SPOTTING (DIFFERENT TYPES OF) WORDS IN (DIFFERENT TYPES OF) CONTEXT

James M. McQueen and Anne Cutler

Max-Planck-Institute for Psycholinguistics
Nijmegen, The Netherlands

ABSTRACT

The results of a word-spotting experiment are presented in which Dutch listeners tried to spot different types of bisyllabic Dutch words embedded in different types of nonsense contexts. Embedded verbs were not reliably harder to spot than embedded nouns; this suggests that nouns and verbs are recognised via the same basic processes. Iambic words were no harder to spot than trochaic words, suggesting that trochaic words are not in principle easier to recognise than iambic words. Words were harder to spot in consonantal contexts (i.e., contexts which themselves could not be words) than in longer contexts which contained at least one vowel (i.e., contexts which, though not words, were possible words of Dutch). A control experiment showed that this difference was not due to acoustic differences between the words in each context. The results support the claim that spoken-word recognition is sensitive to the viability of sound sequences as possible words.

1. INTRODUCTION

Speech is continuous, and must be segmented into its component words to be understood. Psycholinguistics has developed a laboratory task which is especially designed to study the segmentation of spoken words from their immediate speech context: the word-spotting task [1]. In a word-spotting experiment, listeners do not know in advance what the input might be; they respond as soon as they hear a real word - any real word. Since this could arguably count as a characterisation of normal listening, the task offers a window on spoken-word recognition which is somewhat more realistic than many other laboratory tasks.

Word spotting has provided evidence that speech segmentation and recognition involve competition between candidate words [2,3]. Word spotting has also shown that segmentation is based on cues (e.g., metrical [4,5] and phonotactic [6]) to the location of likely word boundaries in the speech signal. Most recently, word-spotting studies have led to the discovery of the Possible Word Constraint (PWC), now proposed as a general mechanism by which segmentation and recognition in continuous speech is achieved [7]. The PWC computes the viability of sound sequences as possible words in the listener's language. Computer simulations have shown how the PWC could operate on the basis of knowledge of possible words, and the cues in the signal to likely word boundary locations, in the competition-based framework of the Shortlist model [8].

In the word-spotting experiments which motivated the PWC [7], English listeners found it harder to detect words in consonantal contexts (e.g., *egg* in *fegg*; *apple* in *fapple*) than in syllabic contexts (e.g., *egg* in *maffegg*; *apple* in *vuffapple*). This, according to the PWC, is because the stretch of speech between the word and the boundary cued by the silence preceding the string is an impossible word of English in the case of the single consonant (f in *fegg*) but a possible though non-existent word in the syllabic case (*maff* in *maffegg*). Thus in the implementation in Shortlist, the PWC acts to penalize the activation of the lexical candidate (*egg*) only in the impossible word context. Further support for the PWC has come from a Dutch word-spotting study [6]. Listeners found it easier to spot words such as *rok* (skirt) in *fim.rok*, where phonotactics signal a likely word-boundary at the onset of the target, than in *fi.drok*, where the phonotactic boundary is misaligned with the onset of the target by one consonant (i.e., by an impossible word). Again, in the computer implementation, the PWC penalizes the activation of the candidate word (*rok*) only when there is an impossible word between the phonotactic boundary and the word's onset.

In both of these studies, however, the target words represented only a small set of possible word types. In the English study [7], monosyllabic targets were compared with trochaic (Strong-Weak) bisyllabic targets. The Dutch study [6] used monosyllabic targets only. In both studies, as in all word-spotting experiments, the vast majority of targets were nouns. Thus both phonologically and syntactically there was relatively little variation.

The main aim of the present study was thus to test the generality of the PWC using a wider variety of target words both phonologically and syntactically. We sought to establish whether Dutch listeners, like English listeners, use the PWC in segmenting bisyllabic words from nonsense contexts. In a word-spotting task, target words were presented in impossible-word contexts (e.g., *lepel*, spoon, in *blepel*) and possible-word contexts (e.g., *kulepel*). In an extension of the English study, however, we contrasted trochaic with iambic (Weak-Strong) targets. In both English and Dutch, words beginning with strong syllables are more common than words beginning with weak syllables [9,10]. If the PWC is truly general however, it should apply both to targets beginning with weak syllables and to those beginning with strong syllables. Similarly, a general PWC should apply not just to the recognition of nouns (the most word-like words) but also to the recognition of words from other syntactic categories. We therefore tested not only noun targets in the present study, but also verb targets.

In addition, we aimed to examine exactly the question of what constitutes a possible-word context. Thus in a further extension of the previous work, strong (CV) and weak (Cə) monosyllabic and bisyllabic (CVCə) possible-word contexts were contrasted with single-consonant impossible-word contexts (C). This manipulation allowed us to test whether the PWC operates in graded or all or none fashion; that is, is one possible word as good as any other?

2. EXPERIMENT 1

2.1. Method

Subjects. Thirty-six members of the Max-Planck-Institute subject panel, native speakers of Dutch, were paid to take part.

Materials. Nine types of item were constructed, involving three types of bisyllabic word (24 Strong-Weak infinitive verbs, e.g., *wonen*, to live; 24 Strong-Weak nouns, e.g., *lepel*, spoon; and 24 Weak-Strong words of several syntactic categories, e.g., *begin*, begin) and four types of preceding nonsense context (Single consonant, C; Weak syllable, Cə; Strong syllable, CV; and Strong-Weak bisyllable CVCə). As shown in Table 1, each word type appeared in three contexts, allowing separate comparisons between noun and verb targets, between trochaic and iambic targets, and between longer (CVCə), shorter (CV and Cə) and impossible-word (C) contexts.

These materials were split into three counterbalanced subsets, for presentation to three groups of subjects. Each subject heard all 72 targets: eight of each of the three word types in each of the three contexts associated with that word type. Three lists were therefore made in which each set of target-bearing items was mixed with the same 144 filler nonwords, which did not contain embedded words. The fillers were constructed to match the target-bearing items, from nonsense contexts and following bisyllables: 96 trochaic "base" nonwords in four sets of 24 with the preceding contexts C, Cə, CV, or CVCə; and 48 iambic "base" nonwords, in three sets of 16 with the preceding contexts Cə, CV, or CVCə. The target-bearing and filler items were mixed in pseudo-random order. The order of fillers and targets was identical in all three lists; the lists varied only in the context in which a given target appeared.

Procedure. The materials were recorded by a female native speaker of Dutch. The primary stress on each target-bearing item was placed on the strong syllable of the target word. Each item was labelled and measured using the Xwaves speech editor. Word onsets were labelled at or near a zero-crossing closest to the onset of the first phoneme of each target word, as established by both visual and auditory criteria. NESU software controlled stimulus presentation and data collection. Items were presented over headphones, with an inter-item interval of 3.5 seconds. Listeners were tested individually or in pairs, in separate booths. They were asked to try to spot real words, which they were told would be embedded at the end of some of the nonsense words. They were asked to press a button as fast as possible if they spotted a word, and then to say aloud what that word was.

2.2. Results and Discussion

The oral responses of each subject were checked. All button-press responses associated with incorrect or missing oral responses were then treated as errors (a total of 6.6% of the data). Overall error-rates were then computed for each item, and all items which in any one condition were missed by 50% or more of the subjects who heard that item were then excluded. Eight items failed this criterion in Experiment 2: four which failed the criterion here, plus four more (see below). For all eight of these words, at least one token of that word in one context was clearly not easily recognisable. In the analyses reported here, therefore, all eight were excluded (two iambs, five trochaic verbs, and one trochaic noun). Other analyses, in which only the four items which failed the criterion in this experiment were excluded, found the same reliable effects as those given here. Raw Reaction Times (RTs), measured from item onset, were adjusted so as to measure from target word offset by subtracting the total duration of each item from each response to that item. The RTs and error rates were submitted to Analyses of Variance (ANOVAs). The mean RTs for correct responses and mean error rates are given in Table 1.

CONTEXT	C	Cə	CVCə	CV
VERBS (Trochaic)	dWONEN	keWONEN	dukeWONEN	
RT	739	413	432	
Error	16%	5%	8%	
NOUNS (Trochaic)	bLEPEL	seLEPEL		kuLEPEL
RT	667	380		435
Error	9%	4%		5%
IAMBIC WORDS		seBEGIN	zaseBEGIN	geeBEGIN
RT		390	360	419
Error		3%	2%	3%

Table 1: Mean RT (in msec measured from word offset) and mean error-rates (proportion of missed targets), Experiment 1.

Nouns and Verbs. The first set of analyses compared performance on the trochaic nouns and verbs in impossible-word (C) and possible-word (Cə) contexts. There was a highly significant effect of context in both the RTs ($F1(1,33) = 190.4$, $p<0.001$; $F2(1,40) = 60.9$, $p<0.001$) and errors ($F1(1,33) = 18.4$, $p<0.001$; $F2(1,40) = 17.1$, $p<0.001$): listeners were slower and less accurate in spotting words in impossible- than in possible-word contexts (average differences of 307 msec and 8%). There was a weak tendency for responses to be slower to verbs that to nouns (53 msec, on average: $F1(1,33) = 3.0$, $p=0.09$; $F2<1$), and a slightly stronger tendency for responses to be less accurate to verbs than to nouns (5%, on average: $F1(1,33) = 6.4$, $p<0.05$; $F2(1,40) = 3.6$, $p=0.07$). The interaction of context and word type was not significant in any of these analyses. Once again, therefore, listeners appear to be sensitive in on-line spoken-word recognition to the viability of sound sequences as possible words in their language. Although performance was somewhat poorer on verbs than on nouns, these differences were not reliable; furthermore, the impossible-context effect was equivalent for nouns and verbs. This suggests that nouns and verbs are recognized via the same processes.

Trochaic and Iambic Words. The second set of analyses compared performance on the iambic words and the trochaic verbs, in Cə and CVCə contexts. In both RTs and errors, there was an

effect of word type, significant only by subjects (RTs: $F1(1,33) = 5.1$, $p<0.05$; $F2(1,39) = 2.3$, $p>0.1$; Errors: $F1(1,33) = 7.9$, $p<0.01$; $F2(1,39) = 2.6$, $p>0.1$): Iambs were, on average, spotted 47 msec faster and 5% more accurately than trochaic verbs. There were no other significant effects in the RTs. In the errors, the only other significant effect (and that only by items) was the interaction of word type and context: $F1(1,33) = 2.7$, $p>0.1$; $F2(1,39) = 5.9$, $p<0.05$. Pairwise comparisons showed that this interaction was due to the 6% difference between trochaic verbs and iambic words in CVCə contexts (no other pairwise tests were significant). The third set of analyses compared performance on the iambic words and the trochaic nouns, in Cə and CV contexts. The only reliable effect was one of context in RTs. Subjects were, on average, 42 msec faster to spot words in Cə contexts than in CV contexts ($F1(1,33) = 6.5$, $p<0.05$; $F2(1,43) = 4.6$, $p<0.05$). The final analysis compared all three word types in the Cə context; there was no reliable effect of word type. There was thus no overall difference between trochaic and iambic target words.

Longer and Shorter Possible Word Contexts. In all of the above analyses there was only one reliable difference between different types of possible-word contexts. Although spotting words was as easy in longer contexts (CVCə) as in shorter contexts (Cə), listeners were faster to spot words in Cə contexts than in CV contexts.

There was one clear effect in Experiment 1: as predicted by the PWC, words were much harder to spot in contexts which themselves were impossible words than in contexts which, though not words, were possible words of Dutch. Because the materials were natural utterances, however, the target words were not acoustically identical across contexts. It was therefore possible that the difference observed between possible- and impossible-contexts (and indeed that between Cə and CV contexts) was due to differences between the targets, rather to the contexts. This concern was addressed in Experiment 2, in a way which is standard practice in word-spotting studies. The target words were excised from their contexts, and presented to listeners, together with nonwords excised from the Experiment 1 fillers, in a go/no-go lexical decision task. The listeners' task was to press a button every time they heard a real word. If the differences observed in word spotting were due to differences between the targets, then such differences should reappear in lexical decision.

3. EXPERIMENT 2

3.1. Method

Subjects. A further 36 members of the Max-Planck-Institute subject panel were paid to take part.

***M*aterials**. New versions of the three experimental lists from Experiment 1 were made by excising either the target words from the target-bearing items, or the "base" nonwords from the fillers. Excision was made at the labels already marked in the target-bearing items, or, for the fillers, at equivalent points at the onsets of the base nonwords. The result was an experiment with exactly the same materials, design and running order as Experiment 1, with the exception that each target word was presented without its context, and each filler, though still a nonword, was now a subcomponent of the original item.

Procedure. The only differences to the procedure of Experiment 1 were that listeners were told that they would hear a list of words and nonwords and that they were to press the button whenever they heard a real word. As in Experiment 1, they were asked to say the word aloud after they had pressed the button.

3.2. Results and Discussion

	TAKEN FROM CONTEXT			
	C	Cə	CVCə	CV
VERBS (Trochaic)	(d)WONEN	(ke)WONEN	(duke)WONEN	
RT	248	233	215	
Error	3%	3%	4%	
NOUNS (Trochaic)	(b)LEPEL	(se)LEPEL		(ku)LEPEL
RT	292	257		278
Error	4%	1%		4%
IAMBIC WORDS		(se)BEGIN	(zase)BEGIN	(gee)BEGIN
RT		218	240	264
Error		3%	4%	6%

Table 2: Mean RT (in msec measured from word offset) and mean error-rates (proportion of missed targets), Experiment 2.

Button-press responses associated with incorrect or missing oral responses were again treated as errors (6.9% of the data), and responses to any words which in any one condition were missed by 50% or more of the subjects who heard that item were ignored. As mentioned above, the data from eight items were excluded in this way. The duration of each item was subtracted from each appropriate raw RT, so as to measure from item offset. Table 2 shows the mean RTs for correct responses and error rates.

Parallel ANOVAs to those performed in Experiment 1 were carried out. In the first set of analyses, comparing trochaic nouns and verbs taken from possible (Cə) and impossible (C) contexts, there were no significant effects in the errors. In the RTs there were two weak effects: an effect of word type (responses to verbs a mean of 34 msec faster than those to nouns), significant by subjects but not items ($F1(1,33) = 8.8$, $p<0.01$; $F2(1,40) = 1.3$, $p>0.2$); and an effect of context (responses to items from C contexts a mean of 25 msec slower than those to items taken from Cə contexts), significant only by items ($F1(1,33) = 2.6$, $p>0.1$; $F2(1,40) = 4.5$, $p<0.05$). The weak and not fully reliable context effect suggests that at most a very small component of the robust context effect observed in Experiment 1 may have been due to acoustic differences between the targets in the possible- and impossible-word contexts. The

effect in Experiment 1 therefore appears to be due almost entirely to the contexts per se: listeners have difficulty spotting the words in the consonantal contexts because those contexts fail the PWC. The weak word type effect is in the opposite direction to that observed in Experiment 1: verbs were slightly easier to process than nouns in lexical decision, but slightly harder than nouns to detect in word-spotting. Clearly, nouns do not necessarily have a processing advantage over verbs.

In the second set of analyses, comparing iambic words and trochaic verbs taken from Cə and CVCə contexts, there were no significant effects. In the third set, however, comparing iambic words with trochaic nouns taken from Cə and CV contexts, there was an effect of context: responses to words taken from Cə contexts were, on average, 33 msec faster and 3% more accurate than those to words taken from CV contexts (RTs: $F1(1,33) = 9.1$, $p<0.005$; $F2(1,43) = 7.7$, $p<0.01$; Errors: $F1(1,33) = 13.6$, $p<0.001$; $F2(1,43) = 3.6$, $p=0.06$). No other effects were significant. In the fourth set of analyses, comparing the three word types taken from Cə contexts, the only significant effect was one of word type in RTs, significant by subjects but not by items ($F1(2,66) = 3.7$, $p<0.05$; $F2<1$). The results of these analyses support the conclusion that there is no principled difference in the ease of processing of trochaic and iambic words. The context effect observed in the comparison of words taken from Cə and CV contexts explains the equivalent difference in Experiment 1: the targets in Cə contexts were easier to spot than those in CV contexts not because of a difference between those contexts, but because of a difference between the words themselves.

4. GENERAL DISCUSSION

Words, irrespective of their stress pattern or syntactic category, are harder to recognise when they occur in a context which is not a possible word of Dutch (a single consonant) than when they occur in possible-word contexts. Furthermore, different possible-word contexts appear equally acceptable: longer contexts (CVCə) do not differ from shorter contexts (Cə or CV). These results are as the PWC would predict, and replicate findings in English [7] and related findings in Dutch [6]. Work in progress has found further support for the PWC in several other languages: French, Sesotho and Japanese. The goal of this research program is to define the nature of the cues which are used to signal likely word boundaries and to establish, cross-linguistically, what units of speech constitute possible words.

Previous word-spotting studies have used mainly noun targets, on the assumption that nouns would be easier to spot. The present results, however, show that verbs can be used in the word-spotting task. The similarity of the noun and verb results also provides a validation of the task: it appears to reveal the operation of general word-recognition processes, not only processes which are specific to noun recognition. In particular, it would appear that (as argued in [7]) the PWC is a prelexical mechanism which applies generally in continuous speech recognition, irrespective of the syntactic categories of the candidate words.

Although words beginning with weak syllables are much rarer in Dutch than words beginning with strong syllables [10], this does not mean that they are harder to recognise: there was no difference between the trochaic and iambic words. This finding runs counter to the recent suggestion that lexical stress provides Dutch listeners with a segmentation cue [10]. On such a view, trochaic words ought to have been easier to spot than iambic words. Instead, the stress-pattern results suggest that segmentation and recognition do not depend on lexical stress: the PWC applies uniformly to monosyllabic and bisyllabic words, and to bisyllables with trochaic and iambic stress.

5. ACKNOWLEDGEMENTS

We thank Mattijn Morren for his work in preparing and running these experiments. Correspondence to either author at the MPI for Psycholinguistics, PB 310, 6500 AH Nijmegen, The Netherlands; Email: James.McQueen@mpi.nl, Anne.Cutler@mpi.nl.

6. REFERENCES

1. McQueen, J.M., "Word spotting", *Language and Cognitive Processes* 11: 695-699, 1996.

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

3. Norris, D., McQueen, J.M., and Cutler, A. "Competition and segmentation in spoken word recognition", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21: 1209-1228, 1995.

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

5. Vroomen, J, van Zon, M., and de Gelder, B. "Cues to speech segmentation: Evidence from juncture misperception and word spotting", *Memory & Cognition* 24: 744-755, 1996.

6. McQueen, J.M. "Segmentation of continuous speech using phonotactics", *Journal of Memory and Language* 39:21-46, 1998.

7. Norris, D., McQueen, J.M., Cutler, A., and Butterfield, S. "The possible-word constraint in the segmentation of continuous speech", *Cognitive Psychology* 34:191-243, 1997.

8. Norris, D.G. "Shortlist: A connectionist model of continuous speech recognition", *Cognition* 52: 189-234, 1994.

9. Cutler, A., and Carter, D.M. "The predominance of strong initial syllables in the English vocabulary", *Computer Speech and Language* 2: 133-142, 1987.

10. Vroomen, J., Tuomainen, J., and de Gelder, B. "The roles of word stress and vowel harmony in speech segmentation", *Journal of Memory and Language* 38: 133-149, 1998.