DO METRICAL AND PHONOTACTIC SEGMENTATION CUES COOPERATE IN SPOKEN WORD RECOGNITION?

Bacri N. & Banel M-H.,
Laboratoire de Psychologie Expérimentale, Université René Descartes & CNRS, Paris, France

ABSTRACT

A Metrical Segmentation Strategy has been proposed to account for the segmentation of stress-timed language like English [3] whereas for syllable-timed languages like French segmentation should depend on syllabic structure. Three word-spotting experiments supported a different assumption. Detection of words embedded in nonsense strings either in initial (Exp. 1 & 3) or in final position (Exp. 2) provided evidence of an effect of prosodic structure on the phonotactically based mechanism of parsing.

Speech segmentation is basically supported by two types of cues: prosodic cues [3, 5, 6] and phonotactic cues [7]. Both are determined by language-specific constraints. Lexical parsing is assumed to be stress-based for languages contrasting stressed and unstressed syllables. It is founded on distributional phonotactic properties as evidenced by juncture misperception data [2]. Syllabification was almost equally affected by the type of phonotactic string and by the stress pattern of the stimuli [7]. In French, in which stress falls regularly on the lengthened final syllable, durational inter-syllabic differences may have a functional role in speech segmentation. The parsing of bisyllabic words into their monosyllabic constituents showed that the usual short-long (iambic) pattern produced recognition of bisyllables more often than did the reverse long-short (trochaic) pattern [1]. Subjects' behavior was well adapted to the trailer-timed structure of French language [8] and to the fact that 85% of lexical items are polysyllabic. Syllable internal structure had no clear effect on parsing, but the parsing of long-long spondees was facilitated by CVC syllables. Listeners were induced to apply a syllable-based device when lacking metrical cues. However, data were compatible with two different accounts. The efficiency of prosodic cues might be due to rhythmic expectancies and intervene post-lexically. On the other hand, a metrical segmentation device might have had a direct effect on parsing, either cooperating or conflicting with phonotactic cues.

Well-formedness conditions for syllable-final consonant clusters imply, at the phonological level, that the first but not the second consonant belongs to the same syllable as the preceding vowel: French codas cannot contain more than one consonant. "garde" would be syllabified as /gar # dØ/. But in the specific case of the obstruent + liquid legal clusters, whatever their position, the two consonants are tautosyllabic: "livre" is syllabified /li # vrØ/ [4]. If we consider an illegal medial cluster, e. g. /VpzV/, the first assumption leads to syllabify it as /Vp # zV/, whereas the second one forbids the processing of /pz/ as tautosyllabic, and blocks a /V # pzV/ parsing. These phonotactic constraints constitute powerful segmentation cues.

The aim of the present experiments was to determine whether or not metrical and phonotactic segmentation devices are simultaneously available. A word-spotting task offers a good opportunity to address this question [3]. Listeners have to detect real monosyllabic words in nonsense bisyllabic strings. The medial consonant cluster, when present, was illegal, thereby providing a segmentation cue. Metric structure realized an iambe, a trochee or a spondee. According to the metrical segmentation hypothesis, the bisyllable string will be more easily segmented when bearing a trochaic pattern, whereas it will be processed as a whole when bearing an iambic pattern. If a metrical segmentation device intervenes prelexically, detection of the word in initial position will be facilitated in the first case when both cues cooperate, and interfered with in the second case by the conflict between phonotactic and prosodic cues, as compared with the prosodically neutral condition.

EXPERIMENT 1

Method:

Subjects:

37 native speakers of Parisian French.

Materials:

The materials consisted in two sets of nonsense bisyllables. The 18 items of the first set had an embedded high frequency CVC word target in initial position. Word final consonant was an obstruent. Consonant sequence in medial position cannot form a syllabic coda or onset, e. g. /bk/, /bv/, /pz/, /tl/, /tn/. The 18 CVC-CVC items in the second set were nonsense fillers without embedded word target.

Three metrical patterns were realized for each test item: A short-long iambic pattern, a long-short trochaic pattern, and a long-long spondee. All the fillers were spondees. All the stimuli were recorded by a Parisian French female speaker, at a regular rate and without salient F0 movements. They were stored and digitized at 10 kHz with 12 bit resolution. Monosyllabic mean duration was about 520 ms when long, and about 350 ms when shortened with a compression rate of 35%.

Procedure and design:

Subjects were told that they would hear nonsense sequences, some of them having a real CVC word inside, in initial position. On detecting a word, they had to press a button as quickly as possible, and to say it aloud. Detection times were measured from the burst of the word final consonant. Three experimental tapes were constructed, one for each metrical pattern, each having 32 stimuli (18 word targets plus 18 fillers). The order of stimuli was balanced across the three tapes. The three practice tapes included 32 different bisyllabic strings (16 with, and 16 without a CVC word), bearing the same metrical pattern than the corresponding test tape. Subjects were given a feedback on their performance for the first half of the practice tape. Each experimental condition combined an illegal cluster with (1) a neutral pattern ("one cue" control condition), (2) an iambic pattern (conflicting cues), (3) a trochaic pattern (cooperating cues). Each subject heard the three metrical patterns, but each item was presented bearing only one pattern.

Results and discussion:

Seven subjects, having detected less than 50% of the targets, were discarded, leaving 10 subjects for each tape. Missing data (3.9%) and reaction times (RTs) greater than 1500 ms (2.4%) were replaced. RTs were faster for a trochaic pattern than for either an iambic pattern or a neutral spondee (Figure 1, left panel). The main effect of Condition was significant (ANOVA by subject: $F1(2, 58) = 9.3$, $p < .001$; by item: $F2(2, 34) = 21.2$, $p < .001$). Pairwise comparison showed that RTs were reliably faster for a trochaic pattern than for an iambic and a neutral pattern, respectively. The tendency towards slower RTs for a neutral pattern than for an iambic pattern did not reach significance.

The order of performance levels supported the predictions derived from the application of a metrical segmentation device. Word extraction is facilitated when a long-short pattern is correlated with an illegal cluster. In as much as the cluster cannot be tautosyllabic, it is parsed into its constituent units, and an attempt at lexical access, initiated near the beginning of the input in the case of a trochaic pattern, can be easily achieved. In the case of iambic patterns, lexical access is delayed, since the nonsense string is processed as a whole on the basis of its metrical structure. However, conflicting cues did not reliably increase RTs as compared with the control condition. The lack of a reliable interference when cues were conflicting suggests that the two types of information have been processed separately. But the redundancy gain when both cues cooperate is consistent with the view that both dimensions may nevertheless influence each other. The purpose of the next experiment was to determine whether this influence is symmetrical or not.

EXPERIMENT 2

To address this issue, the functional relations between the two segmentation cues were reversed, by locating the target word in final position. The former "conflicting cues" (an illegal cluster plus an iambic pattern) might now facilitate a lexical access attempt on the final syllable, whereas the former "cooperating" cues (an illegal cluster plus a trochaic pattern) should now deter the listeners from pursuing a lexical search after having parsed the bisyllable and found a nonword. The presence either of interferences (slower RTs with a trochaic pattern than in the control condition) or of a redundancy gain (faster RTs when iambic than in the control condition) will permit to decide whether both cues are processed conjointly and symmetrically. On the other hand, if the two cues are processed separately, results will show neither interference, nor redundancy gain.

Method:

Subjects: 30 Parisian French speakers.

Materials and procedure:

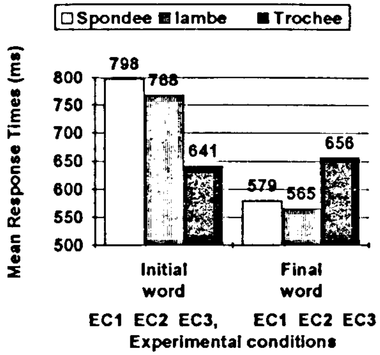
18 new nonsense bisyllabic strings were created by reversing the order of the first and the second syllable. Medial illegal clusters assembled obstruents and liquids so that they could not be tautosyllabic in French, e. g. "lamp(e) + zok" became "zor + lamp(e)" to avoid a /kl/ legal cluster. The apparatus and procedure were similar to those of the preceding experiment.

Results and discussion:

Missing data (6%) and RTs exceeding 1500 ms (3.1%) were replaced. As in the previous experiment, the main effect of Condition was significant ($F1(2, 58) = 5.2$, $p<.01$; $F2(2, 34) = 7.1$, $p<.01$). The 14 ms difference between RTs obtained for iambic patterns and RTs for the control condition did not reach significance, but RTs for trochaic patterns were reliably slower than RTs for the two other conditions (Fig. 1, right panel). The order of performance levels was reversed by comparison with the preceding experiment. There was no redundancy gain when iambic patterns induced subjects to process the whole string. But the convergence of both cues to parse the input increased RTs and slowered lexical access. These results suggest a mutual interference between prosody-based and phonotactic-based segmentation devices.

EXPERIMENT 3

The purpose of the next experiment was to evaluate the weight of the phonotactic cue, respective of metrical pattern, by comparing the processing of CVCCVC strings and of CVCVC strings.



EC1: one segmentation cue
EC2: conflicting cues
EC3: cooperating segmentation cues

Figure 1: Mean RTs as a function of experimental conditions and word position (left panel Exp.1, right panel Exp. 2).

In the later case, word coda is assembled with the following vowel, e. g. "lampoc" is syllabified /lā # pok/. The target word is in initial position. A lexical access requires a resyllabification that is impeded by an iambic pattern indicating not to parse the bisyllable, but also by a trochaic pattern which induces to parse the string in such a way that the parsing produces two nonsense syllables.

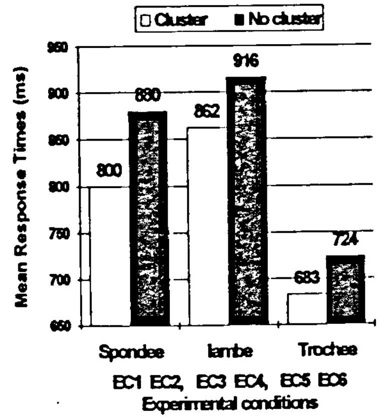
Method:

Six groups of ten subjects were tested on 18 test items, with or without a medial cluster, depending on the condition, and bearing one of the metrical patterns. The apparatus and procedure were as previously described.

Results:

Two test items were missed by 50% of the subjects. Corresponding data were discarded. The overall analysis of variance showed a significant main effect for both the phonotactic structure ($F1(1, 54) = 4.4$, $p<.04$; $F2(1, 15) = 21.8$, $p<.001$) and the metrical pattern ($F1(2, 54) = 15.7$, $p<.001$; $F2(2, 30) = 22.5$, $p<.0001$). The interaction between the two factors did not reach significance. The order of RTs was the same for both types of phonotactic structure (Fig. 2). It approximated the order obtained in Experiment 1, and yielded the same levels of significance. The weight of a

phonotactic cue is more important in the control condition. However, the 54 ms (iambic pattern) and the 41 ms lengthenings (trochaic pattern) indicate merely a tendency towards an impairment of lexical access attempt when listeners are lacking a strong phonotactic segmentation cue.



EC1: one relevant segmentation cue.
EC2: one irrelevant segmentation cue.
EC3: conflicting cues.
EC4: converging irrelevant cues.
EC5: cooperating segmentation cues.
EC6: cooperating irrelevant cues.

Figure 2: Mean RTs as a function of experimental conditions (Exp. 3).

GENERAL DISCUSSION:

The ability to segment speech input and to detect a word merged into a nonsense string crucially depends on the position of the word in the string. The target position determines the functional relations between metric and phonotactic structures, and thus the efficiency of the metrical segmentation device. However, together with the redundancy gain when both cues are cooperating (Exp. 1), and with a mutual interference when they give competing information (Exp. 2), the last experiment clearly demonstrates a joint influence of metrical contrast and of phonotactic constraints. This joint influence cannot be the result of interactions in the perceptual system (Exp. 3). We suggest that metrical structure might activate the segmentation procedure which is in turn triggered by a phonotactic disjuncture. McQueen et al.' proposal (1994) should be modified to take the specificity of French metrics into account. A metrical segmentation device would permit to anticipate word boundaries, but makes no assumption about the mechanism of parsing. The interference effect, although limited to certain condition, would be best interpreted if metrical and phonotactic information are processed in parallel.

ACKNOWLEDGEMENTS

Financial support for this research was provided by the GDR 957 "Sciences Cogntives de Paris" (CNRS).

REFERENCES

[1] Banel, M-H., Bacri, N. (1994), "On metrical patterns and lexical parsing in French", *Speech Communication*, vol. 15, pp. 115-126.

[2] Cutler, A., Butterfield, S. (1992), "Rhythmic cues to speech segmentation: Evidence from juncture misperception", *J. of Memory and Language*, vol. 31, pp.218-236.

[3] Cutler, A., Norris, D. (1988), "The role of strong syllables in segmentation for lexical access", *J. of Experimental Psychology: Human Perception and Performance*, vol. 14, pp. 113-121.

[4] Dell, F. (in press), "Consonant clusters and phonological syllables in French", *Lingua*, 95.

[5] Grosjean, F., Gee, J. P. (1987), "Prosodic structure and spoken word recognition", *Cognition*, vol. 25, pp. 135-155.

[6] McQueen, J., Norris, D., Cutler, A. (1994), "Competition in spoken word recognition: Spotting words in other words", *J. of Experimental Psychology: Learning, Memory and Cognition*, vol. 20, pp. 621-638.

[7] Treiman, R., Danis, C. (1988), "Syllabification of intervocalic consonants", *J. of Memory and Language*, vol. 27, pp. 87-104.

[8] Wenk, B., Wioland, F. (1982), "Is French really syllable-timed?", *J. of Phonetics*, vol. 10, pp. 193-216.