French Intonational Rises and their Role in Speech Segmentation

Pauline Welby

The Ohio State University, Columbus, OH, United States
and
Institut de la Communication Parlée, UMR CNRS 5009, INPG, Univ. Stendhal, Grenoble, France
welby@ling.ohio-state.edu

Abstract

The results of two perception experiments provide evidence that French listeners use the presence of an early intonational rise and its alignment to the beginning of a content word as a cue to speech segmentation.

1. Introduction

How do people engaged in a conversation know where one word begins and the next word ends? Spoken sentences do not have the convenient spaces that separate words in written sentences. This lack of clear boundaries can lead to misperceptions. The photograph in Figure 1, taken at the Musée de la Révolution Française in Vizille, France, illustrates an example of such a misperception. The maker of the plate clearly intends to proclaim l’abondance ‘abundance’, but missegments the presumably unfamiliar word as la bondance. The la on the first line is one form of the definite article, but bondance is not a French word. Such missegmentations are the exception rather than the rule. The current study investigates how intonational patterns in French help listeners effectively segment speech.

A large body of research shows that listeners exploit their tacit knowledge of a wide range of patterns in their native language to help them segment speech, making the process automatic and seemingly effortless. For example, Dutch listeners have been shown to be sensitive to the fact that the sequence [mr] is impossible both syllable-initially and syllable-finally in their language. Such a sequence can only occur across a syllable boundary and therefore a possible word boundary [1].

Listeners also use language-specific metrical cues in speech segmentation. English listeners, for example, know that the vast majority of English content words (nouns, adjectives, etc.) begin with stressed syllables [2], and their perception errors reflect this pattern [3, inter alia]. Errors in which a word boundary is placed before a stressed syllable (l’illegal → an eagle) are much more common than those in which a word boundary is placed before an unstressed syllable.

French, however, does not have contrasive word stress. The results of a number of experiments suggested that French listeners use a syllable classification strategy in speech segmentation [4, 5]. Participants in those studies detected sequences like ba more quickly in words like balance ‘scales’ than words like balcon ‘balcony’. In the first case, the target sequence ba matches the syllabic structure of bal lance, while in the second, there is a mismatch (bal con). The reverse was found if the target sequence was bal. A recent study, however, calls into question these results [6]. The study successfully replicated the critical interaction effect only for items that were segmentally similar to those used in the earlier experiments, that is, items with liquid “pivotal consonants.” This result shows that the apparent syllable effect is highly dependent on segmental and cannot be a general strategy.

French listeners have been shown to exploit a phrase-level metrical pattern in their language. In French, the last syllable of a phrase or group of words is lengthened and has special prominence. Listeners use this lengthening as a cue to a word end, interpreting ambiguous sequences like [ka.gaq] as two words (bas, gage ‘low, pledge’) when the first syllable is lengthened and as one word (bagage ‘luggage’) when the second is lengthened [7]). If a prosodic phrase is not utterance-final, the prominence or accenting of its last syllable is marked not only by lengthening, but also by a rise in fundamental frequency (f0), as illustrated in Figure 2a (labelled “late rise”). An f0 rise is also sometimes found near the beginning of the prosodic phrase, as illustrated in Figure 2b (“early rise”). The starting point of the optional early rise is consistently realized at the beginning of a

Figure 1: A 211-year old speech segmentation error. The intended message is l’abondance ‘abundance’.

Figure 2: Illustration of a prosodic phrase (et les gamins sages ‘and the good kids’) realized (a) without the optional early rise and (b) with the early rise.
content word (noun, adjective, etc.) ([8, 9] See also [10]).

Although other researchers have suggested that the early rise may help listeners to identify the beginning of a content word (e.g., [10], [11], this potential role had not been empirically tested. Two perception experiments were therefore conducted to explore how French listeners might use the presence and timing of intonational rises as cues to speech segmentation.

2. Experiment 1

The following hypothesis was tested:

**Presence of Early Rise as Cue to Content Word Hypothesis:** Listeners interpret the presence of an early rise as a cue to a content word beginning.

2.1. Methods

2.1.1. Materials

Stimuli were 24 minimal pairs or near-minimal pairs of phrases like the ones in (1). Each phrase contained seven syllables, with the last four syllables constituting the target. One member of the pair contained a target with two function word syllables followed by a two-syllable content word (e.g., *et mes galops* ‘and my galops’ [e.mе.га.lо]). The second member contained a target with one function word syllable followed by a three-syllable content word (e.g., *et mегало ‘and megolomanic’* [е.ме.га.lо]). An early rise beginning at the second syllable of the phrase *et mегало* is possible, since it is the first syllable of a content word. By contrast, an early rise is unlikely for *et mes galops*, since *mes* is a function word, and early rises are rarely realized on function words. Nonsense phrases were used to minimize the effect of context on segmentation preferences.

(1) a. [prodigieux et mes galops] 
   [продижъ и мез гало] ‘prodigious and my galops’

b. [prodigieux et mегало] 
   [продижъ и мегало] ‘prodigious and megolomanic’

These phrases were inserted into carrier sentences like *Ils écrivent X dans les marges ‘They write X in the margin*, read by a female native speaker of Parisian French in a sound attenuated chamber and recorded onto digital audio tape (DAT). The data were transferred to computer, where each phrase was extracted and saved as a separate file. Fundamental frequency curves were created and inspected to ensure that each phrase had been produced as intended. Two additional conditions for each pair were created through \( f_0 \) resynthesis. For the one-word segmentation items, the early rise was removed and for the two-word segmentation items, an early rise was added. There were therefore four files for each pair: two natural files and two files with resynthesized \( f_0 \), corresponding to four conditions as illustrated in Figure 3. The RMS amplitude of all 96 items was normalized. Noise was added to the normalized stimuli at a signal-to-noise ratio (SNR) of -10dB. At this SNR, listeners can hear prosodic information such as duration and intonation, but cannot identify segments (see [12] and references therein).

2.1.2. Procedures

Participants listened to stimuli played over headphones and indicated what they heard by circling one of two choices on an answer sheet (e.g., (a) *prodigieux et mes galops*, (b) *prodigieux et mегало*). Each participant heard six items in each of the four conditions. There were two short practice sessions (one without noise, one with). No feedback was given.

2.1.3. Participants

Forty native speakers of French participated in the experiment.

2.2. Results

Responses were coded according to whether a function word/content word boundary was perceived at the critical boundary, that is whether a one-word or two-word response was given. For example, *prodigieux et mes galops* is a two-word response (there is a boundary between the syllables [me] and [га] (mes | galops)) and was coded 1. The one-word response *prodigieux et mегало* has no word boundary at the critical position and was therefore coded 0. The percentage of two-word responses for each condition was calculated.

The results are shown in the graph in Figure 4. The bars represent the percent of two-word responses to each of the four conditions, defined by the speaker’s original segmentation and by presence vs. absence of an early rise. The first and the last bars represent responses to original productions while the middle bars represent responses to the resynthesized versions. The category axis shows the original segmentation (as read by the speaker) of the four syllables of the target region (one word (e.g., *mегало*) or two words (mes galops)). The y-axis shows the percentage of two-word responses, responses which included a word boundary at the critical region (e.g., *prodigieux et mes galops*). Dark bars represent items with an early rise in the target, and light bars items with no early rise.

![Figure 3: Naturally produced fundamental frequency patterns (solid lines) and resynthesized versions (dotted lines).](image)

![Figure 4: Percentage of two-word responses in Expt. 1.](image)
galops) responses. Results of ANOVA showed main effects of ORIGINAL SEGMENTATION ($F(1,39) = 112.79$, $p < 0.001$, $F(2,123) = 7.404$, $p < 0.05$) and RISE STATUS ($F(1,39) = 16.983$, $p < 0.001$, $F(2,123) = 19.748$, $p < 0.001$). The interaction between the two main factors was not significant by subjects or by items (both $p's > 0.1$). Listeners were more likely to perceive two words when the speaker had produced two words: for the two-word segmentation conditions, 56.7% of responses corresponded to the original segmentation. They were more likely to perceive one word when the speaker had produced one word: for the one-word segmentation conditions, 77.9% of responses corresponded to the original segmentation. Listeners were more likely to perceive a long content word (to give a one-word response) when the early rise was present than when it was absent. In the one-word segmentation conditions, which contained a long content word (e.g., mes galops), only 17.1% of responses incorrectly identified the stimuli as two-word productions when an early rise was present, versus 27.1% of incorrect responses when no early rise was present. In the two-word segmentation conditions, which contained a function word followed by a shorter content word (e.g., mes galops), 63.0% of responses correctly identified the segmentation of the stimuli when no early rise was present. Correct identification dropped to 50.4% of responses when an early rise was present.

2.3. Discussion
The results of the experiment support the hypothesis. Listeners used the presence of an early rise as evidence for a content word beginning. It is likely that listeners use their knowledge of the distribution of early rises as a cue in natural conversations, as they do their knowledge of other patterns.

3. Experiment 2
Experiment 2 further examined the nature of intonational cues to speech segmentation. It used non-words, whose segmentation is assumed to partially reflect that of (unknown but real) words, to test the following hypothesis:
ALIGNMENT OF EARLY RISE AS CUE TO CONTENT WORD HYPOTHESIS: Listeners interpret the low starting point of an early rise as a cue to a content word beginning.

3.1. Methods
3.1.1. Materials
Thirty-six minimal pairs with identical segment sequences were designed, as in (2). In (2a), mes ‘my’ ([me]) is followed by the nonsense word lamondines, while in (2b), [me] is the first syllable of the nonsense word mélamonde.

(2) a. mes lamondines
me la.mò.din
‘my lamondines’
b. mélamonde
me.la.mò.din
‘mélamonde’

Each critical string was four-syllables long, with the first syllable interpretable as a determiner: non [me] ‘my’ (masc. sing.), ma [ma] ‘my’ (fem. sing.), mes [me] ‘my’ (pl.), etc. Twenty non-word distractor items were also included.

The target sequences were embedded in the carrier phrase Et X pourrait être une expression utilisée par les Français. ‘And X could be an expression used by the French’ and recorded by a female native speaker of Parisian French. As predicted by the production results of Welby (in prep.) ([9]), the starting point of the early rise was consistently located at the function word/content word boundary. For example, the starting point is located between et and mé- for et mélamonde, but between mes and la for et mes lamondines (see Figure 5).

The alignment of the early rise was manipulated through fo resynthesis, producing two additional versions of each item. For function word+content word items (e.g., mes lamondines), the alignment of the early rise was made compatible with a single four-syllable content word (e.g., mélamonde): the low starting point was moved to the boundary between the first and second syllables ([f] and [me] for et mes lamondines) and the peak was moved to the end of the second syllable. For items produced with a single four-syllable content word (mélamonde), the alignment of the early rise was changed to correspond to a function word followed by a three-syllable content word: the low starting point was moved to the boundary between the second and third syllables ([me] and [la] for et mélamonde) and the peak was moved to the end of the third syllable. The synthesized versions were quite natural: participants routinely reported that they had been unaware that the stimuli had been manipulated.

3.1.2. Participants
Twenty-eight native speakers of French participated in the experiment, immediately before completing Experiment 1.[2]

3.1.3. Procedures
Participants listened to a total of 36 critical items and 20 filler items and “filled in the blank” on a response sheet with the word or words they heard between Et and pourrais. There was a Latin square design with four lists, each containing nine items in each of the four conditions and the same 20 fillers. There was a short practice session.

3.2. Results
Responses were coded for the presence of a word boundary between the second and third syllables: 1 for a response containing a boundary (e.g., (et) mes lamondines or (et) me lamondine)

Figure 5: Identical segment strings with different alignment patterns. The pattern in (a), with a L elbow at the [me]/[la] boundary is consistent with the interpretation of [me] as the function word mes ‘my’ (et mes lamondines ‘and my lamondines’). The pattern in (b), with an elbow at the [f]/[me] boundary is consistent with the interpretation of [me] as the first syllable of a content word (et mélamonde ‘and mélamonde’).

2Data for an additional fourteen participants were excluded from the analysis. For two participants, examination of language background questionnaires revealed that they were not speakers of Hexagonal French. Data for an additional 12 participants who gave all or almost all one-word (mélamonde) or two-word (mes lamondines) responses were excluded from the analysis. These participants had clearly adopted one segmentation and clung to it throughout the experiment.
Percentage of responses with a word boundary was calculated. Results are shown in Figure 6.

Analyses of variance (ANOVA) were performed. There were two factors, each with two levels: POSITION OF RISE STARTING POINT (SYL 1/2 BOUNDARY, SYL 2/3 BOUNDARY) and ORIGINAL SEGMENTATION (1 WORD, 2 WORDS). The dependent measure was the arcsine of the percentage of responses with a word boundary between the second and third syllables. There were main effects of ORIGINAL SEGMENTATION \( F(1,27) = 12.392, p < 0.01 \), \( F(2,35) = 10.684, p < 0.01 \) and POSITION OF RISE STARTING POINT \( F(1,27) = 94.732, p < 0.001 \), \( F(2,35) = 79.320, p < 0.001 \). The interaction between the two main factors was not significant by subjects or by items (both \( p’ s > 0.1 \)). Listeners were more likely to perceive a word boundary at the critical second/third syllable boundary when the speaker had actually produced a boundary (e.g., \( et \) mes lamondines). They detected a boundary 58.5% of the time in those cases versus only 48.8% of the time when she had produced a single word (e.g., \( et \) mélamondine). They perceived a word boundary at the critical region (giving responses like mes lamondines) 66.7% of the time when the starting point of the early rise was located at the later position, but only 40.7% of the time when it was located at the earlier position.

3.3. Discussion

The results provide strong support for the hypothesis. Given ambiguous sequences of syllables, listeners use the alignment of the low starting point to a syllable boundary as a cue to the beginning of a content word. Although the early rise is not consistently present in all or even most prosodic phrases (recall that it is optional), when it is present, the alignment of the low starting point of the rise is consistent, and listeners exploit this consistency in speech segmentation.

4. Conclusions

The results add to our knowledge of the range of cues used in speech segmentation. The presence and alignment of the early rise serve as powerful cues to speech segmentation in French. The results also have an application to the development of speech technologies. A better understanding of segmentation strategies can improve automatic speech recognition systems, which lag far behind humans in segmentation accuracy. The problem is particularly acute in noisy conditions, so that intonational cues, which are robustly conveyed, even in noise, should be particularly valuable.

5. Acknowledgements

I thank the following for their help with various aspects of the project: Hélène Balladur, Mary Beckman, Mariette Bessac, Chris Brew, Marie-Agnès Cathiard, Saveria Colonna, Robin Dautricourt, Ludovic Gourgas, Keith Johnson, Christiane Laeufer, Hélène Levenbruck, Amandine Penel, Stephanie Shaheen, Shari Speer, Anne Vilain, and the Phonies and Psychoids discussion groups at Ohio State. I gratefully acknowledge support from the Institut de la Communication Parlée, whose generosity in hosting me greatly facilitated the carrying out of the experiments. I am grateful for financial support from the Graduate School, the Office of International Affairs, and the Department of Linguistics of the Ohio State University. This material is based on work supported by the National Science Foundation under Grant no. 0234667.

6. References