Specific contribution of tonal and duration cues to the syntactic parsing of French

Amandine Michelas [1,2]*, Mariapaola D'Imperio* [2]

[1] Division of Linguistics and Multilingual Studies, Nanyang Technological University, Singapore
[2] Université d'Aix-Marseille & Laboratoire Parole et Langage, CNRS, Aix-en-Provence, France
michelas@lpl-aix.fr, mariapaola.dimperio@lpl-aix.fr

Abstract

This study addresses the question of the separate contribution of tonal and duration cues in prosodic boundary marking and their effect in syntactic parsing in French. We used pairs of Noun Phrases whose segmental structure was identical but differing in the potential placement of an Accentual Phrase (AP) or Intermediate Phrase (ip) boundary at their right edge. We artificially manipulated the acoustic cues at AP/ip boundaries to obtain a set of stimuli marked either by one type of cue (either tonal or duration cues) or by both indices. In a two forced-choice completion task (PP-choice or VP-choice), participants gave more PP responses when an AP boundary was marked, either by both cues or only by duration indices, relative to when an ip boundary was present. Specifically, for the ip-boundary level to be identified, participants needed the presence of both types of cues to correctly complete the sentence. These results offer a new understanding of the separate role of acoustic cues in prosodic boundary marking.

Index Terms: prosodic boundary, syntactic parsing, Intermediate Phrase, Accentual Phrase, preboundary lengthening, pitch accent scaling, French.

1. Introduction

A fundamental question which remains to be answered regarding the role of prosody in speech processing is how listeners are able to pick out the individual words and build larger syntactic phrases out of them. It has been shown in several languages that phrasal prosody is an obvious candidate for constraining syntactic parsing, and several experiments have shown that adults exploit it to resolve syntactic ambiguities (for English [1], [2], [3] and for French [4], [5]). While two levels of phrasing (Accentual Phrase, or AP, and Intonation Phrase, or IP) are generally accepted for French, a number of recent studies have raised the possibility of the existence of an intermediate level of phrasing (the Intermediate Phrase, ip).

[6] et [7] have shown that an ip right boundary appears to be aligned with a major syntactic break (i.e. a NP/VP break) and is marked by systematic scaling relationships between the *fo* peaks of its AP-final rises. Specifically, it appears that the ip right boundary is cued through a process blocking the iterative downstep of successive AP-final *fo* peaks in addition to preboundary lengthening.

Figure 1 shows two renditions of *La nana du sauna* ("The girl who managed the sauna") extracted from the sentence *La nana du sau**na**]AP d'Héléna devenait vraiment méchante ("The girl who managed Héléna's sauna became really nasty")* or *La nana du sau**na**]ip devenait vraiment méchante ("The girl who managed the sauna became really nasty")* where the 6th syllable is either associated to an AP boundary contained within a Noun Phrase (1a) or an ip boundary depending on the alignment with the N(oun) P(hrase)/V(erb) P(hrase) break (1b). Note that the second *fo* peak is realized at the same level as the first peak in 1b, whereas it is downstepped in 1a, and that the last syllable is longer in 1b than in 1a because of the presence of the ip boundary.

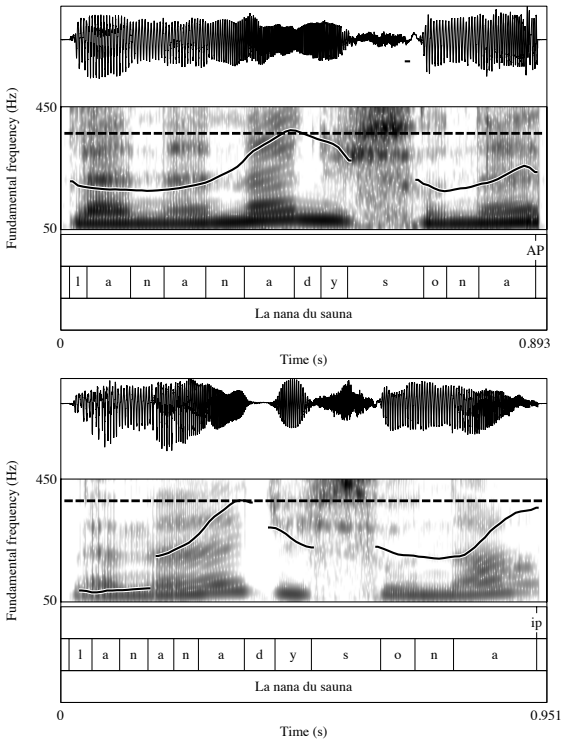

Figure 1: *fo contours for two utterances containing the NP La nana du sauna "The girl who managed the sauna" whose last syllable is either associated to an AP (1a) or an ip (1b) boundary.*

It appears that French listeners are able to differentiate prosodic cues at both AP and ip boundaries and exploit this difference in order to build expectations about the morphosyntactic categories of the upcoming syntactic phrase. [8] employed pairs of NPs whose segmental structure was identical until the 6th syllables but differed in the potential placement of an AP or ip boundary at their right edge. In a two forced-choice completion task (P(repositonal) P(hrase)-choice vs. VP-choice, they showed that participants preferentially completed the NPs by a VP (e.g. *devenait* "became") when an ip boundary was present in the stimulus relative to when an AP boundary was present**.** On the contrary, participants preferentially completed the sequences by a PP (e.g. *d'Héléna* "of Héléna") when they heard an AP boundary in the stimulus.

[8] thus shows unambiguously that the acoustic cues at AP

and ip boundaries can influence participants' syntactic processing. However, one cannot establish what type of specific phonetic and/or phonological cue intervene in this kind of processing.

Among the cues marking the location of prosodic boundaries, we find increased duration of preboundary syllables (preboundary lengthening), as well as of earlier segments within the preboundary word [9], segmental variation in preboundary or postboundary segments [10], strengthening of articulatory correlates such as linguopalatal contact or nasal airflow [11], presence of optional silent pauses, or tonal cues (including the realization of a specific intonational contour and pitch range resetting before and after the break) [12]. We also know that in some cases durational cues do not match with tonal cues in prosodic boundaries marking. For example in the English ToBI system, break index 2 is employed in order to signal a mismatch between tonal properties and the perception of an ip break. For French, the observation of both read and spontaneous speech highlights a great number of cases in which the perception of a clear boundary is not supported by tonal cues (especially at fast speech rate) [6]. Thus the relevance of each cue in prosodic boundary marking is an open question.

In this paper, we analyze the separate role of tonal and duration cues at two prosodic levels (AP and ip) in the syntactic parsing of spoken utterances within the Autosegmental-Metrical framework. In that framework, the structure of APs in French can be described by a sequence of H(igh) and L(ow) tones. The AP is minimally defined by an obligatory final rise consisting of a L plus a H tone (denoted by LH*) associated to a metrically strong syllable, which typically occurs at the right edge of the phrase [13]. Thus the AP right boundary is delimited by a pitch accent (L)H* at its right whereas the ip boundary is cued by a right-edge tone H- responsible for blocking the downstep of subsequent rising pitch accents at the ip boundary [7].

Because there seems to exist a strong correlation between duration cues and the perception of a prosodic break in English [14] and French [15], we hypothesize that both AP and ip boundaries can be minimally marked by duration cues. Such a marking would be sufficient for French listeners to perform syntactic parsing of spoken utterances. We hence replicated the experiment conducted by [8] by artificially manipulating the prosodic cues associated to AP and ip right boundaries. As shown in [8], we expected more VP responses when an ip boundary was present in the stimulus relative to when a simple AP boundary was present. We also hypothesized the same result when AP was marked only by durational cues. Alternatively, we expected more PP responses correlating with AP boundary identification.

2. Method

2.1. Speech material

20 pairs of NPs whose segmental structure was identical up to the 6th syllable but differing in the potential placement of an AP or an ip boundary at their right edge were used to build the stimuli. Consider the following French utterances:

1a. La nana]AP du sauna]AP d'Héléna]AP devenait vraiment méchante. "The girl who managed Héléna's sauna became really nasty".

1b. La nana]AP du sauna]AP/ip devenait vraiment méchante. "The girl who managed the sauna became really nasty".

If only the segmental material is taken into account, the two sequences are identical up to *sauna*. However, the prosodic characteristics of the second AP are different between the two utterances: (i) the last syllable of the second AP shows a greater degree of preboundary lengthening in 1b than in 1a and (ii) the second AP is lowered in pitch relative to the first one in sentence 1a while not in 1b (Fig. 1). Except for this difference on the last syllable of the second AP, the two sentences are strictly identical.

We also checked whether both completions (VP or PP) were plausible within the sentences. To do so, a group of 10 French participants read all experimental sentences and judged their overall plausibility on a 0 (completely implausible) to 7 (highly plausible) scale. PP completions obtained the same result as VP completions (mean of 6.1 for PP completions, 6.2 for VP completions).

A native speaker of French produced the 40 utterances thus constructed. For each utterance, we checked if the expected phrasing was produced. We then cut the utterances after the 6th syllable, which could be either associated to an AP or ip boundary. Out of the sequences thus created, a set of stimuli was built in which the acoustic indices cuing the AP and ip boundaries were artificially controlled. The 20 NPs whose target syllable was associated to an AP boundary were used as source utterances. We hence artificially modified the tonal and duration cues associated to target syllables using PSOLA (Pich Synchronous Overlap and Add) on Praat [16] in order to obtain six prosodic boundary conditions (Table 1) for a total of 120 stimuli (20 NPs x 6 prosodic conditions).

	Tonal cues	Duration cues
AP-ton	H*	Ambiguous duration cues
ip-ton	H*H-	Ambiguous duration cues
AP-dur	Ambiguous tonal cues	AP preboundary lengthening
ap-dur	Ambiguous tonal cues	ip preboundary lengthening
AP	H*	AP preboundary lengthening
ip	H*H-	ip preboundary lengthening

Table 1: *Six types of resynthesized stimuli.*

In both the AP-ton and ip-ton conditions, AP and ip boundaries were cued only by tonal cues, i.e. in the AP-ton condition the last pitch accent was downstepped relative to the preceding one, while in the ip-ton condition the last pitch accent was at the same level as the preceding one. In the two AP-dur and ip-dur conditions, prosodic boundaries were only cued by durational cues (i.e., a greater degree of preboundary lengthening for ip than for AP boundaries) and finally for the AP and ip conditions, boundaries were cued by both tonal and durational cues.

For the AP-ton and ip-ton conditions, we altered the duration of target syllables to make them durationally ambiguous. In each case we adjusted the duration of the syllable to a value that was the average of its duration for the AP and ip conditions in the NPs read by our speaker. In addition, for the ip-ton condition we manipulated the height of the final vowel in order to introduce a H- phrase accent responsible for blocking the downstep at the ip boundary.

In both AP-dur and ip-dur conditions we artificially modified durational cues of target vowels. For each sentence we adjusted the duration of the vowel to a value that was the corresponding vowel value of AP and ip conditions in NP pairs read by our speaker. We also altered the height of target vowels to make them tonally ambiguous. In each case we adjusted the height of the *fo* peak to a value equal to the average between the height of the first pitch accent and the height of the second pitch accent. Finally for AP and ip

conditions, prosodic boundaries were cued through both tonal and duration cues. For the AP condition we simply stylized the *fo* curve while we manipulated both the height of the target pitch accent (to introduce a H-) and adjusted the duration of the target syllables (at the value obtained in the NPs read by our speaker) for the ip boundary condition.

In order to enusre that the stimuli thus created did not sound too artificial, we asked 5 French native speakers to judge the naturalness of the stimuli on a 1 (completely natural) to 7 (highly unnatural) scale. The listeners judged the NPs read by our speaker as natural as the 120 stimuli artificially created by resynthesis (means of 5.1 for the natural NPs; means of 5.158 for the artificial NPs; t(290.855)=0.311, p=0.6667).

On the basis of the 120 stimuli thus created, six lists of sentences were constructed so that each stimulus created from the same utterance base appeared in a different list. Each list contained 20 stimuli and 40 fillers. The fillers were the same in all lists and were randomly cut at a word boundary. Within each list the order of presentation of the stimuli was random. Each participant heard only one list of stimuli so that he heard only one version of each stimulus.

2.2. Participants and procedure

60 native speakers of French took part in the experiment, 10 for each list. Each participant was tested individually in a quiet room. Participants were seated in front of a computer screen with headphones. Instructions written on the screen informed them that they had to listen to a sentence beginning and had to complete it by choosing one of two possible sentence segments written on the screen (either a PP or a VP). Listeners were aware that both choices were plausible but were forced to choose the most appropriate response. A trial began with the auditory presentation of a sequence and 2 seconds later the two alternative completions appeared on the screen. Participants had to press either the right or the left button to indicate their choice. Before the experiment began, participants performed an 8-item training.

3. Results

We measured both the response types given by participants (PP-choice or VP-choice) as well as their reaction times. Figure 2 presents the percentage of PP-responses in the six prosodic conditions.

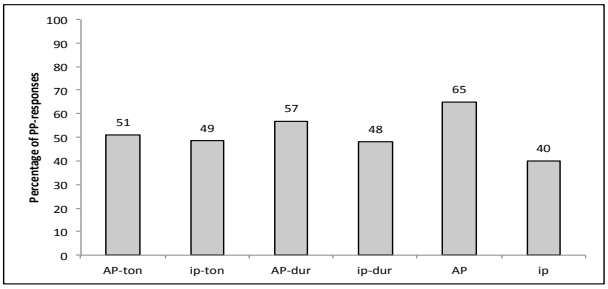

Figure 2: *Percentage of PP-responses given by participants for the six prosodic conditions (AP-ton, ip-ton, AP-dur, ip-dur, AP, ip).*

A Mixed Logit Model (MLM) was run on the responses given by participants. MLM is a statistical model for examining discrete choices with values coded as either 1 (for VP-choice) or 0 (for PP-choice). The fixed effect was the prosodic condition (AP-ton, ip-ton, AP-dur, ip-dur, AP, ip). Participants and items were included as random factors. The model included 1200 values (60 participants x 20 stimuli). The MLM results showed a main significant effect of prosodic boundary condition. Participants gave significantly more VP-choices than PP-choices for the ip-condition (β=0.5306, se=0.1914, z=2.2374, p<0.01). The ip-ton, ip-dur and AP-ton conditions had no significant effect on the response type given by participants (ip-ton condition: β=0.06352, se=0.18724, z=0.339, p=0.73442; ip-dur condition: β=0.01652, se=0.1871, z=0.88, p=0.92965; AP-ton condition: β=-0.04935, se=0.18725, z=-0.2864, p=0.792135). Finally, the AP and AP-dur conditions had a significant effect with significantly more PP-choices than VP-choices (AP condition: β=-0.6598, se=0.1926, z=-2.427, p<0.001; AP-dur condition: β=-0.412, se=0.1892, z=-2.174 p<0.01).

We also analyzed participants' Reaction Times (RTs, Figure 3). For each participant, RTs below or above 2.5 standard deviations of the mean were excluded from the analyses (4.08% of the data).

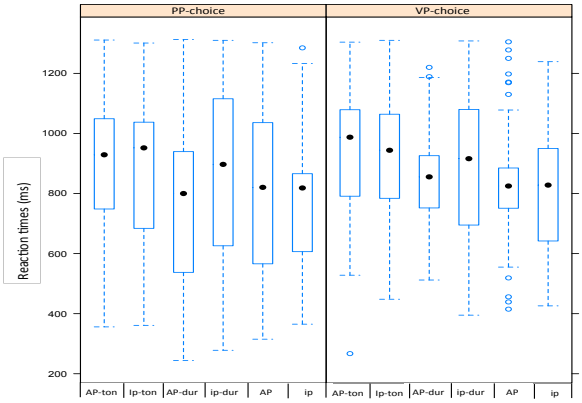

Figure 3: *Reaction times (in ms) for each prosodic boundary condition (AP-ton, ip-ton, AP-dur, ip-dur, AP, ip) relative to response type given by participants (PP-choice or VP-choice).*

A linear-mixed model (LMM) was fitted on RTs, with prosodic boundary type (AP-ton, ip-ton, AP-dur, ip-dur, AP, ip) and response type (PP-choice, VP-choice) as fixed effects. Random intercepts were included for participants and items. The model included 1151 values. The LMM results showed a main significant effect of prosodic boundary condition. RTs were significantly longer in the AP-ton (β=77.731, se=30.987, t=2.508, pMCMC<0.05), ip-ton (β=64.7463, se=29.447, t=2.189, pMCMC<0.05) and ip-dur conditions (β=59.720, se=29.889, t=2.089, pMCMC <0.05) than in the AP and ip ones. In contrast, RTs were not significantly longer in the AP-dur (β=-79.870, se=28.787, t=-0.343, pMCMC=0.738) than in the AP one. The effect of response type (β=-1.893, se=31.147, t=-0.057, pMCMC=0.9543) and the interaction between the two fixed factors were not significant (MCMC p-values were all greater than 0.05).

4. Discussion

In the experiment presented here, French listeners appeared to exploit the difference in the acoustic cues present in the stimuli (duration and/or tonal indices) marking AP and ip-boundaries to infer the morphosyntactic category of the

upcoming syntactic phrase. These results are line with those observed in [8]. More specifically, in a two forced-choice completion task, participants gave more VP responses when an ip-boundary was present in the stimulus relative to when an AP-boundary was present. The type of response (PP-choice vs. VP-choice) also lead to a number of conclusions on the specific role of tonal and duration indices in participants' response patterns. When AP and ip-boundaries were only cued by tonal indices, the prosodic boundary did not significantly influence participants' choice. On the contrary, when AP-boundaries were merely cued by duration indices, participants responded with significantly more PP-choices, as it was the case for AP boundaries cued by both types of indices. These results suggest that duration cues are sufficient for French listeners to perceive an AP boundary, which are exploited as indicating the presence of an upcoming minor syntactic break. On the contrary, for the ip level participants needed both type of cues (tonal and duration cues) to correctly infer the morphosyntactic category of the upcoming syntactic phrase.

RT analyses were in line with the results relative to response type. Specifically, participants responded more slowly when prosodic boundaries were cued only by tonal cues relative to when prosodic boundaries were marked by both duration and tonal cues, independent of prosodic boundary condition (AP or ip). Note also that participants took the same time in responding when AP-boundaries were cued by both tonal and duration cues than when they were marked only by duration cues. This was not the case for ip-boundaries, since participants responded more slowly when these were only cued by duration.

These results suggest that duration cues are more relevant than tonal cues for the purposes of AP-boundary identification. It is generally agreed that the relevance of prosodic cues is language-dependent. A number of authors have previously underlined the importance of duration in the prosodic marking of French boundaries [15] and, in line with these studies, our results support the idea that duration indices are the most relevant type of cue in French prosodic boundary marking. However our results also suggest that the relevance of each cue is not only language-dependent but might also depend on the level of the prosodic constituent in the hierarchy. In fact, although duration indices were sufficient for French listeners to infer the morphosyntactic category of words immediately following AP-boundaries, both duration and tonal indices were necessary for perceiving and exploiting the presence of an ip-boundary.

Moreover, our results bear implications for speech processing models. The question of when prosodic information intervenes within the syntactic process is still unclear in the literature. One of the primary methodologies used to address the question of the syntactic parsing of spoken utterances is to study the processing of sentences containing syntactic ambiguities. When participants are faced with ambiguous sentences that allow multiple structures to be built, the parsers encounter problems in assigning syntactic structure. Prosodic cues must be activated in the first stages of speech processing or at a later stage when the syntactic parser detects the ambiguity to be resolved. In our study, the data did not contain a clear ambiguity. When participants heard an initial NP, they were capable to build expectations about the syntactic category of the upcoming sequence. This shows that prosodic boundary cues are not only used when the syntactic parser faces an ambiguity. Rather, listeners employ those cues immediately in order to perform syntactic parsing.

5. Conclusions

Results from a sentence completion study, with synthetic stimuli containing duration and/or tonal cues to prosodic boundary marking, show that the perception of either an AP or an ip-boundary can influence French syntactic parsing. While both tonal and duration indices seem to be necessary to identify and exploit ip-boundaries as cueing major syntactic breaks, French listeners are able to capitalize on duration indices alone (when tonal cues were kept ambiguous) to identify an AP-break and interpret it as an indicator of a minor syntactic break. These results lead to a better understanding of the specific role of acoustic cues to prosodic edge marking in French, while also underlying the dependency of cue marking on both language type and prosodic level within the hierarchy.

6. References

[1] Kjelgaard, M. M., & Speer, S. R., « Prosodic facilitation and inter- ference in the resolution of temporary syntactic closure ambiguity ». *Journal of Memory and Language, 40,* 153–194, 1999.

[2] Schafer, A. J., Speer, S. R., Warren, P., & White, S. D., « Intonational disambiguation in sentence production and comprehension ». *Journal of Psycholinguistic Research, 29,* 169–182, 2000.

[3] Ferreira, F., Horine, M. D., & Anes, M. D., « Exploring the use of prosody during language comprehension using the auditory window technique ». *Journal of Psycholinguistic Research, 25,* 273–290, 1996.

[4] Millotte, S., Wales, R., & Christophe, A., « Phrasal prosody disambiguates syntax ». *Language and Cognitive Processes*, 22, 898-909, 2007.

[5] Millotte, S., René, A., Wales, R., & Christophe, A., « Phonological phrase boundaries constrain the on-line syntactic analysis of spoken sentences ». *Journal of Experimental Psychology: Learning Memory and Cognition*, 34, 874-885, 2008.

[6] Michelas, A., « Caractérisation phonétique et phonologique du syntagme intermédiaire en français: de la production à la perception ». Thèse de Doctorat de l'Université d'Aix-Marseille I, 2011.

[7] D'Imperio, M., & Michelas, A., « Embedded register levels and prosodic phrasing in French ». Proceedings of the Vth Speech Prosody Conference, 2010.

[8] Michelas, A., & D'Imperio, M., "Uncovering the role of the intermediate phrase in the syntactic parsing of French". Proceedings of the 17th International Conference of Phonetic Science, 2011.

[9] Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, F., & Price, P.-J., "Segmental durations in the vicinity of prosodic phrase boundaries". *Journal of the Acoustical Society of America*, 91, 1707-1717, 1992.

[10] Byrd, D., & Saltzman, E., " The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening". *Journal of Phonetics*, 31, 149-180, 2003.

[11] Fougeron, C., "Articulatory properties of initial segments in several prosodic constituents in French". *Journal of Phonetics,* 29, 109-135, 2001.

[12] Beckman, M.-E., & Pierrehumbert, J., "Intonational structure in Japonese and English". *Phonology Yearbook*, 3, 255-309, 1986.

[13] Jun, S.A., & Fougeron, C., "Realizations of accentual phrase in French". *Probus*, 14, 147-172, 2002.

[14] Ostendorf, M. & Veilleux, N., " A hierarchical stochastic model for automatic prediction of prosodic boundary location. *Computational Linguistics", 20,* 1-27, 1994.

[15] Delattre, P. Les dix intonations de base du français. *The French Review*, 40(1), 1-14, 1966.

[16] Boersma, P. & Weenink, D., Praat: doing phonetics by computer (Version 5.1). www.praat.org, 2009.