SEQUENTIAL PROBABILITIES AS A CUE FOR SEGMENTATION

Arie H. van der Lugt Max Planck Institute for Psycholinguistics, Wundtlaan1, 6525 XD Nijmegen, The Netherlands Tel. +31243521911, FAX +31243521213, E-mail: vdlugt@mpi.nl

ABSTRACT

A large amount of psycholinguistic research, phonetic research and research in speech technology has been dedicated to the problem of segmentation: how is speech segmented into words? The work reported here extends earlier findings by McQueen & Cox ([1]), who found that phonotactics are used by listeners as a cue to the location of word boundaries. The present investigation addresses the question of whether people can also use less extreme sequential probabilities as a segmentation cue. Hearing a combination of sounds that often occurs at the end of a word or syllable may facilitate recognition of a following word; hearing a combination of sounds that occurs often at the beginning of a word or syllable may facilitate recognition of a preceding word. In a word-spotting task some indications were found that people are sensitive to sequential probabilities. However, no effects were found that strongly support the hypothesis that people do indeed use these distributional properties of the lexicon in the segmentation of spoken language.

1. INTRODUCTION

In written language spaces unambiguously indicate word boundaries, but in speech such a clear demarcation does not exist. Nevertheless, listeners hear speech as a series of words. How is speech segmented? Possible boundary cues are acoustic-phonetic cues, metrical cues and phonotactics. In the literature we find evidence for all these boundary cues, and although these cues are not always available, listeners seem to use them when they do occur. Acoustic-phonetic cues in English are, for instance, lengthening of onset syllables and segments ([2], [3]) and aspiration of word-initial stops ([4]). A large amount of crosslinguistic research has been dedicated to the study of metrical cues for segmentation; listeners are able to use the dominant rhythmic structure of their native language to help locate word boundaries ([5] for English; [6] for Dutch; [7] & [8] for French; [9] for Japanese).

Another possible cue for word boundaries is provided by the phonotactic permissibility of sound sequences; if people perceive a combination of sounds that isn't allowed within a syllable or word in their language, they may assume a boundary between those sounds. McQueen and Cox ([1]) used a word-spotting task ([5]) to investigate the role in segmentation of syllable boundaries which are mandatory on phonotactic grounds. They found that detection of the Dutch word 'rok', skirt, is much harder in *fie-drok*, where the target is misaligned with the syllable boundary of the bisyllable nonword (voiced stops like /d/ must be syllable-initial in Dutch), than in *fiem-rok*, where the target is aligned with the syllable boundary forced by the phonotactic rule that /mr/ is an illegal cluster in Dutch.

If listeners can judge the legality of a perceived combination of sounds in their native language and hypothesize possible word boundaries in continuous speech according to those phonotactics, perhaps they are also able to use less extreme sequential probabilities of their language. In Dutch there are lots of words beginning with the cluster /sp/, but only very few words starting /ks/. Conversely there are lots of words that end with the cluster /ts/, while there are very few words ending /sk/. Do people use these distributional properties of the lexicon in the segmentation of continuous speech? This is the general question addressed in the present investigation. Hearing a combination of sounds that often occurs immediately before a syllable or word boundary may facilitate recognition of a following word but may interfere with recognition of a preceding word; hearing a combination of sounds that occurs often immediately after a syllable or word boundary may facilitate recognition of a preceding word but may interfere with recognition of a following word. A word-spotting experiment was designed to test this hypothesis. Will it be easier to detect the Dutch word 'bloem' (flower) in bloem-spuum than in *bloem-ksuum*? Will it be harder to spot 'vriend' (friend) in duusk-vriend than in duuts-vriend?

2. METHOD

2.1 Materials

Six clusters (/sk/, /sp/, /st/, /ks/, /ps/ and /ts/) from the full set of 21 types that occur both as an onset and as a coda in Dutch, were selected. The six clusters are optimal with respect to two important properties: firstly, they are minimally different with respect to acoustic features and secondly, they cover the entire range of onset/coda ratios that were found in an analysis of the Celex computerised database of Dutch ([10]).



Figure 1 Relative frequency of occurrence of consonants and consonant clusters as onsets and codas of syllables and words. Onset/coda ratios are given on a logarithmic scale by type. An onset/coda ratio larger than one means that a consonant (cluster) is onset dominant, a ratio smaller than one means that a cluster is coda dominant.

All occurrences of consonants and consonant clusters as onsets and codas of syllables and of words were counted, both by type and by token. From this list the 21 types that can occur both as an onset and as a coda in Dutch were selected. For these 21 types the relative frequency of occurrence as onset versus coda was calculated, expressed as the ratio of the number of onset occurrences by the number of coda occurrences. Figure 1 shows these ratios, computed over syllables and words. The selected six consonant clusters range from onset dominant (e.g. the cluster /sk/ is 7.5 times more likely as an onset than as a coda) to coda dominant (e.g. the cluster /ks/ is 6.8 times more likely as a coda than as an onset).

Each cluster was placed in the onset of a nonsense syllable, after an initial target word (e.g. bloem, flower, in, for example *bloem-ksuum* or *bloem-skuum*) and in the coda of a nonsense syllable, before a final target word (e.g. vriend, friend, in, for example duuks-vriend or duusk-vriend). The target words were selected in a way that there was a phonotactically mandatory syllable boundary between the target word and the cluster as far as possible. To obtain a large enough set of materials some target contexts were used that do not form a phonotactically illegal sequence with the cluster or the adjacent phoneme of the cluster in Dutch; in those cases the cluster (or the phoneme of the cluster that was adjacent to the target) and the target context never occur in a single onset (for final targets) or coda (for initial targets) in Dutch.

All of the target words were paired with all of the six clusters. The six target-cluster pairs were then counterbalanced across six different versions of the experiment, so that each version contained a different cluster in the context of each particular target word. Each version therefore contained a total of 96 target-bearing bisyllables (48 initial and 48 final, 8 for each cluster) and 192 bisyllabic fillers with no embedded words.

2.2 Recording

The items were recorded onto DAT tape in a sound attenuated booth. All items were spoken three times in random order by a female native speaker of Dutch. After the recording all items were redigitized onto a computer. All the materials were measured and spliced into separate speechfiles, using the Xwaves speech editor. For each item the most natural utterance was selected. The target items were inspected to make sure that the target-bearing bisyllables were all syllabified correctly between target word and cluster. The individual speechfiles, one for each nonsense bisyllable were then transferred to the hard disk of a personal computer for use in the experiment.

2.3 Subjects and Procedure

The presentation of the stimuli, the timing of the manual responses and the collection of the data were performed by the NESU experimental system. developed at the M.P.I. for Psycholinguistics. Subjects were tested two at a time in individual sound attenuated booths. Seventy two subjects participated. Each of the six different versions of the experiment was heard by



Figure 2 Mean reaction times for the six clusters in initial and final position

twelve subjects. They were asked to press a response button with their preferred hand whenever they heard a real word embedded in one of the bisyllables, and to say which word they had heard. Their answers were monitored on-line for false alarms (trials where they spotted a word other than the intended target word).

3. RESULTS AND DISCUSSION

All manual responses that were accompanied by an incorrect spoken response were set to zero and treated as missing responses, just like the trials where subjects failed to respond entirely. Subjects did not respond correctly in 20.8% of the trials with target-bearing bisyllables; they detected words that were not the intended targets in 2.9% of trials. A total of 15 items was excluded from further statistical analysis, because subjects failed to respond correctly to those targets in more than 50% of the cases.

Reaction times were first measured from target word onset. Subjects responded faster to final targets (941 ms) than to initial targets (1115 ms). This is in line with earlier findings ([1]) that initial targets are harder to detect in a mixed design with both initial and final targets.

The RTs were submitted to Analyses of Variance (ANOVAs) with both subject (F1) and items (F2) as the repeated measure. No clear effects were found that support the hypothesis that word recognition in this task was influenced by the relative frequency of either

Proportion errors



Figure 3 Mean error rates for the six clusters in initial and final position

following or preceding clusters. Because the predicted effect of a certain cluster would be opposite in the two different target positions, the interaction of a cluster with target position is the source of variation that is relevant here (F1(5,330)=1.13, p>0.1; F2(5,405)=0.3, p>0.5; see Figure 2 for the mean reaction times in the subjects analysis). To exclude possible interfering effects of the duration of the target word the target durations were subtracted from the reaction times measured from target onset, thus obtaining reaction times from target offset, and the data were reanalysed. Again, the results show no reliable effects of the relative probability of occurrence of complex clusters on the detection latency of words (F1(5,330)=1.64, p>0.1; F2(5,405)=0.65, p>0.5).

Although no overall effects in the reaction times were significant, there were trends in the error data (see Figure 3 for the mean error rates in the subjects analysis). For example, the error analysis by subjects showed that subjects made most errors on initial targets that are followed by the cluster /ks/ (.30 relative to an overall mean for initial targets of .24), and subjects made most errors on final targets that are preceded by the cluster /sp/ (.26 relative to an overall mean for final targets of .18). These effects are in line with the predictions: /ks/ is the most coda-dominant cluster, and /sp/ is the most onset-dominant cluster. The overall results of the error analyses show a significant effect of the relative probability of occurrence of the consonant clusters in a certain position on the proportion of errors (F1(5,330)=6.62, p<0.01; F2(5,405)=2.77, p<0.05).

In this experiment one might expect that the conditions

with a /s/ directly preceding or following the target would be more difficult than the conditions with a stop (/k/,/p/,/t/) adjacent to the target. There would be two plausible explanations for such an effect. An acousticphonetic explanation would be where recognition of a target is facilitated when it is followed or preceded by a stop, because stops are acoustically more salient than fricatives. A second explanation would be a phonological account in the same direction: targets following or preceding an /s/ may be harder to recognise due to the role of the /s/ in complex consonant clusters; the /s/ has a high probability of occurrence both as an onset and as a coda. The data were reanalysed in terms of the phoneme next to the target, stop versus /s/. It was found that subjects' detection latencies did not differ significantly in these two conditions (F1(1,66)=1.77, p>0.1; F2(1,81)=.11, p<1).

4. CONCLUSION

There are two possible interpretations for these results. A strong interpretation of this experiment would be that people do not use sequential probabilities in the segmentation of speech. The other interpretation would be that the trends found in the error data are indicative of a weak sensitivity to sequential probabilities, and that the experiment lacked sufficient power for these effects to be observed clearly. A second experiment has therefore been designed, in order to increase power by controlling for some factors that were impossible to control for in the first experiment.

The predicted effects of the relative probability of a cluster as an onset or as a coda might have been obscured by effects of the absolute frequencies of occurrence in both positions. Moreover, the complexity of the syllable boundaries and, related to that, possible ambiguities in syllable boundaries because of the role of the segment /s/ in complex consonant clusters in Dutch (the /s/ has a high probability of occurrence in clusters, both as an onset and as a coda) may have also interfered with the predicted effects of the relative probability. In the follow-up experiment, the possible use of the absolute probabilities of CV and VC sequences as boundary cues will be studied in a word-spotting task similar to the one used in the first experiment. However, instead of a range of relative probabilities, two strictly bimodal sets of low versus high probability CV onsets (boom, tree, in, for example boom-douf versus boom-dif) and low versus high probability VC codas (veer, feather, in, for example buul-veer versus beel-veer) will be tested as possible boundary cues in the context of the same target word.

The results from this experiment suggest that people are sensitive to the distributional properties of the lexicon that were investigated, that is to the relative probability of occurrence of complex consonant clusters at the beginnings and ends of syllables. These effects were, however, very small, and the present results do not provide us with strong evidence that this kind of information is used by the listeners in the segmentation of spoken language.

5. ACKNOWLEDGEMENT

I would like to thank Rinske Goettsch for her help in the preparation of this experiment. Correspondence can be addressed to A.H. van der Lugt, MPI for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, or by email (vdlugt@mpi.nl).

6. REFERENCES

[1] J. M. McQueen, & E. Cox, "The use of phonotactic constraints in the segmentation of Dutch", Proc. EUROSPEECH'95, Vol. 3, 1707-1710, Madrid, 1995.

[2] D. W. Gow & P. C. Gordon, "Lexical and prelexical influences on word segmentation: Evidence from priming", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 21, 344-359, 1995.

[3] I. Lehiste, "The timing of utterances and linguistic boundaries", *Journal of the Acoustical Society of America*, Vol. 51, 2018-2024, 1972.

[4] I. Lehiste, "An acoustic-phonetic study of internal open juncture", *Phonetica*, Vol. 5 (Suppl. 5), 1-54, 1960.

[5] A. Cutler & D. Norris, "The role of strong syllables in segmentation for lexical access", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 14, 113-121, 1988.

[6] J. Vroomen & B. de Gelder, "Metrical segmentation and lexical inhibition in spoken word recognition", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 21, 98-108, 1995.

[7] J. Mehler, J.-Y. Dommergues, U. H. Frauenfelder & J. Segui, "The syllable's role in speech segmentation" *Journal of Verbal Learning and Verbal Behavior*, Vol. 20, 298-305, 1981.

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

[9] T. Otake, G. Hatano, A. Cutler & J. Mehler, "Mora or syllable? Speech segmentation in Japanese", *Journal of Memory and Language*, Vol. 32, 58-278, 1993

[10] G. Burnage, "CELEX: A guide for users", Centre for Lexical Information, Nijmegen, 1990.