The role of the syllable in spoken word recognition: access or segmentation?

FRAUENFELDER, Ulrich Hans, CONTENT, Alain


Available at: http://archive-ouverte.unige.ch/unige:83992

Disclaimer: layout of this document may differ from the published version.
The role of the syllable in spoken word recognition: Access or segmentation?

Uli H. Frauenfelder¹ & Alain Content¹,²

1 Laboratoire de Psycholinguistique Expérimentale, Faculté de Psychologie et des Sciences de l’Éducation, Geneva, Switzerland
2 Laboratoire de Psychologie Expérimentale, Université libre de Bruxelles, Belgique

Abstract. Spoken word recognition involves the classification and the segmentation of a variable and continuous speech input. Different proposals have been made to characterize the relation between these two operations and the nature of the resulting units. One well-known proposal for French assumes that the syllable constitutes the basic perceptual unit of speech processing. The present overview summarizes recent data from three different experimental paradigms that aim to clarify the processing role of the syllable. These data underscore the primary segmentation function of syllabic structure in word recognition. More specifically, it is argued that listeners use syllable onsets as alignment points for lexical matching.

1. INTRODUCTION

The ability of listeners to identify correctly and almost instantly a word from amongst the tens of thousands of other words stored in their mental lexicon constitutes one of the most extraordinary human cognitive feats. The rapidity with which the listener can recognize words is all the more astonishing given the nature of the input and the content of the lexicon. The speech signal indeed presents a formidable challenge. First, speech is variable; every word takes on a different phonetic shape each time it is produced. The existence of large numbers of highly similar words in the lexicon makes this variability even more troublesome. Second, speech is continuous. Unlike written text, it contains no systematic “spaces” or reliable markers to indicate where one word ends and the next one begins. The existence of embedded lexical entries (“corps” in “corbeau”) increases the difficulty of lexical segmentation, since multiple possibilities often must be pursued in parallel.

How do listeners process the incoming speech signal so efficiently despite its variable and continuous nature? It is generally assumed that some intermediate representations serve as an interface between the speech signal and the phonological representations stored in the lexicon (see for review, Kolinsky, 1998). The postulation of a prelexical input representation is more than simply a convenient line of demarcation between psycholinguistics and phonetics. This representation is assumed to play a key role in solving the problems of classification. A phonetic or phonological representation provides a more abstract and less variable format for representing the information in the input and for accessing the lexicon. Instead of trying to match each unique sensory input directly onto a huge number of lexical entries, listeners first recode the variable input into
a smaller number of abstract units like phonemes or syllables that in turn serve to contact the lexicon.

The intermediate representation based upon units can potentially guide the segmentation process. For instance, the onset of prosodic units (Grosjean & Gee, 1987) or strong syllables (Cutler & Norris, 1988) could be used as starting points for the lexical matching process. To the extent that these segmentation points are likely to correspond to word boundaries, such heuristics would be helpful in reducing wasteful attempts to match the input with misaligned lexical candidates.

In a framework that attributes a central role to intermediate levels of representation, we are led to search for the nature of the units making up this representation. We find considerable variety in the types of perceptual units that have been proposed in the literature. These include distinctive features (Lahiri & Marslen-Wilson, 1991), phonemic segments (Pisoni & Luce, 1987), and syllables (Mehler, 1981). However, the notion of "perceptual unit" is ambiguous, since in principle segmentation and classification units are not necessarily identical. Furthermore, segmentation and classification units appear to vary from one language to another, depending upon the phonological properties of the language. In what follows we will briefly contrast some conclusions drawn for these units in French and English.

2. PERCEPTUAL UNITS IN FRENCH AND ENGLISH

It has been claimed upon the basis of various experimental findings (Seguí, Dupoux & Mehler, 1990) that the syllable constitutes the perceptual unit in French. According to this view the continuous speech stream is segmented and classified into syllabic units to construct an intermediary syllabic representation. In a seminal study (Mehler, DommGregues, Frauenfelder & Seguí, 1981) using the fragment detection procedure (Frauenfelder & Kearns, 1997, for description), French subjects detected consonant-vowel (CV) or consonant-vowel-consonant (CVC) targets in spoken carrier words which varied in syllabic structure (CVCV- "pa.lace" vs. CVC.V- "pal.mier"). Detection latencies were faster when the target corresponded to the initial syllable of the carrier word than when it did not. This result, the so-called syllable effect, provides evidence for the salience of the syllable-sized units in French. These authors (Seguí et al., 1990) go on to argue on the basis of other findings that words are accessed by their initial syllable.

The situation for English appears to be somewhat different. Indeed, in an experiment using the same fragment detection paradigm with comparable stimuli in English, English listeners did not show the same sensitivity to syllabic structure as French listeners (Cutler, Mehler, Norris & Seguí, 1986). These authors concluded that syllabic segmentation was more appropriate for French than for English, citing as the underlying reason, differences in the phonological and rhythmic properties of the two languages. These authors go on to argue for the existence of a language-universal mechanism that helps locate possible word onsets in the continuous signal. They claim that the precise manner in which this mechanism operates varies across languages as a function of the sources of information that are available in each language, giving rise to language-specific segmentation strategies and units. For English a Metrical Segmentation Strategy
is proposed according to which the metrical foot and its initial strong syllable is assumed to constitute the unit of segmentation that serves as the point of alignment with the lexicon (Cutler & Norris, 1988). Thus, the goal of segmentation is less to permit or facilitate the speech classification process than to find the likely alignments between the classification units and the lexical entries. A clear distinction is made then for English between the operations of classification and segmentation; the units are not the same for each operation.

In contrast, there is some ambiguity concerning the exact role of the syllable in the proposals for French, due perhaps to the fact that the syllable is proposed as both the segmentation and classification unit. We can oppose a view in which, like for English, syllable onsets are located to identify potential points of lexical alignment and another view in which the delimitation of syllabic units guides the classification process by identifying the end-points of syllabic chunks that are in turn classified. In what follows, we present empirical work that will help decide between these two alternatives. The goal of the present paper is thus to present novel data in an attempt to clarify the role of the syllable in the recognition of French words.

3. IS THE SYLLABLE A CLASSIFICATION UNIT?

Surprisingly, despite the multitude of fragment detection experiments in various languages with word carriers, few studies have tested for the “syllable effect” in pseudoword carriers, not even in French. The use of pseudowords has the advantage of making it possible to increase the number and diversity (with a greater variety of vowels and pivotal consonants) of experimental stimuli. Moreover, an experimental demonstration of the syllable effect with pseudowords would help confirm its prelexical locus and generality. Finally, this study provides sufficient data for evaluating the relation between the RTs and the acoustic information in the stimuli. A detailed investigation of this relation should provide further insights into the nature of classification units.

Four fragment detection experiments examined the syllable effect with pseudoword carriers. The experiments differed in the presence or absence of foils (Exp. 2 was without foils) and the organization of the stimuli into blocks. The foils (e.g., target GU, carrier GALOR; target GUL, carrier GURIP) were devised to ensure that subjects could not respond on the basis of a partial match on the initial segments, but had to process all of the segments in the target. Subjects were required to press a key as rapidly as possible if the target (PA or PAL) was present at the beginning of the carrier (PA.LOUNE / PAL.NUFFE). Four different vowels (/a/, /i/, /u/, /y/) as well as three classes of pivotal consonants (liquids, stops and fricatives) were used in the construction of the thirty-two pairs of pseudowords. The results for these experiments are shown in Figure 1.
These experiments failed to replicate the syllabic effect with no hint of the familiar crossover interaction. Two effects were obtained with 1) the detection of CVC targets requiring more time than the detection of CV targets and 2) the detection times to CVC carriers being faster than to CV carriers. Taken together, the results are consistent with a phonemic processing hypothesis according to which the target is detected as soon as all of its segments are identified in the signal. To test this hypothesis further, the detection latencies were correlated with 1) phonetic measurements and 2) perceptual estimates of the vowel onset temporal position (VO₁ and VO₂) and of the pivotal consonant onset temporal position (PCO₁ and PCO₂) on the carriers. In the perceptual gating experiment, participants (N=21) heard increasing portions of signal (by increments of 15 ms) and wrote down what they heard. Estimates of VO₂ and PCO₂ were computed as the averaged duration of the signal for which the vowel and pivotal consonant, respectively were first reported accurately.

For each experimental and each type of target (CV or CVC) separately we ran forced order regression analyses to examine the proportion of the variance across items accounted for by the temporal position of the vowel (step 1) and by the delay between the onsets of the pivotal consonant and the vowel (step 2). The outcomes based on acoustic and perceptual measurements were highly consistent. The location of vowel onset accounts for a large part of the item variance (50 to 80%), and more so for the detection of CVC targets than CV targets. Total variance explained is between 50 and 90%, except for Exp. 1 (without foils). Furthermore, and more interestingly, the temporal position of the pivotal consonant accounts for 20-25% of the remaining variance, but only for CVC targets.

The absence of the syllable effect on the mean detection times and the results of these regression analyses confirm that the classification processing underlying the detection
decision is determined primarily by phonemic throughput. The tight fit between the physical measures of the signal and the RTs suggests that input is classified continuously into infrasyllabic units rather than discontinuously in syllable-sized chunks. In sum, the results of the preceding detection experiments do not provide any support for the syllable as a classification unit.

4. IS THE SYLLABLE A SEGMENTATION UNIT?
Given our failure to find evidence for the syllable as a classification unit, we decided to conduct other studies aimed at assessing the role of the syllable in the segmentation process. Indeed, the fragment detection task is not particularly well suited for determining the nature of the segmentation unit. By presenting subjects with isolated carrier words and targets matching the onsets of these carriers, the experimenter has essentially resolved the segmentation problem for subjects in this task. In a first set of experiments we examined the explicit segmentation of bisyllabic words by adult native speakers; in a second study we used the word-spotting task to assess the role of syllable structure in on-line word recognition.

4.1 Explicit syllabic segmentation
Psycholinguistic evidence about segmentation can come from metalinguistic tasks in which subjects are required to consciously manipulate a given input string in different ways. For example, in a set of studies investigating adults' intuitions about syllabification (Treiman & Danis, 1988), bisyllabic words with intervocalic consonants (liquid, nasal or obstruent) were presented to American subjects who were required to reverse the two syllables. The listeners' performance was not consistent; the intervocalic consonant was sometimes placed in the first syllable (melon -> lon-me), and sometimes in the second (melon -> on-mel). In addition, there was also a considerable number of responses which they called “1-2” where the consonant was placed in both syllables (lon-mel). The more sonorous the consonant, the greater the probability that it was placed in the first syllable rather than in the second syllable, and the greater the probability that it would elicit responses where the consonant was placed in both syllables (liquids are more sonorant than nasals, and nasals more sonorant than obstruents).

The aim of these experiments (Content, Kearns & Frauenfelder, 1998) was to examine the syllabification of singleton intervocalic consonants by French native listeners, and to compare their segmentation preferences with those of American listeners. In order to do so, we employed several syllable manipulation tasks. Figure 2 shows the results of two such studies. Although the participants consistently placed the intervocalic consonant in the second syllable (CV responses for second part), they were much less consistent in giving these CV responses for the first part. They often produced CVC responses, thereby placing the intervocalic consonant in both syllables. Such “1-2” responses, which were more frequent in the words with a more sonorous intervocalic consonant, replicate the findings observed in English by Treiman and Danis (1988) as well as by Schiller, Meyer & Levelt (1997) in Dutch. Overall, these findings suggest that even French listeners are not consistent in their segmentation of syllables, at least for syllable offsets.
Figure 2. Proportion of CV responses as first (left bars) or second part (right bars) of CVCV items, and as a function of task. In “Single” conditions, participants produce either the first or the second part. In “Double” condition, participants produce both parts separately, in either order.

Interestingly, by separating those responses that required the determination of the syllable onsets from those requiring decisions for syllable offsets, we observed a clear dissociation. The former were much more consistent than the latter; responses for the second syllable of the word remained invariable and nearly always began with the consonant, whereas responses for the first syllable varied more, generally ending with the vowel but sometimes with the consonant. This difference in the pattern of responses for the two syllables suggests the existence of two separate processes, one responsible for locating syllable onsets and the other for locating syllable offsets. Onset detection appears to be highly consistent in marking a single intervocalic consonant as an onset, while offset detection appears more variable and is sensitive to sonority and spelling.

The pattern of results obtained in these metalinguistic experiments has been replicated in a speeded segmentation task with adults (Goslin, Content, Goldman & Frauenfelder, 1999) as well as with literate and preliterate children (Content, Dumay, & Frauenfelder 1999). Thus, we have convergent evidence for the conclusion that the detection of syllable onsets has a special status in French.

4.2 On-line embedded word recognition

In order to strengthen this conclusion further, we sought additional experimental results using a more on-line task that required lexical identification. Thus, we conducted a word spotting experiment (Dumay, Banel, Frauenfelder & Content, 1998) in which participants made speeded manual responses when they detected monosyllabic words embedded at the beginning or end of bisyllabic nonwords (e.g., [lac] in CVC.CVC [lac.tuf] or [zun.lac]). In addition, participants also detected the same word in non-aligned versions for both onset and offset position (CV.CCVC [la.gluf] or [zu.glac]). According to the syllable onset hypothesis, the perceptual cost attributable to the mis-alignment should be greater for the condition in which the syllable onset and not the syllable offset is mis-aligned with the word to be detected. This prediction was confirmed by the data, with a
significant perceptual cost for the cases of [zu.glæ], thereby providing additional support for the onset detection hypothesis.

5. CONCLUSIONS AND PERSPECTIVES

The findings summarized in this short overview paper paint a picture of speech processing in which the French listener segments the continuous input by locating syllable onsets to guide the lexical alignment process. This syllabic segmentation strategy, however, does not imply that the syllable is the classification and access unit. Rather the results of our fragment detection experiments for pseudowords suggest that the listener classifies the sensory input into units that are smaller than the syllable. However, given the discrepancy between our findings and those by Mehler and his colleagues in support of syllable as a classification unit, further research into the precise nature of these units is still needed.

One shortcoming of a syllabically-based segmentation strategy is that it does not easily handle cases of resyllabification of the input that result from phonological processes across word boundaries. For example, syllable boundaries are not coterminous with word boundaries in case of liaison, and an incorrect lexical alignment would be made on the basis of syllable onsets. The way in which listeners cope with such problems as encountered in fluent speech is attracting increasing attention, but still needs to be examined more closely. In a recent study (Dumay, Content & Frauenfelder, 1999), we showed that listeners can exploit certain durational cues in the signal to locate word boundaries. In other research in progress using the phoneme monitoring task, we observed that vowel initial words like "animal", become activated despite the presence of an inappropriate initial consonant (/v/ in "animal"). Research along these lines should provide further insight into the role that syllable onsets play in lexical segmentation in these problematic cases. Overall, our findings concur with the conclusions of Cutler and her colleagues and point to the existence of a universal segmentation strategy that has language-specific implementations. In French, this heuristic strategy appears to be based on syllables or, more specifically, on syllable onsets.

6. ACKNOWLEDGEMENTS

The research described in this paper has been financed by grants from the Swiss FNRS (Project 1113-049698.96) and from the Ministère de la Communauté Française de Belgique (A.R.C. 96/01-203).

7. REFERENCES


