

The role of temporal variation in narrative organization

Nassima Fezza

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France nassima.fezza@lpl-aix.fr

Abstract

The aim of this study was to see if temporal variation can be considered a robust cue in the discourse structuring process. If so, at what level (s) of the discursive structure does it operate? In a bottom-up corpus-based approach, we analyze a 58minute corpus of 60 natural French speech narratives. First, the corpus was segmented at the phonemic, syllabic, lexical and inter-pausal unit levels. Second, a narrative segmentation was applied using the criteria of Labov's evaluative model, which is based on semantic and informational criteria. Duration data was then extracted automatically at each level of granularity. The mapping of discourse segmentation to acoustic-phonetic analyses was made on two structural levels: micro and macro. A positive effect of local temporal variation in discourse structuring was not found, however, the existence of a link between the narrative internal segmentation and speech rate variation was identified. This variation is longterm, progressive, and gradual which suggests a manipulation of this feature. In relation to the content, temporal values can be seen as contextual cues: relevant information is presented with a slower speech rate; while minor content is presented with a faster speech rate.

Index Terms: Speech rate, discourse structure, temporal variations

1. Introduction

Human speech rate is characterized by its variability; it varies not only from one speaker to another, but internally in the speech of the same speaker. This paper addresses the nature of variations in temporal domain. Do they occur in response to physiological constraints related to the regulation of respiration, or is it a grammaticalized illocutionary strategy related to discourse content? In other words, are these variations automatic or programmed? Numerous studies have addressed this question and have suggested that there is a direct or indirect link between discourse segmentation and duration variation. For instance, [1] and [2] showed an influence of speech rate and the length of pauses in the identification of parenthetical phrases and reported speech in English. However, their analyses could not define the temporal variations as a self-sufficient index of discourse segmentation

[3] observed a lengthening in average syllable duration at the left border of a topic opening and acceleration at the end of paragraphs (on the personal comments and additions) in Dutch. On the basis of this local use named as "premeditated", [4] and [5] speak of "manipulation" of speech rate. The speaker varies his flow deliberately to contextualize the relevance of information [6].

Until the end of the 90s, various studies examined temporal variables in narrative discourse based on textual analysis with

varying levels of success. Oliveira ([7], [8]) approaches the problem differently. Narrative discourse is a particular type of discourse and the analysis of its structure must be specific to it. In his study of Brazilian Portuguese, he tries to identify the constituents of narratives as presented in one of the most well-known models of narrative analysis, the Labov evaluative model ([9], [10]). His results do not demonstrate an effect of speech rate variation in the marking of discourse boundaries at the local level. However, he notes a correspondence at the global level between the Labovian narrative segmentation and variations of speech rate, noticing a "cyclical" way. No time value could be associated with these sections.

This study proposes to analyze rhythmic variation in relation to the segmentation process of narrative discourse in French with the aim of (i) clarifying the nature of this link, (ii) specifying its domain of action, and (iii) evaluating the value of these temporal variations in this segmentation process. In this study, we distinguish the local level, which includes phonemes, syllables and words as opposed to the global level, involving units above the word, therefore, the analyses will be carried out on these two levels. We assume that speakers control acceleration and deceleration in relation to what they have to say. Thus, on the cognitive level, they project information units to be filled, and vary the speech rate, the distribution and the length of pauses according to these blocks. Temporal variations would therefore be premeditated and would correspond to a structuring strategy.

2. Methodology

The corpus consists of the recordings of 20 native French speakers who were asked to narrate the contents of three comic strips. We thus obtained 60 semi-spontaneous narrations (a total of about 58 minutes of speech). The sound files were annotated and segmented, first of all, on different prosodic levels: phonemes, syllables, words and inter-pausal units. Secondly, we applied a discourse segmentation using the Labov evaluative model, which analyses 6 narrative parts, each part answering a question:

- Abstract: What is about?
- Orientation: Who, when, where?
- Complication: and then what happened?
- Evaluation: and so?

_

- Resolution: and finally?
- Coda: bring the auditor to the present.

This model was chosen due to its flexibility, which allows its application to natural non-prepared speech. Indeed, even if a linear temporal order is said to be preferred, this order is not rigid; the presence of some narrative parts is optional, and the model allows the appearance of a category more than once in each story. The following figure shows, using a finite-state diagram the number of transitions observed between each labovian segment and the next segment by including the transitions between stories for the same speaker:



Figure 1 Finite-state diagram representing the number of transitions between segments for the entire corpus.

3. Statistical analyses

3.1. Local analyses

In relation to the aims of this study, it was necessary to locate strong local variations of duration between two consecutive units, considering these rhythmic boundaries as a potential marking of a narrative boundary. The word was used as the basic, unit to be explored and not as more traditionally, the inter-pausal unit (IPU). The IPU is a unit determined solely by acoustic criterion constituted by the presence of a silent pause. It has been shown, however, [11], [12] for read speech in Korean, that the presence of the latter is not a necessary or sufficient criterion for the presence of a discourse boundary. Indeed, a discourse boundary can appear without the marking of a silence, inside an IPU for example, but it will necessarily appear between two consecutive words. For these reasons, the decision was made to exploit local variations between words.

The narrative boundaries were placed based on the evaluation model of Labov [9], [10]. In order to establish the rhythmic boundaries, an adaptation and enrichment of the methodology developed for the prediction of silent pauses in read speech in Korean ([11] and [12])was performed. This study noted the following potential boundaries:

- the normalized duration of the phoneme on the left and right of the potential boundary (z-score)
- the raw duration of the syllable on the left and right of the potential boundary
- the number of phonemes in these syllables
- the mean normalized duration of the phonemes in these syllables
- the raw duration of the word on the left and right of the potential boundary
- the number of phonemes in these words
- the mean normalized duration of the phonemes in these words
- the average normalized duration of the phonemes in a window of 250 ms to the right and to the left

Linear discriminant analysis was performed (LDA). The LDA is a predictive and also an explanatory classifier that separates two or more classes of objects using a dimensionality reduction before the classification. This method was used to 1) establish whether strong variations in duration induce a rhythmic boundary between two words; and 2) to check whether the locus of these rhythmic boundaries corresponds to the locus of discourse boundaries previously annotated according to the criteria of the Labov model.

The 16 measurements of duration were made for 9809 word boundaries. The algorithm predicted 107 rhythmic ruptures.

The following contingency table (Table 1) shows the result of an automatic classification and crossing algorithm:

Table 1: Contingency table of local rhythmic variations

Rhythmic boudaries					
		No	Yes		
Labovian boundaries	No	9379	92		
	Yes	287	15		

- 15 represent the true-positive: the number of Labovian boundaries annotated in the corpus and those predicted by the LDA as a rhythmic boundary.
- 9379 represent the true-negatives, the number of boundaries not annotated and not predicted by LDA.
- 92 represent the false-positives, the number of boundaries not annotated but predicted by the LDA.
- 287 represent the false-negatives, the number of boundaries annotated in the corpus, but not predicted by the LDA.

Calculation of the F measurement, which is a harmonic mean of the recall and precision, yielded a value of F=0.0030, a nonsignificant result. The boundaries predicted from the rhythmic variations do not seem to correspond to those that had been annotated using the Labov model. In that case, to what can they correspond? A manual check for the 92 rhythmic breaks was performed to determine if there was a correspondence with a discursive segmentation other than that proposed by Labov. The results obtained were classified as follows:

- 2% of these variations could not be categorized due to unclassifiable uses
- 18% coincide with the use of co-ordinating conjunctions in a stylistic use in action enumeration
- 23% of the boundaries appear with the use of a discourse particle.
- 57% of the predicted boundaries correspond in reality to repetitions following hesitations, abandonments and vocal lengthening.

The manual verification of these local variations made it possible to associate them to a large extent with reformulative resumptions and the use of discourse particles and to a lesser extent in the process of enumeration. This result therefore excludes any compatibility of local variations of duration with another discourse model.

3.2. Global analyses

On this discourse level, speech rate (SR) was used as a temporal measurement. Expressed in syllables per seconds, the SR allows us to account for language activity, treating pauses as a process included in the construction of the discourse. In this sense, the speaking time includes the pauses. The SR calculation formula was implemented as a Praat script and the measures were extracted automatically.

In order to neutralize intra-speaker variability and allow a comparison of the SRs, the mean- centric method of standardization was used. This method expresses the positive or negative deviation of a value from the average of the rank of the values of the same class. All these standardized SR measurements were grouped according to the Labovian category to which they belonged, all speakers and stories combined, we counted in total in our corpus 323 narrative segments: 55 Abstracts, 73 Orientations, 69 Complications, 75 Resolutions, 24 Codas, 27 Evaluations.

A one-factor analysis of variance was performed to see if the division into Labovian parts had an effect on the variation of

the overall SR. A link between the SR variation and the Labovian division was identified p<0.0001. Post-Hoc analyses using the Tukey method allowed us to perform a multi-comparison of the mean values of each Labovian part with the SR measurement, the table below shows the significant differences.

 Table 2: Multi-comparison of SR mean values using the Tukey

 method

	Estimate	Std. Error	t value	Pr (> t)
Coda - Abstract	1.8949	0.2596	7.300	<0.001
Complication – Abstract	0.8810	0.1918	4.594	<0.001
Evaluation - Abstract	1.4105	0.2493	5.657	<0.001
Orientation – Abstract	0.6863	0.1895	3.622	0.00434
Resolution - Abstract	1.2379	0.1884	6.572	<0.001
Complication – Coda	-1.0139	0.2514	-4.032	< 0.001
Orientation – Coda	-1.2087	0.2497	-4.841	<0.001
Orientation - Evaluation	-0.7242	0.2390	-3.030	0.02995
Resolution - Orientation	0.5517	0.1745	3.162	0.02012

For better readability, these averages are represented in figure 2



Figure 2:SR variation on the different Labovian parts

The first result that we can emphasize is that there is no significant difference in average SR between two successive units; but each narrative unit is realized with a distinguished SR from the rest of narrative units. Moreover, it should also be noted that the SR evolves along the narrative components in a very progressive way, accelerating between the opening unit and the closing units. Apart from this, the interaction between the Subject factor and the domain factor is significant for both (p=1.522E-13). We can then ask what are the domains of these interspeaker fluctuations? In order to answer this question, we carried out a descriptive statistical analysis of the speaker/domain interaction for the SR, classifying the centered mean values by subject and Labovian part. The results are presented in Figure 3.

Coda and Evaluation are the narrative domains that show the largest variations in inter-speaker values. No line can connect the values of the twenty speakers, and the intra-speaker variability is relatively large.

Abstract, Orientation, Complication and Resolution have a more stable SR, despite some points of variability; no significant deviation in this line is to be noted. The error bars on the averages are smaller, indicating relatively low intrinsic error margins.



Figure 3: SR variation per narrative part and speaker

3.3. Speech resetting

By extension to the temporal domain, we will speak of resetting in duration or speech reset following Oliveira's terminology [8]. This phenomenon consists of a rhythmic jump between two adjacent units. Following the definition of tonal resetting, the chosen unit of analysis was the syllable, even if this phenomenon has never been observed.

The analysis of temporal variability at the narrative boundaries started on the syllable level using the KS-test which demonstrates a significantly distinct temporal realization of the first accented syllable (narrative opening) in comparison with the rest of the syllables (p=2.251E-07), the first syllable being longer than the average. We extended these analyses to the first two syllables then to the three first syllables, and the result remains significant (respectively p=8.635E-11 and 7.624E-09). The initial syllable lengthening is therefore not restricted to the level of the syllable. The span of speech resetting is not the syllable, which encourages us to explore higher levels.

At the word level, a series of KS-tests were performed, comparing the mean value of the normalized duration of the first narrative opening word with that of the other words. The result shows that the initial narrative word is produced with a significant lengthening in comparison with the mean duration of the rest of the words (p=0.002). This result is found for the following initial words (cumulatively integrated at least until the third word (p=0.0004)) (see Figure 4). Again, these results suggest a resetting phenomenon that goes beyond the word boundaries, analysis on higher levels is therefore necessary.



Figure 4: KS-tests, comparing word mean value

At the IPU level, a comparison was made of the mean speech rate (centered corrected) of the first unit with the value of the rest of the narrative units by means of the KS-test. The results show that the opening IPU is significantly different from the other IPUs (p=5.985E-08). This test was extended to the first two IPUs, the result remains significant (p=4.322E-07), indicating a linear and progressive effect. However, the application of a t- test demonstrates a significantly different temporal realization between the first IPU and the second p=0.006, the second and the third p=0.007. The t-test is no longer significant between the third IPU and the fourth p=0.062.

The temporal reset span has not been specified, which is why temporal variations were analyzed on the Labovian segments. The narrative division is based on semantic criteria. The analysis of a structural phenomenon such as the resetting on semantic units can be argued in favor of its functional role. Thus, at each narrative change, we compared the mean values of SR (centered corrected) of Abstract with the SR of the preceding narrative transitions (Coda and Resolution):

H1 - H2: t-test indicate a significant difference between Abstract and Coda p=0.03, and between Abstract and Resolution (p=1.144E-07)

H2 - H3: t-test indicate a significant difference between Abstract and Coda (p=0.005), and between Abstract and Resolution (p=2.169E-07)

A significant difference in SR is therefore characteristic of narrative change. In other words, a marked rhythmic break is associated with a strong thematic break.

This observation is not the only one to be noted. The Abstract is produced with an SR significantly different from that of the Coda and the Resolution, the Abstract is also distinguished from the Orientation, as the results show Average comparison tests:

H2: The t-test indicates a significant difference between Abstract and Orientation (p=0.005).

H3: The t-test indicates a significant difference between Abstract and Orientation (p=1.928E-09).

The Abstract therefore differs also from the segment that follows it, only in H2 and H3 since the student's test does not indicate a significant difference between Abstract and Orientation (p=0.26) in the first story.

The Abstract thus has a significant place in the narrative transitions not only from the preceding part but also to that which follows it. This double emphasis is marked by both a categorical function and a demarcative function.

We classified the temporal data (centered duration) of each segment by history in Figure 5, in order to observe the rhythmic configurations at the narrative boundaries



Figure 5: SR variation on Labovian parts per story

Narrative change is accompanied by a slowing down of SR, regardless of the Labovian part that closes the narrative. The speech reset is not only performed by resetting with a return to

an average or "comfortable" SR. Indeed, this reset is marked by the passage of fast speech rate to slow rate and this between two Labovian segments separated by a silent pause. These results argue again in favor of a truly functional and directed reset

4. Discussion

Results of temporal short-term variations do not allow us to establish an association between local rhythmic breaks and the narrative division according to the Labovian criteria, nor after verification with another narrative model, in line with the results of [8]

Variations in our data appear in the majority of border uses of reformulation segments that intervene to correct and resume discontinuities or disfluencies, or coincide with the use of discourse particles. This rhythmic detachment is superimposed on the semantic and syntactic detachments that are otherwise attributed to them (see for instance [13])

Secondly, analyses carried out on long-term variations of duration show several results, taking the SR as a measure, we observed the existence of a tempo line throughout the narration, opening the narrative with a slow flow and closing it with a rapid flow, contrary to the results of [8], which rather describes cyclical variations in flow. The speakers adopt the same SR in marking the Labovian parts except for the Coda and the Evaluation (figure 3) which constitute the places of different strategies. This may correspond to their "personalized" contents according to the Labov model and which presents them as semantically imprecise. These findings support the informational contextualization function played by speech rates as described by [4] and [5]

Finally, we tried to characterize the SR values during transient passages from one narrative to another. Descriptive statistical analyses have enabled us to note that the narrative break is accompanied by a break in SR. The Abstract SR differs significantly from those of its immediate environment, not only from the preceding part, but also from that which follows. Abstract, according to Labov's criteria, provides information about the title or the overall content of the story. In our experiment, it plays a dual role: to indicate that the preceding story is finished and to announce the imminent beginning of a new one. This dual discursive marking is also found in the marking of SR value, which is slowed down in this narrative part.

5. Conclusions

The result of this study support the hypothesis that rhythmic variations play a role in the structuring process of narrative in French. Those variations are long-term, progressive and gradual which suggests a manipulation ([4] and [5]) of this prosodic parameter. Strong discourse ruptures are marked by strong prosodic variations expressed by a speech resetting. Furthermore, in relation to the content, rhythmic values can be seen as contextual[6] cues: relevant information will be presented with a slow speech rate in narrative opening; while minor content with a fast speech rate in narration closing

6. Aknowlogement

We thank Daniel Hirst who provided comments and expertise that greatly assisted this research,

7. References

- Hirschberg, J., & Grosz, B. (1992). Some intonational characteristics of discourse structure. In Proceedings of the International Conference on Speech and Language Processing, Banff, Canada.
- [2] Hirschberg, J. & Grosz, B. (1992). Intonation features of local and global discourse structure". In: Darpa Workshop On Spoken Language Systems, Proceedings Arden House, 1992. pp.441-446.
- [3] Koopmans-van Beinum, F. J., & van Donzel, M. E. (1996). Discourse structure and its influence on local speech rate. In Proceedings of the Institute of Phonetic Sciences (Vol. 20 : 1-11. Institute of Phonetic Sciences Amsterdam.
- [4] Fon, J. (1999). Speech rate as a reflection of variance and invariance in conceptual planning in storytelling. In Proceeding of the International Congress of Phonetic Sciences, San Francisco.
- [5] Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. Pragmatics, 6, 371-388.
- [6] Uhmann, S. (1992). Contextualizing relevance: On some forms and functions of speech rate changes in everyday conversation. The Contextualization of Language, 297- 336.
- [7] Oliveira Jr, M. (2000). Prosodic Features in Spontaneous Narratives (Doctoral dissertation, Simon Fraser University).
- [8] Oliveira Jr, M. (2012). A study on speech rate as a prosodic feature in spontaneous narrative. Alfa: Revista de Linguística (São José do Rio Preto), 56(2), 623-651.
- [9] Labov, W. (1972). Language in the Inner City: Studies in the Black English Vernacular (Vol. 3). University of Pennsylvania Press.
- [10] Labov, W., & Waletzky, J. (1967). Narrative analysis: oral versions of personal experience. Essays on the verbal and visual arts, Proceedings of the 1966 annual spring meeting of the American Ethnological Society, ed. by June Helm, 12-44.
- [11] Cho, H. S. (2009). Étude des propriétés acoustiques de la structure prosodique du coréen. (Doctoral dissertation, Aix Marseille 1).
- [12] Cho, H. S., & Hirst, D. (2006). The contribution of silent pauses to the perception of prosodic boundaries in Korean read speech. Proceedings of the Third International Conference on Speech Prosody 2006, Dresden.
- [13] Dostie, G., & Pusch, C. D. (2007). Présentation. Les marqueurs discursifs. Sens et variation. Langue française, (2), 3-