

Prosody and discourse structure in French: a multilayer multi-parametric analysis of comic strip narrations

Nassima FEZZA¹, Cyril AURAN²

¹ *Laboratoire Parole et Langage, CNRS, UMR 6057, Aix-en-Provence, France.*

² *Laboratoire Savoir, Textes, Langage, CNRS, UMR 8163, Villeneuve d'Ascq, France.*

nassima.fezza@lpl-aix.fr ; cyril.auran@univ-lille3.fr

Abstract

This paper fits into a tradition of prosodic analyses of discourse in French, but it proposes among the first multilayer multi-parametric descriptions. The study presented here is based on the analysis of a corpus of narrations from the comic strip *Titeuf* by six French natives. Discourse annotation was carried out, and pitch, intensity, duration and spectral parameters were automatically extracted from the recordings at different levels of the linguistic structure. Our results confirm and further illustrate previous studies on the prosodic marking of discourse structure in general, particularly concerning pitch reset at the beginning of new topics; but we also provide new elements regarding the discursive roles of intensity and speech rate, which somewhat challenge previous work in the field, and favour less strictly hierarchical models of discourse structure.

Index Terms: discourse prosody, discourse structure, French

1. Introduction

Our global project is at the interface between the acoustic, phonetic and phonological descriptive levels of prosody as defined in [1]. In a general bottom-up corpus-based approach, and based on semi-automatically extracted acoustic and phonetic measures, this project aims to describe, compare and, ideally, explain the diverse phenomena related to the prosodic marking of discourse structure in French, with the aim of potentially paving the way for prosodic phonology beyond traditional utterance-level structural units such as intonation phrases.

More specifically, this paper, which constitutes a preliminary part of this broader project, focuses on the acoustic analyses of narrations of comic strips by six native speakers of French. It is divided into three major parts: first, section 2 proposes a brief review of classical conceptions concerning the prosodic marking of discourse structure; section 3 describes the methodology (corpus and analyses) used in this study; finally, section 4 lists and discusses the main results of our multi-parametric and multi-dimensional analyses at different levels of prosodic structure, thereby providing a thorough descriptive model of the prosodic patterns more specifically associated with discourse discontinuity in French.

2. Prosody and discourse

The multi-parametric aspect of this study is motivated by a conception of prosody as a complex multidimensional macro-system with 4 domains: the tonal domain, which concerns short-term and long-term (e.g. level and range) pitch variations; the self-explaining intensity domain; the temporal domain, concerned with unit durations, speech rate and

acceleration; the spectral domain, related to voice quality. Our conception of prosody, identical to that suggested in [2] or [3], thus breaks away from some traditional descriptions, focused only on pitch and its short-term variations.

All these prosodic features have been known to play an important part in the coherent and cohesive structure of discourse, more specifically with relation to the marking of topic units ([2], [4]-[5]). Along these lines, this paper proposes to describe the prosodic configurations at the beginning and end of major topic units, configurations that have been described as frequently converging with the boundaries of paratones ([6], [7]-[8]). However, contrary to previous work that dealt with the prosodic marking of both discourse continuity and discontinuity ([4], or [9], [10], [11] for French) this study focuses on discourse discontinuity, with a classical distinction between initial discontinuity (topic beginnings) and final discontinuity (topic ends).

2.1 Initial discontinuity markers

Previous research dealing with prosodic cues of initial discontinuity tends to converge in describing higher pitch values ([6]; [12]; [13]; [14] and [15]). In English, the first stressed syllable of a paratone has been identified as being realised with a significantly increased pitch, a phenomenon known as *resetting* or *pitch reset* ([9]). In French, it is the first syllable that has been identified as being the locus of pitch resetting (see for instance [11]).

A similar trend has been observed with intensity, with a significant increase early in the first intonation unit in a paratone ([14]; [16]; [17] and [18]).

In the temporal domain, [5] or [19] mention longer syllables and slower speech rate at the beginning of paratones, a trend mainly explained by the introduction of new information.

2.2 Final discontinuity markers

Numerous studies associate the final boundary of discourse units with lower pitch values ([7]; [14]; [20]; [10] and [11] for French). Pitch is thus described as globally following a gradual downward trend from increased pitch values at the beginning of paratones (after pitch reset) down to low values in the speaker's register [21].

This frequency decrease is frequently accompanied by a decrease in intensity (for instance [14]; [16] and [18]) and, in the temporal field, as with initial discontinuity, decreased speech rate, and thus lengthened syllables [5].

3. Corpus and methodology

3.1 Corpus

For the purpose of this preliminary study, we recorded six native speakers of French. After reading three *Titeuf* comic strips, they were asked to retell the three stories in a row without previous notice or preparation. Our aim in this procedure was to allow the production of semantically and lexically controlled speech within a fixed type of interaction avoiding the artificial nature of read speech. Moreover, using a sequence of three short stories helped us to straightforwardly identify three separate discourse topics in each narration.

After this high-level discourse annotation, our data was transcribed orthographically and aligned manually at the level of words (1881 words in total) before undergoing semi-automatic multi-parametric prosodic annotation and statistical analysis. The prosodic part of this procedure is directly derived from that used in the Aix-MARSEC database ([22]).

3.2 Prosodic annotation

The initial stage of the prosodic annotation involved the manual segmentation and alignment of the approximately 9 minutes of speech signal at the level of phones (for a total of 4832 phones).

Higher-level annotations were then automatically derived from both this basic level and the lexical segmentation using Praat scripts [23]. This allowed us to generate aligned annotations at the level of sub-syllabic constituents (4561 SSC in total), syllables (2272 in total) and pause-defined units dubbed inter-silence segments (441 ISSs).

Acoustic-phonetic parameters were then automatically extracted at all these levels of granularity, leading to enriched TextGrids with 44 annotation tiers:

- Tonal domain: raw (Hz) and normalized (semitones around the speaker's mean) F0 values for slope, minimum and maximum, mean, median (a correlate of pitch level) and standard deviation (a correlate of pitch span).
- Intensity domain: raw (dB) values for minimum and maximum, mean, median and standard deviation.
- Temporal domain: raw and z-score normalized durations, speech rate and speech rate variation (acceleration/deceleration).
- Spectral domain: formant values and spectral slope.

3.3 Statistical analysis

The prosodic parameters described above were then automatically extracted by means of a Praat script in the form of a tab-delimited file used as input for the software and statistical environment R.

The measures were pooled for all subjects and all topics. We studied prosodic parameters in connection with the location (from the beginning and from the end) of all lower-ranked units within all their higher-ranked "containers" (from phone to topic). Such a procedure allows a fine-grained identification not only of the parameters involved in the prosodic marking of discourse structure, but also their precise location. To this end, positions were expressed in terms of both *left-position* (from the beginning of the "container") and *right-position* (from the end of the "container").

Our analyses mainly relied on classical statistical tests such as multi-factor analyses of variance, Student tests or

Kolmogorov-Smirnov tests when data distributions significantly deviated from a normal distribution.

4. Results

The statistical analyses were carried out bottom-up, starting from the lowest-ranked unit (the phone) to the highest-ranked one (ISS); we will therefore provide the results of these analyses in the same order. For obvious reasons of space, this paper will focus on results at the level of phones, syllables and ISSs, which are sufficient to provide both a fine-grained and a more global description.

Our aim primarily concerning the prosodic marking of discourse structure, we will mainly restrict this section to the effect of left-position and right-position within topics, taking into account, however, that right-position in the word and in the ISS, for instance, also play a major role in unit durations in French. These interactions were checked for in multi-factor Anovas, before more focused analyses based on post-hoc Student t-tests.

Given the inherent limited durations of phones and syllables, long-term parameters such as pitch level, intensity range or acceleration were not analysed at this level; results for the other (short-term) parameters are provided below.

4.1 Phones

4.1.1 Tonal domain

Concerning initial discontinuity, phone F0 is very significantly influenced by left-position in the topic ($F=129.6948$, $p<2.2e-16$). More specifically, we identified a significant F0 increase on the initial phone in a topic (2.4778889 st) in comparison with following phones (0.3700268 st; $t=3.7365$, $df=17,115$, $p=0.001626$).

With regard to final discontinuity, our results show a significant relationship between F0 and right-position in the topic ($F=57.6446$, $p=3.901e-14$). Repeated t-tests confirmed an F0 decrease on the last phone in a topic (-0.3045556 st; $t=-1.261$, $df=17,261$, $p=0.2241$) compared with all previous phones (0.39416 st). This trend extends beyond the last three phones, which suggests a higher-ranked phenomenon.

4.1.2 Intensity domain

Concerning initial discourse discontinuity, left-position in the topic seems to have only a marginal weakly significant effect on variations within the intensity domain ($F=4.8105$, $p=0.02834$).

The results are somewhat different when we consider final discontinuity. Indeed, right-position of the phone in the topic is highly significant ($F=73.7422$, $p<2.2e-16$), with a significantly lowered value for the last phone by comparison to previous ones ($t=-5.8105$, $df=21,389$, $p=8.47e-06$); t-tests show a large scale significant effect, for example with very high significance down to the fifth phone from the end of the topic ($t=-4.5074$, $df=123,462$, $p=1.503e-05$). This again suggests higher-level phenomena.

4.1.3 Temporal domain

Our analyses suggest that left-position in the topic influences phone durations ($F=13.2578$, $p=0.0002745$); however, post-hoc t-tests show that no significant effect could be located on the first phones. This suggests an indirect effect of other structural effects such as position in the word or in the ISS, or even an effect of right-position in the topic (which is itself obviously at least weakly correlated with left-position).

Right-position in the word ($F=44.0746$, $p=3.532e-11$) and ISS ($F=108.2490$, $p<2.2e-16$) were indeed shown to significantly influence phone durations, but no such clearly significant effect could be observed for topics ($F=4.9053$, 0.02682).

4.2 Syllables

4.2.1 Tonal domain

F0 variations very significantly depend on left-position in the topic ($F=18.3007$; $2.004e-05$); t-tests indicate significantly higher values for the first syllable (1.381352 st vs. 0.3746781 st; $t=2.9178$, $df=16.65$, $p=0.00975$) and for the following syllables (cumulatively integrated at least until the sixth). Again, the question of the span of the phenomenon arises, inviting us to drill up to the level of the ISS.

F0 variations also significantly depend on right-position in the topic ($F=18.3007$, $p=2.004e-05$); t-tests showed that F0 is gradually and linearly lowered over the last syllables, with a significant difference for example until the ninth syllable from the end of the topic (-0.07343972 st vs. 0.4327389 st; $t=-2.9182$, $df=166.53$, $p=0.004007$). This decrease is obviously located on higher-ranked units analysed below.

4.2.2 Intensity domain

Intensity values significantly depend on left-position in the topic ($F=19.7462$, $p=9.466e-06$); post-hoc KS tests show that this influence is not significant in the initial position ($D=0.1811$, $p=0.5368$). This paradoxical result may be due to the influence of word boundaries (with final primary accents in French), a hypothesis confirmed by tests carried out on the influence of right-position in the word ($F=8.8944$, $p=0.002905$).

Concerning final discontinuity, intensity values clearly are sensitive to the right-position of the syllable in the topic ($F=66.7276$, $p=6.348e-16$). KS tests indicate a significant reduction in mean and median intensity for up to the last six syllables pooled together (69.98 dB vs. 72.1 dB; $D=0.1738$, $p=0.006606$); this reduction is even more significant, moreover, for the last syllable compared with all previous syllables (65.40 vs. 72.04; $D=0.5309$, $p=8.81e-05$).

4.2.3 Temporal domain

Syllable durations seem to be weakly influenced by left-position in the topic ($F=5.0461$; $p=0.02482$); t tests show that the duration of the first syllable in a topic is not significantly different from all the other syllables pooled together ($t=-0.782$, $df=19,472$, $p=0.4436$). Here again, this non-significance may be linked to the influence of lexical stresses; in this perspective, post-hoc t tests show that

- the first syllable in a topic is significantly longer than other syllables when the loci of secondary and primary stresses are excluded (0.168 vs. -0.184, $t=3.6492$, $df=266,727$, $p=0.0003164$);
- primary stress loci are significantly longer than other syllables excluding secondary stress loci and topic-initial syllables pooled together (0.244 vs. -0.184, $t=4.32$, $df=273,554$, $p=2.185e-05$);
- secondary stress loci are significantly longer than other syllables with the exclusion of primary stress loci and topic-initial syllables pooled together (0.051 vs. -0.184, $t=2.5248$, $df=250,904$, $p=0.01219$);
- the durations of topic-initial syllables are not significantly different from those of secondary stress ($t=1324$,

$df=355,781$, $p=0.1864$) or primary stress loci ($t=0814$, $df=356,971$, $p=0.4162$).

Unsurprisingly, temporal variations are significantly sensitive to right-position in the word ($F=28.9888$, $p=8.387e-08$) and in the ISS ($F=39.7523$, $p=3.741e-10$). However, there does not seem to be any significant effect at topic level ($F=0.1299$, $p=0.7186$).

4.3 ISSs

4.3.1 Tonal domain

ISS median F0 is significant influenced by left-position in the topic ($F=9.7726$, $p=0.001995$). A series of t tests show significant higher values ($t=-3.8108$, $df=190,878$, $p=0.0001868$) for the first ISS (0.66 st) compared with all the others pooled together (0.119 st); however, there is a significant difference between the first and the second ISS (1.236 vs. 0.303 st; $t=2.6014$, $df=38,003$, $p=0.01316$), which is not the case for the second and the following ones. This indicates, superimposed onto a global declining trend, a localized effect on the first ISS.

In respect to final discontinuity, pitch level (median and mean) are significantly influenced by the right-position of the ISS in the topic ($F=9.7726$, $p=0.001995$); post-hoc t tests up to the fifth ISS from the end are significant (-0.125 st vs. 0.569 st; $t=-3.8108$, $df=190,878$, $p=0.0001868$), which suggests linear supra-declination.

We finally note that ISS pitch span (standard deviation) is significantly influenced neither by left- nor right-position in the topic. Topic type, however, which differentiates between narrations and inter-story comments, does play a significant role here, with an expanded pitch span for inter-story comments ($F=3.629$; $p=0.000351$).

4.3.2 Intensity domain

Left-position of the ISS in the topic influences neither intensity level (median and mean; $F=1.8549$, $p=0.1746$) nor intensity span ($F=1.7276$, $p=0.1900$).

However, right-position of the ISS in the topic is significantly influential ($F=16,957$, $p=5.341e-05$). This effect appears linear: it extends up to the fourth ISS from the end ($t=-3.2481$, $df=140,423$, $p=0.001453$) and no significant difference was found between the last consecutive ISSs (1st vs. 2nd ISS from the end: $t=0.1685$, $df=30,532$, $p=0.8673$; 2nd vs. 3rd: $t=-1.3344$, $df=33,486$, $p=0.1911$; ...). Here again, as within the tonal domain, this decrease seems to be gradual and is compatible with the concept of supra-declination.

4.3.3 Temporal domain

Our analyses show no significant effect of left-position on the durations of ISSs ($F=0.0632$, $p=0.8018$).

Similarly, no significant effect of right-position in the topic was found for ISSs durations ($F=0.9132$, $p=0.3403$).

5. Conclusion and perspectives

This preliminary study confirms previous results concerning the prosodic marking of discourse topic structure. For instance, it supports the conception of a progressive supra-declination ([9]) between successive units (ISSs in our case), in both the tonal and the intensity domains (contra [10]).

However, our results also somewhat conflict with previous descriptions on some points. Within the tonal domain, we notice that the first syllable in a topic does not seem to bear any distinctive quality with regard to the resetting

phenomenon (contra the position implicitly adopted for instance in [11]); rather, it is at the level of the ISS that distinctive features seem to appear, notably with the first ISS clearly realised with a significantly higher pitch level. Moreover, the tonal expansion at the beginning of new discourse segments described for instance in [17] and [18] is absent from our data.

Within the intensity domain, we did not observe any particular behaviour of topic-initial syllables or ISSs, which contradicts the traditional descriptions given for instance in [14], [17] and [18]. These results, however, need to be taken with caution, since our recording protocol did not involve the use of headset microphones which guarantee reliable intensity measures.

Within the temporal domain, finally, contra [5] and [19], we did not observe any lengthening trend at the level of topic-initial or topic-final ISSs and topic-final syllables (beyond the effects related to primary stress and final lengthening); topic-initial syllables, however, were shown to display an amount of lengthening analogous to that observed for stressed syllables.

These complex patterns argue for the complex conception of prosody as a macro-system mentioned above. The different domains do not always work in synchrony, but each domain may specialise in the marking of specific discourse functions, for instance in the case under scrutiny here, concerning initial and/or final discourse discontinuity. Our results seem to further argue against a strictly hierarchical structural conception of prosodic units beyond intonation phrases or similar utterance-related units (the *major paratone* in [7] or [14], the *major tone group* in [9], or, in French, the *paragraphe intonatif* in [12] and [24], the *paquet intonatif* in [20] or the *période intonative* in [4]). Our position seems to be in line with those defended in [25] for English or [10] and [11] for French, who show interesting cases of disjointed patterns, notably between resettings, supra-declination and final lowering at the level of diverse discourse units. However, our results are also compatible with the conception given in [6], in which intonation units may optionally be marked for discontinuity or, conversely, continuity without strict hierarchical constraints.

This preliminary study obviously is to be regarded as work in progress, and requires further exploration of our data together with an extension of our rather limited corpus and a slight modification of the recording procedure permitting more reliable intensity measures. However, we wish to insist on the interesting perspectives provided by this study concerning prosodic phonology beyond utterance-related units, notably with the open question of the phonological status of topic-initial syllables, that seem to share common temporal behaviour with primary and secondary stressed syllables in French.

6. References

- [1] D. Hirst *et al.*, "Levels of Representation and Levels of Analysis for the Description of Intonation Systems," In Horne, M. (ed.): *Prosody: Theory and Experiment. Text, Speech and Language Technology*, Kluwer Academic Publishers, pp. 51-87, 2000.
- [2] A. Di Cristo, "La problématique de la prosodie dans l'étude de la parole dite spontanée," *Revue Parole*, no 15-16, pp. 189-250, 2000.
- [3] A. Di Cristo *et al.*, "Outils Prosodiques et Analyse du Discours", *Cahiers de l'Institut de Linguistique de Louvain* 28, CILL 30, no 1-3, pp. 27-84. 2004.
- [4] C. Auran, "Prosodie et anaphore dans le discours en anglais et en français: cohésion et attribution référentielle," Ph.D. dissertation, Dept. Lettres et Sciences Humaines, Université de Provence, France, 2004.
- [5] C. Smith, "Topic transitions and durational prosody in reading aloud: production and modelling," in *Speech Communication*, no 42, pp. 247-270, 2004.
- [6] D. Hirst, "Intonation in British English," in D. Hirst & A. Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*, Cambridge, 1998, C.U.P., pp. 56-77.
- [7] G. Yule, "Speaker's topic and major paratones," *Lingua*, no 52, pp. 33-47, 1980.
- [8] A. Lacheret-Dujour & B. Victorri "La période intonative comme unité d'analyse du français parlé : modélisation prosodique et enjeux linguistiques," *Verbum*, vol XXIV, no1-2, pp. 55-72, 2002.
- [9] A. Wichmann, "Intonation in Text and Discourse: Beginnings, Middles and Ends," London, Long, 2000.
- [10] A-C. Simon, "La structuration prosodique du discours en français : une approche multidimensionnelle et expérimentelle," Berne, Peter lang, 2004.
- [11] A.C. Simon & A. Grobet. "Réinitialisations (resetting) et unités prosodiques maximales : une évidence ?," in V. Aubergé, A. Lacheret & H. Loevenbruck (éd.), *Journées prosodie 2001*, Presses Universitaires de Grenoble, pp. 79-84, 2003.
- [12] I. Lehiste, "The phonetic structure of paragraph," in A. Cohen & S.G. Nootboom (eds.), *Structure and Process in Speech Perception*, New York, Springer Verlag, pp. 195-206, 1975.
- [13] D. Brazil *et al.*, "Discourse intonation and language teaching," Harlow, Longman, 1980.
- [14] G. Brown & G. Yule, "Discourse Analysis," Cambridge, Cambridge University Press, 1983.
- [15] E. Couper-Kuhlen, "An Introduction to English Prosody," London, Arnold, 1986.
- [16] G. Brown *et al.*, "Questions of intonation," London, Croom Helm, 1980.
- [17] B. Grosz & J. Hirschberg, "Some Intonational Characteristics of Discourse Structure," in *Proceedings of the International Conference on Spoken Language Processing*, Vol. 1, pp. 429-432, 1992.
- [18] J. Hirschberg & C. Nakatani, "A prosodic analysis of discourse segments in direction-giving monologues," in *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, pp. 286-293, 1996.
- [19] F.J. Koopmans-Van Beinum & M.E. Donzel-Van "Relationship between discourse structure and dynamic speech rate," in *Proceedings ICSLP96*, 4th International Conference on Spoken Language Processing, Vol. 3, Philadelphia, pp. 1724-1727, 1996.
- [20] P. Mertens, "De la chaîne linéaire à la séquence de tons," *TAL* no 38, pp. 27-51, 1997.
- [21] M. Liberman & J. Pierrehumber "Intonational Invariance under Changes in Pitch Range and Length," in M. Aronoff and R. Oehrle, eds, *Language Sound Structure*, MIT Press, Cambridge MA. 157-233, 1984.
- [22] C. Auran, C. Bouzon & D. Hirst, "The Aix-MARSEC project: an evolutive database of spoken English," In Bel, B. & Marlien, I. (eds) *Proceedings of the Second International Conference on Speech Prosody*, Nara, Japan, pp. 561-564, 2004.
- [23] P. Boersma & D. Weenink, 2010, *Praat: doing phonetics by computer* [Computer program]. Version 5.3, retrieved from <http://www.praat.org/>
- [24] M.-A. Morel & L. Danon-Boileau, "Grammaire de l'intonation, l'exemple du français oral," Paris, Ophrys, 1998.
- [25] K. Hird, "The Relationship between Prosody and Breathing in Spontaneous Discourse," *Brain and Language*, no 80, pp. 536-555. 2002.