boundaries in continuous speech. This study also confirms that the process of listening is language-specific. Furthermore, this study shows that phonotactic constraints of a non-native language also influence listeners' segmentation strategies, although to a lesser degree than native phonotactic constraints. Thus, listeners do not only make use of phonotactic information about their own language. Although listeners are most strongly influenced by information about their own language, they can also use information they have learned about another language, in which they are proficient but not native, in order to segment speech of that language.

5. ACKNOWLEDGEMENTS

This research was supported by the Max Planck Institute for Psycholinguistics and by NIDCD 00323. I thank Anne Cutler, James McQueen and Natasha Warner for their comments on this work. Please address all correspondence to: Andrea Weber, Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands. Email: andrea.weber@mpi.nl.

6. REFERENCES

[1] McQueen, J.M., Norris, D.G. & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 621-638.

[2] Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, **20**, 331-350.

[3] Cutler, A. & Carter, D.M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, **2**, 133-142.

[4] 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.

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

[6] Cutler, A, Mehler, J., Norris, D.G. & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, **24**, 381-410.

[7] Cutler, A., Mehler, J., Norris, D.G. & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25**, 385-400.

[8] Altenberg, E. & Cairns, H. (1983). The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects. *Journal of Verbal Learning and Verbal Behavior*, **22**, 174-188.