Perceptual investigation of prosodic phrasing in French

Laury Garnier1,2, Lorraine Baqué3, Anne Dagnac2, Corine Astésano1
1 URI Octogone- Lordat (E.A 4156), Université de Toulouse, UT2J, France
2 CLE (UMR 5263), Université de Toulouse, CNRS, UT2J, France
3 Laboratori fLexSem, Universitat Autònoma de Barcelona, Spain
laury.garnier@univ-tlse2.fr, lorraine.baque@uab.cat, dagnac@univ-tlse2.fr, astesano@univ-tlse2.fr

Abstract

The aim of this paper is to investigate prosodic phrasing and more precisely the use of prosodic cues in the marking of morphosyntactic units in French. As a first step towards this goal, a perception study was conducted on 27 listeners, who had to perform 3 distinct perceptual tasks on 32 syntactically controlled phrases read by a female speaker: a prominence strength judgment task, a boundary strength judgment task, and a task where listeners had to choose between 4 different phrase groupings intended to reflect the potential choices of prosodic phrasing. The corpus consists of syntactically ambiguous structures manipulating high and low adjective attachment on 2 coordinated nouns. It was designed to specifically test the role of prominence and boundary cues in the marking of prosodic constituency. Our results show that listeners use prosodic cues to discriminate between the two syntactic structures, with boundary cues being more readily used to capture morphosyntactic structuring. More interestingly, our results indicate that prominence and boundary cues are used to distinguish finer-grained grouping levels than those predicted by traditional descriptions on French prosodic structure.

Index Terms: prosodic phrasing, French, perception, prominence, boundary, grouping.

1. Introduction

One of the most important functions of prosody is to segment the speech flow into organized units. Prosodic cues such as prominence and boundaries are used to structure speech and partake in what is referred to as ‘prosodic phrasing’. In doing so, prosodic cues help linearize constraints arising from other linguistic levels such as semantic-pragmatic constraints and syntactic rules [1; 2]. A long line of research has been designed specifically interested in establishing the relationship between prosody and syntax. While most prosodists do not deny some interdependencies between prosody and syntax, it is arguably difficult to find common grounds on a precise relationship between the two linguistic components. Whether one assumes a mapping between syntax and prosody, whereby prosodic cues mark morphosyntactic units, or not, it is difficult to find a clear consensus on the levels reflecting prosodic hierarchy.

Indeed, there is a wide array of propositions ranging from 2 levels (Accentual Phrase or ap, and Intonational Phrase or IP; [3]) to 5 levels (including according to models: Prosodic/Phonological Word or pw, Clitic Group, Phonological Phrase or PP, Intonational Phrase and Utterance; [4]–[7]). Some authors, however, propose to envisage the possibility of recursive prosodic units in order to better capture the link between prosody and syntax [8; 9]. In these proposals, Super-Major Phrase (corresponding to an IP containing several IPs) can be found to posit an intermediate level between the Utterance and the Major Phrase. This recursion principle has been captured in more recent propositions revising the Strict Layer Hypothesis in the frame of the Optimality Theory [10; 11]. Also, a revision of the initial model by Selkirk [4] proposed to divide the PP level into 2 distinct levels: a Minor PP and a Major PP [12], equivalent to the ap and the Intermediate Phrase (ip) levels, respectively, as proposed in [13]. Discussions on the existence of an intermediate level in French have also arisen, with propositions that the ip exists for specific syntactic structures, such as tag questions and dislocated structures [14] (see also the ‘segment d’unité intonative’ in [15]). Further evidence points towards a true intermediate level applying to longer constituents and situated between the ap and the IP in French [16], more in line with the initial proposition in [13].

For French, this question about which levels are necessary to account for prosodic phrasing is of particular interest. While stress and boundary tones are clearly dissociated in English, (final) stress is syncretic to boundary tones in French, both aligning to the right edge of prosodic domains. Stress in French is also said to be post-lexical, marking the ap rather than the pw. Both these characteristics have led some authors to question the phonological status of the final accent (FA) and to characterize French as a ‘language without accent’ [18] or a ‘boundary language’ [19; 20]. Following up on this view, FA is described as a pitch accent (H*) found to mark the lowest level of the hierarchy (ap), which however disappears in favor of the boundary tone (H%) at the higher (IP) level of the hierarchy [3]. Thus, intonation pre-empt stress at higher levels of the prosodic hierarchy. Another type of accent has also been described to mark the ap level in French: the initial accent (IA), said to be secondary and optional as opposed to the primary FA. IA has been described as a mere rhythmic device occurring on longer constituents (see among others [19; 21; 22]). Descriptions of its role in the marking of the prosodic hierarchy are scarce and its functions not well established. Its functions were specifically studied on a corpus controlling for constituents’ lengths and syntactic structure [23]. This acoustic study revealed that the function of IA is more one of structuration than rhythmic balancing. It was also shown to

1 pw sometimes refers to the Clitic Group of [5; 6], hence similar to the ap. Here, we refer to the sense of [5; 6] and [17], corresponding to the Lexical Word.
more readily mark structure than FA at lower levels of prosodic constituency, close to the AP and possibly the PW. A perception study on a subset of the same corpus confirmed these results [24]. They showed that French listeners can perceive IA and FA independently from boundaries, with IA being perceived as consistently stronger than FA throughout the prosodic hierarchy.

The present study is a follow up of [24]. It is not intended to address syntax/prosody mapping per se; rather, it is specifically designed to tackle the issue of which levels of prosodic structure are necessary to account for prosodic phrasing in French. Perception is used here as an interface between the acoustic signal and phonology, in line with propositions in [25]–[27]. It is also used to circumvent the variability or ‘flexibility’ of prosodic cues found in the signal to instantiate constituency at predicted similar levels, as described in [1; 2; 28].

2. Method

2.1. Corpus

The linguistic material used for this perception study derives from the Edinburgh Corpus [23]. This corpus is composed of syntactically ambiguous structures that can be disambiguated via prosodic cues (prominences, boundary tones, pauses). These structures are composed of two coordinated nouns (N1 & N2) and an adjective (A). Syntactic ambiguity is created by manipulating the adjective scope, with low or high syntactic attachment of the adjective, yielding 2 syntactic conditions and the following phrasing (with predicted structure’s depths):

**Condition 1** - The adjective qualifies only the second noun ([N2] A), predicting a low syntactic attachment of the A with an Intermediate Phrase (ip) boundary between N1 and N2 and a Prosodic Word (pw) boundary between N2 and A.

**Condition 2** - The adjective qualifies both nouns ([N1] [N2] A), predicting a high syntactic attachment of A with an Accentual Phrase (ap) boundary between N1 and N2 plus an Intermediate Phrase (ip) boundary between N2 and A.

The prosodic structure proposed here is composed of 3 levels under the IP: ip, ap and pw. While ap is largely accepted in French, ip is more controversial and pw is hardly ever mentioned. However, in earlier acoustic/perception studies [23; 24] the question of the relevance of the pw unit was raised. The present study aims at further investigating this finer granularity in the prosodic hierarchy. The corpus is also manipulated with regards to constituents’ length; nouns and adjectives increase from 1 to 4 syllables, in all possible combinations. N1 and N2 lengths always co-vari. The original corpus is composed of 4 sets of phrases. Each phrase is extracted from a carrier sentence, the syntactic structure of which enforces a major prosodic boundary after A, realized as a non-terminal, high IP boundary. Morphology and semantics are controlled so that the adjective can apply to both nouns (see [23] for more details on the corpus).

In this perception study, one set of phrases read by one female speaker was used: 32 phrases with 4 lengths of N and 4 lengths of A in the 2 different syntactic conditions. Our data analysis however excludes phrases with monosyllabic words, because they do not allow for the distinction between initial accents (IA) and final accents (FA). Our results were thus computed on 24 phrases, namely 12 phrases by condition.

2.2. Participants and experimental tasks

Each sequence N1+N2+A was perceptually judged by 27 French native listeners. They performed 3 perception tasks:

**Boundary task:** listeners had to judge the degree of break between each word (5 potential sites) on a scale from 0 (no break) to 3 (strong break).

**Prominence task:** listeners had to judge the degree of salience of each syllable of the phrases (ranging from 6 to 15 potential sites depending of the combination of constituents’ length), on a scale from 0 (no salience) to 3 (strong salience).

**Grouping task:** listeners had to judge how the sequences N1+N2+A are divided into groups by selecting between 4 different phrase groupings (G1, G2, G3, G4) intended to reflect the possible choices of prosodic phrasing (see Figure 1).

![Figure 1: The 4 groups proposed in the Grouping task.](image-url)

The order of presentation of the 3 tasks was counter-balanced between listeners. For each task, phrases were presented via headphones on a computer and their presentation was randomized across listeners. Participants could listen to the same phrase up to 5 times in order to perform their scoring, by pressing a ‘play’ button on the computer. A training period was undertaken before the start of the experiment, using 4 phrases from another set of N1+N2+A sequences than the one used for the present perception study.

2.3. Predictions

As exposed in the introduction, our perception study is designed to clarify the role of prosodic cues in the marking of morphosyntactic units.

In the first two perception tasks, boundaries and prominences are scored separately to capture the relative effect of these two distinct phonological events on phrasing. In the third task, groups are used to more specifically investigate prosodic phrasing strategies and to help capture the underlying factors explaining potential mismatches with syntax. G1 and G2 correspond to the predictions of the syntactic Conditions 1 and 2. By contrast, G3 and G4 were chosen to propose alternative groupings for Conditions 1 and 2, so as to help uncover potential linearized boundary levels. G3 can be found in both syntactic predictions, but would reflect different relative boundary strengths in Condition1 (N1 || N2) and 2 (N1 | N2 || A). G4 may also correspond to both conditions but cannot reflect gradual boundary levels within the IP. G3 and G4 were also intended to get perceptual insight on previous acoustic results [23] on the same corpus, showing preferential marking of structure by IA over FA. G3 will be used to perceptually test these findings, where IA was found to largely mark N2 (83%) in Condition2 (after a predicted ap boundary), and marginally mark A (23%) in Condition1 (after a predicted pw boundary). G4 is intended to test the potential perception of no boundary between constituents, despite the common IA marking of N2 and A.

Ultimately, the choice of one grouping over another will be correlated to the results emerging from the boundary and prominence tasks, with the goal to clarify the respective or conjoined implication of these prosodic phenomena in speech structuration.

1154
2.4. Data selection and analysis

The design of our data supposes that we first investigate listeners’ interpretation of Conditions 1 and 2 in terms of the 4 grouping choices.

2.4.1. Syntactic condition effect on Groups

The relationship between Conditions and Groups was investigated using a chi-square test of independence. Results show a significant effect of Condition on Group ($\chi^2(3,648) = 530.89, p < 0.001$). Post-hoc analyses show that G1 is highly associated to Condition1 (93.5% vs. G2: 0.6%, G3: 5.9% and G4: 0.0%; adjusted p-value < 0.001). While Condition2 is significantly more frequently associated to G4 (42.9%), G2 (32.1%) and G3 (21.0%) than to G1 (4.0%; adjusted p-value < 0.001).

For lack of data, we thus decided to exclude from further analyses the cases in which Condition1 was perceived as G2, G3 or G4, and those in which Condition2 was perceived as G1. Listeners are able to perceptively distinguish the two syntactic conditions: G1 massively corresponds to Condition1, while G2 is associated to Condition2. Finally, whereas we expected G3 and G4 to be distributed equally over the two conditions, they are mostly associated to Condition2.

2.4.2. Statistical models

On the basis of these results, we analyzed the effect of Group (G1, G2, G3 and G4) on Prominence and Boundary scores (ranging from 0 to 3). Constituents’ lengths were taken into account for N only, while A lengths were colluded in order to increase statistical power. Analyses focused on the sites in the phrases reflecting structuring, i.e. between N1/N2 and between N2/A. Boundary scores were measured after each content word, and Prominence scores were measured immediately before boundaries (FA on N1 and N2) and at the beginning of content words after boundaries (IA on N2 and A). Separate Mixed Linear Models were computed for dependent variables Boundary and Prominence respectively, and with subjects and phrases as random variables. The predictors were Groups and N length (2, 3 and 4 syllables). The non-significant variables or interactions were excluded from the final models.

3. Results

3.1. Boundaries

3.1.1. N1|N2 boundary scores

There was a main effect of Group (F(3,91.25)=94.4, p<0.001) and of N length (F(2,37.54)=10.59, p<0.001), and an interaction between the two predictors (F(6,89.32)=4.04, p<0.001). Post-hoc analyses show significant differences (p<0.05) of boundary scores between groups, for all N lengths’ conditions: the boundary in G1 (mean score (ms) range: 3.38-3.62) is perceived as much stronger than in the other groups (ms ranges: G2=1.44-1.83; G3=1.43-2.21; G4=1.36-1.60). In addition, the boundary in G3 for 2 and 4 syllables N is also perceived as stronger (ms: 2.10 and 2.21 respectively) than in G2 (ms: 1.58 and 1.83 respectively) and G4 (ms: 1.40 and 1.60 respectively). For 3 syllables N, however, no difference is observed between the boundaries in G3 (ms: 1.43), G2 (ms: 1.44) and G4 (ms: 1.36). Altogether, these results indicate that listeners perceive 3 different boundary levels between N1 and N2: the strongest in G1, an intermediate one in G3 and the weakest one in G2 and G4.

3.1.2. N2|A boundary scores

Our analysis shows a main effect of Group (F(3,111.64)=30.95, p<0.001). Post-hoc analyses show significant differences (p<0.05) for all comparisons except between G2 and G3. The boundaries in G3 (ms: 2.52) and G2 (ms: 2.42) are perceived as much stronger than in G4 (ms: 1.99), with the lowest boundary score for G1 (ms: 1.77). Again, it seems that listeners perceived 3 different boundary levels between N2 and A: the strongest in G2 and G3, an intermediate one in G4 and the weakest one in G1.

3.1.3. Relative strength of N1|N2 and N2|A boundaries

Prosodic phrasing is better accounted for when capturing the syntagmatic relationship between boundaries. This relationship was inferred in our data by subtracting Boundary scores perceived between N2|A to those perceived between N1|N2. Results show main effects of Group (F(3,106.96)=7.28, p<0.001) but no effect of N length. Post-hoc analyses show that the only significant contrasts (p<0.001) are between G1 (ms=-1.78) and the other three groups (ms: G2=0.96; G3=0.47 and G4=0.50). For G1, the second boundary (N2|A) is perceived as much weaker than the first boundary (N1|N2). By contrast, for G2, G3 and G4, the second boundary (N2|A) is perceived as stronger than the first boundary (N1|N2). It seems that listeners perceive two types of syntactic/prosodic grouping with a different amount of balance of boundaries: strong + weak boundaries (N1|N2 > N2|A) in Condition1 associated to G1, and weak + strong boundaries (N1|N2 < N2|A) in Condition2 associated to G2, G3 and G4.

Figure 2: Perception results as a function of N length and Groups for A) N1|N2 boundary scores’ strength, and B) Relative strength of N1|N2 and N2|A boundaries.

3.2. Prominences

3.2.1. Final and Initial accents

For FA on N1, our analysis shows main effects of Group (F(3,568.71)=6.24, p<0.001) and of N length (F(2,556.49)=8.14, p<0.001). Post-hoc analyses show that FA on N1 in G1 (ms: 2.39) is perceived as stronger (p<0.035) than in G2 (ms: 2.09) and G4 (ms: 2.17), while the difference with G3 (ms: 2.13) doesn’t reach significance (p=.087). The other contrasts are not significant. As far as N length effect is concerned, FA on N1 is globally perceived as significantly weaker (p<0.014) when N is composed of 3 syllables (ms: 2.03) as compared to nouns of 2 or 4 syllables (ms: 2.25 and 2.31 respectively). No difference is observed between N of 2 and 4 syllables. For FA on N2, our analysis shows no effect of Group and a main
effect of N length ($F(2,42.64)=3.87$, $p=.029$). Post-hoc analyses show that FA on N2 is perceived as weaker when N is composed of 3 rather than 4 syllables ($ms: 1.92$ vs $2.14$ respectively, $p<.022$). For IA on N2, there are no significant main effects of Group and N length ($p>.05$; $ms$ range: 2.26-2.80). The same goes for IA on A ($ms$ range: 2.00-2.72).

3.2.2. IA and FA relative prominence strengths

The question remains as to the relative weight of prominences in the marking of the structure. We computed a paired two-tailed t-test with IA and FA scores around the two boundary sites, for each group and independently from N length. Results show a significant difference between FA on N1 and IA on N2: IA is perceived as stronger than FA in G3 ($ms: 1.63$ vs. $1.07$; $t(67)=3.30$, $p<.002$) and in G4 ($ms: 1.56$ vs. $1.07$; $t(138)=5.55$, $p<.001$). IA is also marginally stronger than FA in G2 ($ms: 1.34$ vs. $1.11$; $t(103)=1.96$, $p<.053$). There is however no effect for G1 ($p=.338$). Our results also show a significant difference between IA on A and FA on N2 for all groups: IA is perceived as stronger than FA in G1 ($ms:1.23$ vs. $0.92$; $t(302)=4.83$, $p<.001$), in G2 ($ms:1.38$ vs. $0.99$; $t(103)=3.30$, $p=.001$), in G3 ($ms:1.45$ vs. $0.82$; $t(67)=3.80$, $p<.001$) and in G4 ($ms:1.42$ vs. $1.04$; $t(138)=4.06$, $p<.001$).

4. Discussion & Conclusion

The aim of this study was to investigate prosodic phrasing in French, and more precisely the use of prosodic cues in the marking of morphosyntactic units. Indeed, as exposed in the introduction, there is no clear consensus on how many levels are necessary to reflect prosodic hierarchy, especially for French descriptions. Investigating prosodic phrasing is best done manipulating syntactic ambiguity, as shown in the seminal work by Lehiste [29; see also 28]. Our corpus uses low and high adjective attachment to two conjointed nouns in order to elicit prosodic disambiguation in French. It also manipulates constituents’ length. If structure depth can be uncovered by means of acoustical analyses, some authors have however demonstrated the difficulty of matching acoustic features to syntactic structures [28]. Perceptual investigations of such a link might thus be a better gateway towards this end [25]–[27], a useful interface between acoustic cues and phonology. Our participants were asked to perform three separate perceptual judgments: two tasks (Boundary and Prominence strengths’ scoring) were designed to investigate the role of these prosodic cues in structure marking. Our statistical design aimed at interpreting these results with regard to the Grouping task, in order to more specifically test their relationship to prosodic phrasing.

Taken together, our results indicate that listeners are able to distinguish the two intended syntactic conditions: while Condition1 was exclusively associated with G1, Condition2, however, was distributed across G2, G3 and G4. Prosodic phrasing is thus perceived with different degrees of granularity in Condition2. This granularity could be investigated by testing the effect of each Grouping choices on Boundary and Prominence scores independently. Altogether, our results indicate that morphosyntactic structuring is better accounted for by Boundary than Prominence perception. Indeed, while Grouping choices very marginally affect Prominence perception, they largely explain Boundary scores. Indeed, Grouping choices indicate that listeners perceive various Boundary levels on each site (N1 | N2; N2 | A). However, a syntagmatic investigation indicates that listeners adjust relative boundary strengths in order to maintain the intended syntactic structure: strong + weak boundary (N1/N2 > N2/A) in Condition1 and weak + strong boundary (N1/N2 < N2/A) in Condition2. Confronting the syntagmatic analysis with the analysis by site allows for finer interpretation of the granularity of these two boundary strengths: listeners perceive up to three boundary levels on each site below the IP level. Between N1 and N2, the strongest boundary perceived in G1 (N1/N2 > N2/A) could reflect an ip boundary. The weaker one perceived in G3 (N1/N2 < N2/A) would correspond to an ap boundary. The lowest boundary is perceived in G2 and G4 (N1/N2 < N2/A), implying potentially a pw boundary. Between N2 and A, the strongest boundary perceived in G2 and G3 (N1/N2 < N2/A) could reflect an ip boundary, and the weaker one perceived in G4 (N1/N2 < N2/A) could correspond to an ap boundary. The lowest boundary perceived in G1 (N1/N2 > N2/A) could also reflect a pw boundary. These results are particularly interesting when considering the alleged propositions for French prosodic constituency. In most theoretical descriptions, indeed, French accent is said to be post-lexical, which makes it impossible to recognize a boundary level lower than the ap, namely the pw level close to the lexical word [5; 6; 17]. Our results however indicate that such a constituency level is actually perceived: between N2/A in G1 and between N1/N2 in G2 and G4.

Our investigations on Prominence also confirm these assumptions. Two main results emerge: first, FA was clearly used to mark structures perceived as G1 at the ip boundary, which can be interpreted as an independent and “supporting role” [30] to the strong boundary scores found at this site. This result questions the proposition in [3] according to which boundary tones (H%) supersedes FA pitch accent (H*) at higher prosodic boundaries. Namely, the syncretic occurrence of FA prominence and intonation boundaries does not block the independent perception of both these prosodic events. The second important result concerns IA. The perception of IA does not depend on Grouping choices. More surprisingly with regards to the literature in French, it is not linked to constituents’ length, contradicting the hypothesis that IA essentially plays a rhythmic role in French (see among others [19; 21; 22]). Rather, our results indicate that IA is perceived as stronger than FA, with consistent high scores throughout the prosodic structure. Previous acoustic results suggested that IA preferentially marks the lowest ap level over FA [23]. The present perceptual results rather point towards preferential marking of the pw. Indeed, results show that IA after pw boundaries in G1 (N2/A), and in G2 and G4 (N1/N2) was perceived as more salient than FA (situated before the pw boundary): here, IA is a left boundary marker of constituency, starting as early as the pw level.

We wish to elaborate on these results, which are particularly interesting to address the question of prosodic phrasing in French. New analyses are currently under way on a much larger database, using the same corpus (now 4 speakers) and the same tasks run on 80 listeners.

5. Acknowledgements

This study is supported by the Agence Nationale de la Recherche grant ANR-12-BSH2-0001 (PI: Corine Astésano) and by the Spanish Ministry of Economy and Competitiveness grant FFI2013-40419-P (PI: Lorraine Baqué). The authors also wish to thank Rafèu Sichel-Bazin for useful comments on an earlier version of this text.
6. References


