The Structure of French Intonational Rises: A Study of Text-to-Tune Alignment

Pauline Welby

Institut de la Communication Parlée, CNRS UMR 5009, INPG, Univ. Stendhal, Grenoble, France
welby@icp.inpg.fr

Abstract

A production study examined the structure of French intonational rises. Prosodic phrases with a two-rise pattern (LHLH) were most common. Phrase length, expressed either in number of syllables or in clock time, was the best predictor of the realization of the two-rise pattern. Several other patterns were observed, including one not reported in the literature. I argue, following [6], that the early and late rises are structurally different: the LH of the late rise is a bitonal pitch accent, while the the LH of the early rise is a bitonal phrase accent. I revise my account ([9]) of the association of the L of this phrase accent.

1. Introduction

> [D]ans aucune langue il n'est aussi facile de déterminer la place de l'accent que dans le français. . . .
>
> 'In no other language is it as easy to determine the position of the accent as in French. . . .'
>
> Friederich Diez, writing 1838-1843, quoted in [3]

Perhaps the most striking characteristic of French prosody to speakers of Germanic languages (like Diez) is that French words seem to be stressed on their last syllable, a situation unlike that of unpredictable word stress in the Germanic languages and in other Romance languages. Yet the characterization of French intonation turns out to be a problem complex enough to have interested many researchers.

French intonation is characterized by an obligatory fundamental frequency (f_0) rise on the last syllable of a phrase (Fig. 1a) that is not utterance final and an optional early rise occuring somewhere before the late rise (Fig. 1b). The late rise is a marker of the "primary accent" and is accompanied by syllabic lengthening. The early rise is a marker of the "secondary accent" and may accompanied by strengthening of the onset consonants. The two pronunciations do not differ in meaning.

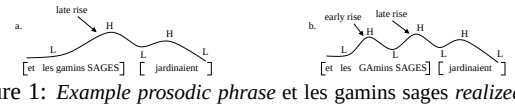

Figure 1: *Example prosodic phrase* et les gamins sages *realized with (a) a late rise and (b) an early rise and a late rise. A gloss for the sentence is* 'And the good kids were gardening'.

Although there is general agreement that French prosody includes a primary and a secondary accent (although not all researchers use these terms), there is considerable disagreement about the structure of these accents. I review three well-known accounts of French prosody and some of the predictions made by these accounts, before describing the current study.

1.1. Di Cristo and Hirst

Di Cristo and Hirst's model assumes three basic units: the Tonal Unit (TU), the Rhythmic Unit (RU), and the Intonation Unit (IU) ([4, 1, 2]). Tonal Unit boundaries are determined by three post-lexical grouping rules that account for the primary and secondary accent and prevent stress clash. TUs are delimited by a L(ow) tone to the left and a H(igh) tone to the right. Both the early rise and the late rise are Tonal Units. Rhythmic Units, which are not tonally defined, account for the observation that the primary accent, but not the secondary accent, is accompanied by syllabic lengthening. RU-final syllables are lengthened. Intonation Units are delimited by a L tone to the left and by a L or H tone to the right. An example is given in (1).

(1) $|\,\{(\text{Sa SE})_{TU}\,(\text{crétaire})_{TU}\}_{RU}\,\{(\text{m'a te})_{TU}\,(\text{léphoné})_{TU}\}_{RU}\,|_{IU}$
'His/her secretary phoned me.'

1.2. Jun and Fougeron

Jun and Fougeron (J&F) argue for an Autosegmental-Metrical (AM) account of French intonation ([6]). In the AM framework (see, for example, [7]), H(igh) and L(ow) tones are associated to metrically strong syllables or moras (pitch accents, marked with a '*') or to the boundaries of prosodic units (boundary tones or phrase accents (collectively, edge tones)). The f_0 of non-specified stretches is determined by interpolation. In the J&F account, the early rise and the late rise together form the Accentual Phrase (AP), which has the default structure /LHiLH*/. The early rise (LHi) is a phrase accent with an association to the left edge of the AP. The late rise is a LH* pitch accent whose H* tone is associated to the last full syllable of the AP and whose preceding L is unassociated. The model predicts five additional surface shapes formed by the absence of one or more tones of the underlying /LHiLH*/, as illustrated in Fig. 2. APs are grouped into Intonation Phrases (IPs), marked at the right edge by a L% or H% boundary tone. J&F (2002) reject the claim of

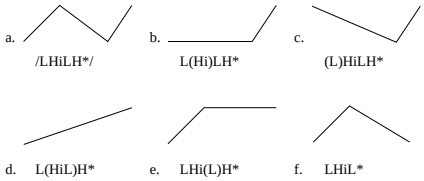

Figure 2: *Six predicted surface realizations of AP.*

an earlier version of their model ([5]) that each tone is associated to a single syllable. Although they do not report new data, this change is probably motivated by a re-analysis of the original data, which were not consistent with the claimed tone/syllable correspondence. For example, two tones were sometimes realized within a single syllable.

1.3. Post

Post (2000) proposes an Autosegmental Metrical account of the phonology of French intonation in which patterns of pitch accent distribution are accounted for by a set of ranked Optimality

Theoretic well-formedness constraints, which refer to the syntactic, morphological, and rhythmic structure of an utterance ([8]). The H tones of both the early and the late rise are treated as pitch accents (H*). The early L tone is treated as an initial boundary tone (%L), while the second L tone is described as a tone that is "inserted" between two H tones. The lowest level of prosodic grouping corresponds to a syntactically derived Phonological Phrase (PP), which is the domain of accent assignment. The larger Intonation Phrase is delimited by initial and final boundary tones and can contain one or more PPs. The tonal specifications of the model are shown in Fig. 3.

$$\left\{ \begin{array}{c} \%L \\ \%H \end{array} \right\} \; (H*(L))_0 \; \left\{ \begin{array}{c} H* \\ H+H* \end{array} \right\} \left\{ \begin{array}{c} L\% \\ H\% \\ 0\% \end{array} \right\}$$

Figure 3: *Tonal specifications of IP (Post (2000)). Parentheses indicate optionality and curly braces a set from which a tone can be selected.*

1.4. Research questions

The production study was designed to address a number of open questions in the literature. What intonation patterns are found and are these patterns predicted by various models? Do the early and late rises have the same status, as claimed by Di Cristo and Hirst, for whom both rises are Tonal Units? Are the H tones of the early and late rises pitch accents associated to syllables? Post makes this claim. For Jun and Fougeron, only the H of the late rise is a pitch accent, while Di Cristo and Hirst reject the Autosegmental Metrical assumption of association to syllables. Are the LH of the early rise and the late rise bitonal units, as claimed by Di Cristo and Hirst and by Jun and Fougeron?

2. Methods

2.1. Materials

A set of 108 sentences was designed. The following four factors were manipulated: number of syllables in target word (2, 3, 4), position of target word in sentence (initial, medial), number of preceding function word syllables (1, 2), form of definite article preceding target word (*le, la, les*). There were 27 target words (27 x 4 factors = 108). An example set with targets underlined is given in (2). An English gloss is given for sentence (c), the longest sentence of the example set.

(2) a. Le minimum sera calculé par Manon.
b. Et le minimum sera calculé par Manon.
c. Le maximum, le minimum et les écart-types seront calculés par Manon.
d. Le maximum et le minimum seront calculés par Manon.

'The maximum, the minimum, and the standard deviations will be calculated by Manon.'

2.2. Participants

Participants were 7 female native speakers of Parisian French.

2.3. Procedures

The sentences were randomized in order to avoid contrastive focus effects. Participants read the corpus out loud two times, first in a self-selected normal rate and then in a fast rate, and were recorded onto digital audio tape (DAT).

2.3.1. Data analysis

Recordings were transferred to computer using a digital-to-digital cable and downsampled to 22.05 kHz. The soundfiles were segmented and each utterance saved as a separate file. F_0 curves and spectrograms were created. After the exclusion of items containing disfluencies or f_0 perturbations, there were 790 tokens available for the analysis.

Word and syllable boundaries were tagged for each target phrase, using waveforms and spectrograms as a guide. Certain intonational features were tagged by hand: beginning f_0, late H of preceding phrases (for medial targets), H of early rise, and H of late rise. The early H tone and more rarely the late H tone were sometimes realized as high plateaux. The plateau realization of the early H conveys no apparent difference in meaning, while that of the late H often sounds continuative. The start and end points of these plateaux were labelled. The positions of L inflections or "elbows" in the f_0 curve could not be consistently determined by hand and and thus an automatic tagging procedure was used (see [11]). Time values for all tags and f_0 values were automatically extracted from the label files.

3. Results

3.1. Accent patterns found

Fig. 4 shows examples of the accent patterns found. The LHLH pattern, Jun and Fougeron's LHiLH*, was produced on 50% of the target phrases. The pattern, defined by the presence of an early and a late rise, has a clear elbow between the two H tones. Binary logistic regression analyses were performed to examine the factors influencing the realization of the LHLH pattern. Duration in clock time (milliseconds) and length in syllables were both excellent predictors of the realization of the LHLH pattern, accounting for close to 80% of the variance (for length in syllables, $R^2 = .795, \chi^2 = 292.02, p < .001$; for duration, $R^2 = .762, \chi^2 = 323.01, p < .001$). Length in syllables (but not duration) has been proposed as a predictor by some models.

The LH pattern, the L(HiL)H* pattern of J&F, was produced on 20% of target phrases. It is defined as a rise from a L at the beginning of the phrase or content word to a H in the last full syllable of the phrase The LLH pattern, the L(Hi)LH* of J&F, accounted for an additional 18%. This pattern has a H on the last full syllable of the phrase, with a leading low plateau ending at an f_0 elbow in a non-initial content word syllable (late in the phrase). The LHH pattern, J&F's LHi(L)H*, occurred in only 4.5% of cases. The pattern is defined by a high plateau stretching from one of the first two syllables to the final full syllable, with no intervening late L elbow.

Jun and Fougeron (2002) predict that the pattern (LHi)LH*, where there is undershoot of both tones of the initial LHi, should not be found: "...one of [the tones of the LHi] must be realized to make the AP initial boundary." Though rare, this pattern is attested in the data, accounting for 2% of cases. The pattern, was labelled the L2H, is characterized by a fall from the late H of a preceding phrase to a L in the last syllable of the target phrase, followed by a rise to the late H.

Another rare pattern was the LHi pattern, in which there is only an early rise, but no late rise. The LHi has a rise from a L at the content word beginning to a H on one of the first two content word syllables. The timing and duration facts make it clear that this is a bitonal pattern, not a LHL pattern: while there is a fall from the early (only) H of the AP, the fall ends at a L in the following AP. In addition, as expected, the last syllable of these phrases is lengthened, although there is no late rise. This pattern is appropriate for non-final elements of a list.

3.2. Text-to-tune alignment of the early and late rises

The text-to-tune alignment of the target phrases, the timing of important parts of the intonational contour with segmental land-

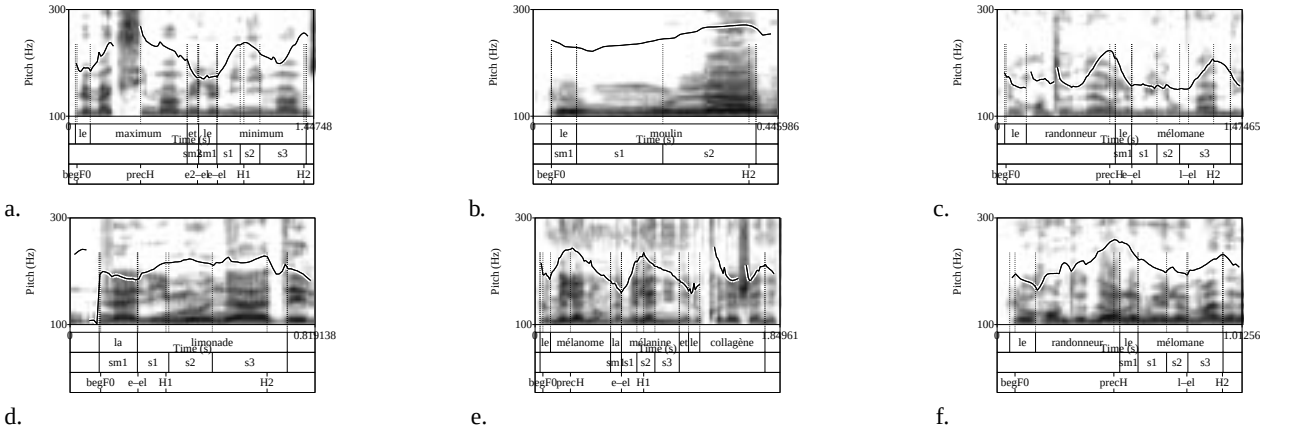

Figure 4: *Examples of accent patterns found. (a) LHLH, (b) LH, (c) LLH, (d) LHH, (e) LHi, (f) L2H. Syllables are labelled only for target phrases (*s1-s4 *for content word syllables;* sm1, sm2 *for function word syllables). Tags on the tone tier are* precH*, peak of preceding phrase;* begF0*, f0 at beginning of utterance;* H1*, peak of early rise;* H2*, peak of late rise;* e-el*, elbow of early rise;* e2-el*, additional early elbow;*l-el*, elbow of late rise.*

marks such as syllables, was examined. This information was assumed to provide information about the intonational strucure. As Pierrehumbert and Beckman (1988) ([7]) put it:

> The [phonetic] realization process does not make arbitrary use of the information in the phonological representation; instead, it uses the information in particular ways, which give rise to generalizations supporting one sort of representation over others. (p. 160)

3.2.1. L elbow of the early rise

The early L elbow was consistently realized at the function word-content word boundary (as shown in the histogram in Fig. 5a), with an optional low plateau stretching leftward from the Low elbow. Welby (2002) argued that this L elbow is doubly associated to the left edge of the phrase and to the left edge of the first content word ([9]). A pattern in the current data lead to a revision of the formulation of this double association. In some medial phrases, the early L plateau does not stretch to the left boundary of the phrase, but stops at another boundary (as in Fig. 4a). To explain such cases, we propose that the L of the early rise is *edge-seeking*. It is associated to the left edge of the first content word of the phrase and to the left edge of the phrase. When this second association is unachievable, it defaults to the nearest syllable boundary.

A stepwise multiple regression analysis was conducted to see what variables contributed to the alignment of the early L. The dependent variable was *latency from the left edge of the last function word syllable*. The model accounting for the greatest amount of variability in the training data included the independent variables *duration of last function word syllable* and *position* and accounted for 59% of the variation ($F = 382.13, p < .001$). The vast majority of that variation (58%), however, is accounted for by a model with *duration of last function word syllable* as the only independent variable ($F = 715.67, p < .001$). Although I had suspected that elbows in the fast rate or those in utterance-medial position might be less precisely placed than elbows in the normal rate or those in initial position, the results shows that the position of the L elbow is quite stable, across speakers, speaking rates, and position in the utterance.

3.2.2. H of Early Rise

The H of the early rise was produced in the first content word syllable 46% of the time and in the second content word syllable

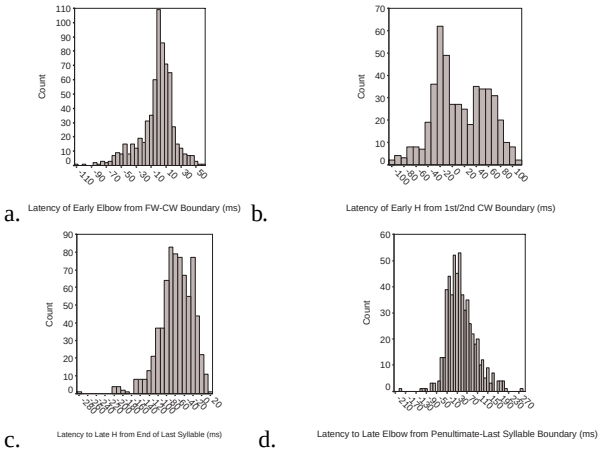

Figure 5: *Histograms illustrating the tonal alignment of the tones of the early and late rises.* CW *stands for* content word *(noun, etc.) and* FW *stands for* function word *(determiner, etc.).*

54% of the time (as shown in Fig. 5b). The position of the early rise peak has been reported to be less predictable than that of the late rise, typically appearing in the first, second or sometimes even third syllable of the phrase. Unlike the early elbow, the position of the early H cannot be specified with respect to a syllable boundary.

3.2.3. Timing of early H with respect to preceding L

Since the L tone of the early H is clearly associated to an edge, we investigated the possibility that the early H tone was the trailing tone of a bitonal edge tone. If this is the case, according to traditional AM assumptions, there should be a fairly invariant distance between the two tones. An examination of the distances between the two tones revealed no such invariance—rise time varies from under 50 ms to nearly 300 ms.

3.2.4. H of late rise

The late H was realized late in the last full syllable of the phrase or, rarely, just beyond it (3% of the time in the current data), as shown in Fig. 5c. The late H is realized an average of 22 seconds later in the syllable for phrases with the LHLH than for phrases with fewer tones. This later placement of the H tone

is perhaps attributable to tonal crowding—the late H may be pushed farther to the periphery to allow realization of all four tones. Results of stepwise multiple regression analyses showed that a model with independent variables *duration of the last syllable of the AP* and *IS-LHLH* accounted for 74% of the variance in H placement ($F = 807.29, p < .001$).

3.2.5. Late L elbow

The position of the late L elbow is reported in the literature to be less stable than that of the late H, appearing sometimes in the same syllable as the late H and sometimes in the penultimate syllable. In the data, the late L elbow (l-el) was realized in the last syllable 82% of the time and in the penultimate syllable the remaining 18% of the time, as shown in Fig. 5d. Inter-speaker differences were also apparent: some speakers almost never realized the late elbow in the penultimate syllable, while one speaker produced half of her late elbows in the penultimate syllable. No speaker produced the majority of her late L elbows in the penultimate syllable. Each speaker's data were examined to see whether the location of the late elbow could be explained in terms of some syllable landmark of the penultimate syllable of the phrase. There was, however, no correlation for any speaker between duration of the penultimate syllable of the AP and latency from either the left or the right edge of the penultimate syllable. There was also no convincing evidence that the position L of the late rise could be explained in terms of the duration of the last syllable of the phrase.

It is possible that the different timings of the L of the late rise may convey pragmatic differences, with later elbows conveying affirmativeness or indignation, as discussed in [8]. Another possibility is that since the early L and the late H mark edges, their timing is more constrained and less free to vary than that of the early H and the late L elbow, which are phrase-internal.

3.2.6. Timing of late L with respect to following H

As with the early rise, it is clear that there is no invariant rise time for the late rise—rise time varies from under 50 ms to over 300 ms.

4. Discussion and Conclusions

The results lead to a number of conclusions about the structure of the early and late rises. First, while the early rise and the late rise can each be described as a L H tone sequence, the structural similarity ends there. The alignment of the LH sequences of the two rises differ considerably, as described above. In addition, the H of the early rise is often realized as a plateau, with no change in meaning. The Di Cristo and Hirst account explicitly predicts only a duration difference between the early rise and the late rise, accounted for by the Rhythmic Unit. In that model, the early and late rise are tonally identical.

The evidence also argues against Post's claim that the H of the early rise is a pitch accent. The H of the late rise has the structure of a pitch accent, since it is associated to the last full syllable of the prosodic phrase. In addition, the late H can only be realized on full (non-schwa) syllables, in line with cross-linguistic observations that pitch accents are associated to metrically strong syllables. Yet, as we have seen, the early H shows no evidence of an association to a particular syllable.

The results of the current experiment strongly support the hypothesis that the L of the early rise is an edge tone (part of a compound edge tone) with a double association to the left edge of the first content word and to the left edge of the prosodic phrase or another syllable boundary. Post's (2000) distinction between a L boundary tone (%L in her notation) anchoring the early H and another type of L preceding the late H is supported.

The results support an Autosegmental Metrical model along the lines of the model in [6], in which the early rise is treated as a bitonal phrase accent (LHi) and the late rise is as a bitonal pitch accent (LH*). The behavior of the two interior tones of the /LHiLH*/, however, leads to a revision of some assumptions about the nature of the structural units. Standard AM models assume that the tones of a bitonal pitch accent lead or trail their starred tone at a fixed distance. There is no such invariance for the LH of either the early or the late rise. Yet there is reason to believe that the two rises are actually bitonal LHs. Although the position of the L of the late rise varies considerably, this L always realized near the right edge of the AP. One never finds, for example, cases in which the late L is realized just after the early H, in the middle of the AP rather than at the end. The L of the late rise may act to highlight the starred tone. Similarly, the results of a series of perception experiments ([10, 11]) showed that both tones of the LHi sequence of the early rise play a role in marking the left edge of the content word. An early rise (LHi) was perceived as a cue to a content word beginning, but so was an early L elbow, in the absence of an early rise (as in Fig. 4c).

Note that the distinction between phrase accent (LHi of the early rise) and pitch accent (LH* of the late rise) is purely a structural one: both bitonal units mark phrase edges.

5. Acknowledgments

I thank Hélène Lœvenbruck for her help with designing and carrying out the experiment and for valuable discussions, Mary Beckman for her helpful comments on previous write-ups of this experiment, and Hélène Balladur for her help with preparing the experimental materials. I also thank the experiment participants. I gratefully acknowledge support from the Institut de la Communication Parlée, whose generosity in hosting me greatly facilitated the carrying out of the recordings. 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

[1] Di Cristo, A., 1999. Le cadre accentuel du français contemporain : essai de modélisation (première partie). *Langues*, 2(3):184–205.

[2] Di Cristo, A., 2000. Vers une modélisation de l'accentuation du français : seconde partie. *French Language Studies*, 10:27–44.

[3] Fonagy, I., 1979. L'accent français : accent probabilitaire. *Studia Phonetica*, 15:123–133.

[4] Hirst, D. and Di Cristo, A., 1996. Y-a-t-il des unités tonales en français? In *XXIème J.E.P*, 223–226, Avignon.

[5] Jun, S.-A. and Fougeron, C., 2000. A phonological model of French intonation. In A. Botinis, editor, *Intonation: analysis, modelling and technology*. Kluwer, Boston.

[6] Jun, S.-A. and Fougeron, C., 2002. Realizations of accentual phrase in french. *Probus*, 14:147–172.

[7] Pierrehumbert, J.B. and Beckman, M.E., 1988. *Japanese Tone Structure*. Cambridge, MA: MIT.

[8] Post, B., 2000. *Tonal and phrasal structures in French intonation*. Thesus, The Hague.

[9] Welby, P., 2002. The realization of early and late rises in French: a production study. In B. Bel and I. Marlien, editors, *Proc. Speech Prosody 2002*, 695–698, Aix-en-Provence.

[10] Welby, P., 2003. French Intonational Rises and their Role in Speech Seg Mentation In *Proc. Eurospeech*, 2125–2128, Geneva.

[11] Welby, P., 2003. The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation. PhD thesis, The Ohio State University.