# RESOLVING WORD BOUNDARIES IN SPOKEN FRENCH

D. C. Bradley* and B. Dejean de la Bâtie**

*Department of Psychology and **Department of Romance Languages
Monash University

ABSTRACT - For three populations of listeners (native French speakers, and tertiary students in the first or second year of their post-secondary studies of French), the ease with which word boundaries are located in connected speech was investigated using the phoneme monitoring task. With response required to word-initial / t /, liaison environments created an ambiguity about the status of surface / t / which could be resolved only lexically, and performance here was contrasted with that in unambiguous cases. The cost of ambiguity tended to be less in native speakers than in second year students, in line with a presumed difference in processing efficiency; for first year students, knowledge limitations and consequent speed-accuracy tradeoffs resulted in an estimated ambiguity cost which was, if anything, less than that in natives. For all three populations of listeners, monitoring performance was frequency-sensitive: word-initial / t / was detected faster in higher frequency carriers, whether or not these occurred in the ambiguity-creating environment. These findings are discussed with respect to processing models of word-boundary resolution, in native and non-native speakers.

## INTRODUCTION

The research to be reported arises in the first instance as a problem in applied psycholinguistics: what is to be said (and in the long run, done) about the evident difficulties suffered by non-native speakers in listening to spoken French? Such difficulties persist even where the formal knowledge of French is high, and indeed the non-natives who are our experimental subjects are English-speaking tertiary-level students of the French language, with some six years of secondary-level study behind them. The response in the teaching situation to this pervasive problem has recently emphasized listening strategies which exploit high level constraints on sentence well-formedness. That is, students are encouraged to use the sense of what has been grasped as an inference base from which what has been missed can be guessed and what is to come can be anticipated. "Front-end" skills are correspondingly de-emphasized: sound drills give practice in distinctions which are tricky for non-native speakers, but sound structures above the level of the isolated word are not often the focus of instruction in listening - and this is not unreasonable, in the absence of direct evidence around which programmes can be formulated.

But there are at least arguments to suggest indirectly that the front-end problem may be a real contributor to listening failures in English-speaking students of French. These arguments start with the observation that French and English differ in the detail of their characteristic suprasegmental structures. Open syllables predominate in the lexical inventory of French, and external sandhi processes operate in connected speech to maximize this tendency. English shows a broader spread of syllable configurations lexically, and the processes of connected speech work to increase syllable weight, not decrease it. Above the syllable, Pulgram (1965) goes so far as to say that "in French, the lexical word in the context of an utterance has absolutely no phonological reality" (p. 138). This undoubtedly overstates the case - for example, Rochet's (1977) analyses of a number of phonological alternations in modern French critically require mention of word boundary - but we take the consensus to be that it is representations at and above the level of the phonological word which play a major role in shaping the prosody of French. English, with its lexical stress, presents a quite different picture.

It seems more than reasonable to assume that processes early in the chain of spoken language recognition will be intimately related to the morphophonological influences in terms of which the input sequence is structured. The studies of Cutler and her colleagues (e.g., Cutler, Mehler, Norris and Segui, 1986) suggest an initial parse of the speech stream, for French speakers, which is sensitive to syllabic structure, and a parse of another kind for English speakers, without such sensitivity. For levels higher than the syllable, the evidence for processing differences is less immediate, but a good

deal can be gleaned from the issues which have been of concern in psychological modelling. All models of sentence comprehension take the recognition of an utterance's lexical elements to be a critical early stage; this has meant that, one way or another, the pervasive problem of recovering underlying word boundaries in a signal not noted for flagging these junctures must be confronted.

Cole and Jakimik (1980) propose a solution to the problem of locating word boundaries which is stunning in its simplicity. They assume, as do many others, that lexical identification follows the acoustic input very closely indeed, and proceeds, so to speak, strictly left-to-right; given this, they propose that sequential identifications allow each word's recognition to locate the onset of the word which follows. Other less optimistic models notice that in very many instances contributions to the word-level parse will come from both prosodic and lexical knowledge sources; in effect, these models acknowledge the complex texture of the input signal. Cutler (1989) argues that speakers exploit the rhythm of utterances as an essential part of the segmentation upon which the lexical access is predicated; for English, she hypothesizes that word boundary is posited at the onset of any syllable containing a strong vowel. Alternatives see word boundary location emerging more as part of the solution arrived at, with prosodic and lexical constraint in interplay: Grosjean and Gee (1987), for example, suppose that the speech stream is parsed *in the course* of lexical identification. The proposal, again for English, is that stressed syllables initiate lexical search, and that identification arises through the convergence of lexical knowledge and an independently computed phonological structure.

Any word-level segmentation which is based around particular prosodic categories (syllables with strong vowels, stressed syllables) clearly risks being language-specific: prosodic shapes differ between English and French, as we have seen. Thus, if the insights which the models capture are even roughly correct (and there seems good reason to assume this), the English-speaker who studies French has a real problem: there will be systematic mismatch between the first-language parsing routines and the character of second-language inputs. But it is undoubtedly true that adult acquirers can, in the end, become competent listeners in their second language, and it may be that this comes about through an alteration of the priority assigned to lexical and prosodic knowledge types (in favour of the former). Pedagogically, this might be good news. It seems pertinent here that the lexical knowledge base ordinarily continues to expand in adulthood, even in the native language, while for adult acquisition of a second language, the last and notoriously unjumpable hurdle is its phonology.

Our research therefore focussed on listeners' use of specifically lexical knowledge in establishing word boundaries. The ideal situation is one for which phonetic and prosodic contributions are strictly neutral, and Matter (1986) in his study of the issues we are canvassing came up with just the right case, *liaison potentielle*. In liaison, the parade case in French of external sandhi, so-called latent consonants (normally not pronounced word-finally) surface when a following word in certain syntactic contexts is vowel-initial, and are tautosyllabic with that word. This allows the construction of sequences which differ in word boundary location but are otherwise matched phonetically and prosodically. The critical cases involve Adjective-Noun constructions which do or do not constitute liaison environments (e.g., *grand arbre* with obligatory liaison, and *grand talent* with no possible realization of the liaison consonant); here, only lexical knowledge establishes which noun carries word-initial / t /. Comparison cases use non-liaison adjectives (e.g., *vrai arbre, vrai talent)*, for which prior identification of the prenominal adjective is potentially sufficient to make that distinction. Patterns of reaction time (RT) in a phoneme monitoring task employing such materials therefore speak to the efficiency with which lexical knowledge can be brought to bear by different populations of listeners, as do patterns of accuracy. For more detail, analyses directed at the trademarks of lexical identification (e.g., frequency and duration effects on detection RT) allow us to determine how routinely lexical knowledge is deployed, both when it is necessary to resolve word boundary ambiguity and when it is not.

## METHOD

### Subjects

Three speaker groups, each of 30 adult subjects (age range 18 to 50 years), participated in the half-hour experiment. Native French speakers were recruited by advertisement on the Monash University

campus, and through the French-speaking community of Melbourne; non-native speakers, from French language courses at Monash University. A higher competence non-native group were enrolled in the second year of their university course and, for all but two subjects, had previously studied French to matriculation level. A lower competence group of non-native speakers, all of whom had been students of French through secondary school, were in their first year of university study.

Materials and Design

The core materials set comprised 48 critical nouns selected in two groups, / t /-initial (e.g., *talent, territoire, tigre)* and vowel-initial (e.g., *arbre, ecolier, idiot)*, and 6 prenominal adjectives which can potentially set up liaison environments (*grand, petit, excellent)* or cannot (*autre, vrai, mauvais)*. Noun selection was constrained in two ways: across the two groups, the distribution of first syllable vowels was to be matched; and familiarity was to be as high as possible, chiefly for the benefit of non-native subjects. Frequency of occurrence was estimated, for native speakers, from the 20th century samples of the *Dictionnaire de Fréquences*; geometric mean frequency for the critical / t /-initial nouns was 13 occurrences per million (range 1 to 300). For non-native speakers, item frequencies were checked with a ratings procedure, some four weeks after collection of the experimental data; subjects rated the critical nouns on a 5-point scale (value 1 rare, value 5 common). Not surprisingly, mean rated frequency was higher for more senior students (3.41 versus 2.95); ranges were 1.70 to 4.77 for second year students, and 1.27 to 4.67 for first year.

Nouns and adjectives were combined in a fixed frame, to form 96 sentences representing 24 instances of each of 4 experimental conditions; nouns, whether / t /- or vowel-initial, were each placed in combination with a liaison adjective and a non-liaison adjective in a minimal sentence, to give four experimental sentence types as illustrated below:

(1) *C'est un grand talent*
(2) *C'est un vrai talent*
(3) *C'est un grand arbre*
(4) *C'est un vrai arbre*

For a task requiring response to word-initial / t /, the design is thus one in which the presence or absence of the target phoneme is combined factorially with the ambiguity or non-ambiguity created the environment. Types (1) and (2) represent positive instances, and (3) and (4), negative instances. In types (1) and (3), lexical identification of the prenominal adjective is not definitive about the role of surface / t /; only lexical lookup (identifying *talent* but **alent*, and *arbre* but **tarbre)* differentiates between a genuinely word-initial / t / in (1) and a liaison consonant in (3). Types (2) and (4) are unambiguous. These materials were arranged to form an experiment in two versions, each with 48 sentences comprising 12 of each type; within either version, there were 8 uses of each adjective, but no critical noun was repeated. Across versions, all nouns occurred in both ambiguous and unambiguous conditions. Twelve additional practice sentences were constructed, representative of the experimental set. Materials were arranged in a fixed pseudo-random order.

Recording Techniques

The experimental tapes were recorded by a female native speaker of French (Ile de France dialect). In making the recording, the speaker used the formal register which makes liaison obligatory in these Adjective-Noun constructions, and a second native speaker confirmed for each instance that the liaison had been made. The tape for each version began with a reminder (in French) of the task instructions; the chief purpose here was to familiarize subjects with the speaker's voice and intonation pattern. The first four practice items, representing one each of the experimental conditions, were presented with accompanying explanation of the response required. All test sentences were uttered with natural intonation at a slow-to-normal place, at 6 second intervals, and buffer sentences which were subsequently edited out minimized any influence of list-start on prosody. The master tape was digitally recorded, and a final editing cross-recorded material to a second digital system. Only one recording channel was used for speech; on the second, trigger pulses were placed before the critical sentence-final noun, to allow automatic measurement of reaction time for positive instances, and detection of false alarm responses for negative instances. For positive instances, the displacement of the pulse from the / t /-burst was measured with a digital oscilloscope, and the offsets entered as correction factors in the data collection routine.

Procedure

instructions to subjects (in French for native speakers, in English for non-native) outlined the GO / NOGO task involving detection of word-initial / t /, and gave illustrative examples. The fixed form of sentences was emphasized, and the necessity for both speed and accuracy in performance stressed. The recorded materials were presented through headphones, the speech channel of the tape being split binaurally. Via an interface to a personal computer, data were collected automatically; the occurrence of a second-channel pulse started a millisecond clock which was stopped when the subject signalled a detection response with a button press. No feedback was given to subjects.

RESULTS

Table 1 summarizes performance descriptors for each of the subject groups, for cases in which the sentence-final noun was properly identified as / t /-initial, or not. A striking feature of the RT data is the long latency of detection responses, for natives (851 ms) as much as for non-natives (884 ms and 876 ms for first and second year students, respectively). Typically, latencies reported for the phoneme monitoring task are roughly half that seen here.

| Sentence type | Native speakers | | Second year students | | First year students | |
|---|---|---|---|---|---|---|
| | | | Word-initial / t / present | | | |
| Ambiguous | 883 | (0.3) | 925 | (10.8) | 911 | (15.0) |
| Unambiguous | 819 | (0.0) | 828 | (1.4) | 856 | (5.3) |
| | | | Word-initial / t / absent | | | |
| Ambiguous | | (0.0) | | (10.8) | | (18.6) |
| Unambiguous | | (0.0) | | (0.3) | | (0.6) |

Table 1. (Upper panel) Mean RT, in ms, and error rate (misses, in parentheses) for detection responses as a function of speaker and sentence type. (Lower panel) Mean percentage false alarms (incorrect detections).

The first analysis takes native French speakers and the non-native group with higher competence, students in the second year of their post-secondary French course, and compares performance for / t /-detections when the target phoneme is in an environment which makes the word boundary's location ambiguous (e.g., *grand talent)* with those for targets in unambiguous environments (e.g., *vrai talent).* For the two subject groups taken together, there is a marked elevation in RT associated with ambiguity, min $F'(1,42) = 22.38$, $p<.001$. An apparent trend for greater ambiguity cost in the non-native group (97 ms) than in the natives (64 ms) is not, however, secure, $F_1(1,56) = 2.91$, $p<.10$, $F_2(1,22) = 2.63$, $p>.10$; nor do the groups differ in overall latency, $F_1 < 1$, $F_2(1,22) = 2.34$, $p>.10$. The corresponding error data showed a striking difference between the groups that plausibly reflects the relative efficiency or adequacy with which lexical information is recruited; for native speakers, monitoring performance is essentially faultless, while non-natives miss a proportion of the target phonemes in ambiguous environments. (We note, however, that just 3 of the 24 targets attract some 70% of such errors - the magnitude of the ambiguity cost may nonetheless be underestimated.) Mistakes also occur in the other direction, and a liaison consonant is taken as word-initial / t /. In some (but not all) cases, this is reconstructable for speakers without full command of phonetic detail in terms of a lexical competitor; for example, *grand elephant* is often interpreted as *grand téléphone.*

A broadly similar picture emerges for the non-native group for whom we presumed a lower competence, the first year university students. The cost of ambiguity for these subjects is estimated at 55 ms, if anything *less* than that for the native speaker group. But again an underestimate arises from their knowledge limitations, and it is one which is likely to be more extreme than for the senior students: errors in both directions were even more probable, here. The incidence of miss errors is spread broadly enough to justify an item-based analysis of the contribution of frequency; that analysis showed that it is items rated lower in frequency for which target phonemes are more likely to be missed, $r = -0.61$, $p<.01$. False alarms are also broadly spread, but frequency is not a reliable determinant in these cases, $r = -0.30$, $p>.10$.

The elevated response times for all subject groups noted earlier raise an issue which emerged in the first decade of the chequered career of the phoneme monitoring task. In setting up the sentence types which make up the conditions of the experiment, we forced an involvement of lexical processes, for half the cases - only this way can the status of surface / t / be resolved given a potential liaison - and allowed that the remaining cases could be sorted out pre-lexically. This formulation is in line with a re-construction offered by Cutler and Norris (1979) of the phoneme monitoring task's underlying mechanisms. These authors proposed one target detection process which depended only on the identification of the preceding word to establish a boundary (cf. Cole & Jakimik, 1980), and a second process, operating independently, which was driven by lexical identification of the target carrier; monitoring performance would reflect the characteristics of whichever route was first available to support a response. Table 2 summarizes item-based correlational analyses which seek the trademarks we expect if monitoring performance is of the second type: RT for correct detections is correlated separately against the frequency and the duration of the critical / t /-initial nouns. (Note that the frequency values used in the calculation are those which are appropriate for each subject group).

| Sentence type | Native speakers | Second year students | First year students |
|---|---|---|---|
| | | Item Frequency | |
| Ambiguous | -.469 * | -.453 * | -.432 * |
| Unambiguous | -.708 ** | -.503 * | -.511 * |
| | | Item Duration | |
| Ambiguous | .435 * | .223 | .158 |
| Unambiguous | .581 * | .213 | .174 |

Table 2. Coefficients of correlation for itemwise phoneme monitoring RT with item frequency (upper panel) and duration (lower panel). * $p<.05$ ** $p<.01$

The data suggest very strongly that / t /-detection is lexically guided not only in that half of the materials set which require it, by design, but also in the cases for which it is not strictly necessary; detections are made more rapidly for higher frequency words for all three subject groups, and frequency-sensitivity is greater, if anything, in the unambiguous case. This observation can only be strengthened by two features of the materials set: its fixed sentence frame, and the eight-fold repetition of the adjectives which set up the environments. It cannot be the case that there is a problem in recognizing the word which precedes the target. The analyses with duration reinforce the frequency evidence, for native speakers: the detection response is initiated faster for shorter words, and that relation tends to be stronger where the adjective does not offer the possibility of liaison. In the data for the two non-native groups, however, only the barest trend remains in the correlations with duration. Plausibly, this might reflect some kind of fluctuating evidence level in listeners whose criterion for lexical match is opportunistic rather than calculated.

What cannot be resolved here is whether a uniform reliance on lexical identification is a special strategy invoked by the experimental situation, though the fact that it occurs so strongly in native speakers is a hopeful sign. If it is not an artifact, the suggestion must be that Cutler and Norris's (1979) prelexical routine for target detection is not an option - at least in French sentences. This in turn would suggest that word boundaries are not posited on the basis of sequential word identifications which operate from left to right - at least in French sentences.

## CONCLUSIONS

Our findings favour - for French at least - a characterization of spoken word recognition which construes word boundary assignment as something negotiated under the simultaneous constraint of lexical and prosodic analyses, not something projected forward by prosodic or contextual analyses as a preliminary to lexical search. And in training non-natives to become fluent listeners, programmes might well direct themselves to enhancing lexical identification skills in ways that emphasize the appreciation of phonetic detail and prosodic shape as much - if not more than - they do interpretive factors.

## ACKNOWLEDGEMENTS

The work reported forms part of the doctoral research of the second author, and was supported in part by a grant to the first author from the Australian Research Council. We thank the Alliance Française for their co-operation in the recruitment of subjects, and Mrs. Catherine Julien-Kemp for her services as native speaker.

## REFERENCES

Cole, R.A. & Jakimik, J. (1980) *A model of speech perception.* In R.A. Cole (Ed.) *Perception and production of fluent speech,* (Erlbaum: Hillsdale, NJ).

Cutler, A. (1989) *Auditory lexical access: Where do we start?* In W. Marslen-Wilson (Ed.) *Lexical representation and process,* (MIT Press: Cambridge MA).

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

Cutler, A. & Norris, D. (1979) *Monitoring sentence comprehension.* In W.E. Cooper & E.C.T. Walker (Eds.) *Sentence processing: Psycholinguistic studies presented to Merrill Garrett,* (Erlbaum: Hillsdale, NJ).

Grosjean, F. & Gee, J.P. (1987) *Prosodic structure and spoken word recognition,* Cognition, 25, 135-155.

Matter, J.F. (1986) *A la recherche des frontières perdues: étude sur la perception de la parole en français,* (De Wereit: Amsterdam).

Pulgram, E. (1965) *Prosodic systems: French,* Lingua, 13, 125-144.

Rochet, B. (1977) *On the status of the word in French phonology,* International Review of Applied Phonetics, 15, 187-196.