Disfluency and discursive markers: when prosody and syntax plan discourse

Julie Beliao & Anne Lacheret

MoDyCo, Department of Linguistics, Paris-Ouest University, France

Abstract

Hesitations, interruptions within phrases or within words are common in spontaneous speech. Those phenomena are widely known to be observable from a prosodic point of view through disfluencies. From a syntactic point of view, many studies already established that discursive markers such as *hm*, *oh*, *I mean*, etc. are representative of spontaneous speech. In this study, we demonstrate through a joint corpus-based analysis that these prosodical and syntactical features are correlated, without however being equivalent. More precisely, the lack of either disfluencies or discursive markers is consistently shown to be representative of a planned discourse.

Index Terms: disfluency, discursive marker, genres

1. Introduction

Corpora of spontaneous speech are characterized by a massive presence of disfluencies phenomena, whose function is yet to be better described (hesitation, self-repair, formulation, memory search, malaise, style effects, trademarks facilitating the right to speak, etc.). Despite the apparent irregularity of these phenomena ant their diversity, constants emerge through observation of corpora.

The disfluencies represent a very heterogeneous class. In this study, we distinguish between those that correspond to numerous acoustic markers (extensions, crushing registry, etc.), and those that are expressed only morphosyntactically (mainly rehearsals and unfinished segments without associated acoustic markers). The former are the prosodic *disfluencies* which we denote "hes" and the latter are denoted *discursive markers*, abbreviated "DM".

From a prosodic point of view, Corley shows in [1] that repetitions, also called disfluencies or hesitations, are not always accidental and are used by the speaker as a communicative tool. In fact, they are even advocated as a particular way to plan discourse. Apart from varying with the planning of discourse, disfluencies were also shown to depend on the content of speech [2]. For example, longer or complicated statements are much more likely to contain disfluencies [3,4]. Similarly, a speaker unfamiliar with the subject he is talking about will tend to produce disfluencies [5,6].

The importance of disfluencies for the syntactic analysis of spontaneous speech was highlighted by Dell [7], who proposed a syntactical model for natural language that explicitly accounts for disfluency as a fundamental phenomenon in spontaneous speech.

In this study, we build on those considerations and first focus on the identification of a possible correlation between prosodic disfluencies and syntactic discursive markers. To this purpose, we proceed to a large-scale statistical analysis of the Rhapsodie corpus of spontaneous speech, which is independently annotated both in syntax and prosody. Then, we show that the rate of disfluencies and DM is indeed representative of both the type (planned or spontaneous) and context (private or public) of discourse.

This paper is organized as follows. In section 2, we describe the context of this study and its material, namely the Rhapsodie Corpus. In section 3, we present the statistical analysis we performed along with its results. Finally, some conclusions are drawn in section 4.

2. Corpus

The aim of this section is to present the corpus and the various transcription studies that have been conducted during 3 years within the Rhapsodie project [8]. This project aims at providing and testing on large-scale of constructions a new prosodical and syntactical transcription and annotation system. A total of 57 samples were gathered – such as existing samples from corpora [9,10,11] but also new ones – with a wide topological coverage.

The speech database covers various discourse genres and speaking styles and comprises about 3 hours of continuous speech, monologues and dialogues, private vs. public, face-to-face vs. broadcasting, more or less interactive, descriptive vs. argumentative vs. procedural samples.

In this study we present an analysis of disfluencies and discursive markers which are often considered as similar phenomena. Each of those two phenomena has been annotated independently from the other using either only formal syntactical criteria or only acoustical/prosodical criteria.

2.1. Syntactic analysis

2.1.1. Syntactic annotation

Combining the syntactic model proposed by the Aix School [12] and the pragmatic model developed within the Lablita experience [13], two levels of syntactic cohesion have been annotated within Rhapsodie: microsyntax (i.e., syntactic cohesion guaranteed by government) and macrosyntax (i.e., syntactic cohesion guaranteed by illocutionary dependency). Microsyntax describes the kind of syntactic relations which are usually encoded through dependency trees or phrase structure trees.

Macrosyntax, which is of interest to us here, can be understood as an intermediate level between syntax and discourse. This level describes the whole set of relations holding between all the sequences that make up one and only one illocutionary act. The annotation of macrosyntax is essential to account for a number of cohesion phenomena typical of spoken discourse and in particular of French spoken discourse, because of the high frequency of paratactic phenomena that characterizes this language.

A complete annotation and a functional tagging of pile structures are also available [12,13,14]. More generally, a complete categorical and functional tagging for every word was achieved in the Rhapsodie corpus, including discourse markers, which are integrated into the syntactic representation at the macrosyntactic level.

2.1.2. Discursive markers annotation protocol

Discursive markers (DM), also called "associated illocutionary units" [14], are considered as macro-syntactic units. They often come as series of impaired verbal constructions, such as *huh*, *well*, *uh*, *so*, *hem*, etc. These units, which we denote with quotation marks "" are equipped with an illocutionary operator but they do not convey information content that is added to the content of knowledge shared by the interlocutors. They are a special case of illucutionary units described below.

An Illocutionary Unit (IU) is any portion of discourse encoding a unique illocutionary act: assertions, questions, and commands [15,16]. An IU expresses a speech act that can be made explicit by introducing an implicit performative act such as "I say", "I ask", "I order". A test for detecting the Illocutionary Units that make up a discourse consists of the introduction of such performative segments (see below). A segmentation in IUs is particularly important for the study of the connection of prosody and syntax, which is the goal of Rhapsodie, because these units are prosodically marked. For example, consider the following statements:

(1) c'est fils de la Sarce "je crois"
it's son of Sarce "I think"
[Rhap-M011, Corpus Avanzi [17]]

(2) ils sont deux Argentins "hein"
they are two Argentinians "eh"
[Rhap-D2003, Broadcast Corpus]

(3) je lui ai dit "ben" "tu vois" je vends des livres
I told him "well" "you know" I sale books
[Rhap-D2001, Corpus Mertens [18]]

Segments "je crois" (*"I think"*), "tu vois" (*"you know"*), "hein" (*"eh"*) and "ben" (*"well"*) are equipped with an illocutionary operator that permits to recognize them as assertions (*I think, you know*) or exclamations (e*h*). They share internal characteristics with the nuclei, such as being segments organized around a finite verb or reduced to an interjection. However, these segments do not convey information content that is added to the content of knowledge shared by the interlocutors: they can be deleted, for example, without any state of knowledge being changed. They do not have a descriptive function, but rather a function of modal change (as in the first example) or interactional regulation (as in the following two examples). From this point of view, we can say that they lack illocutionary strength in the true sense of the term. Indeed, they are not proper illocutionary acts addressed at an interlocutor, who can not deny or question the content of these segments.

2.2. Prosodic analysis

2.2.1. Prosodic annotation

For prosody, Rhapsodie annotators built on the theoretical hypothesis formulated by the Dutch-IPO school [19] stating that, out of the total information characterizing the acoustic domain, only some perceptual cues selected by the listener are relevant for linguistic communication [20,21]. On this basis, they decided to manually annotate only three perceptual phenomena characterizing real productions: prominences, the cornerstone of the sentence-prosodic segmentation [22,23], pauses and disfluencies [24].

Starting from this annotation, a prosodic structure was automatically generated, organized around rhythmical and melodic components. For each constituent of the structure, prototypical-stylized melodic contours were computed. First, perceptual syllabic salience in speech contexts was annotated using a gradual labeling distinguishing between strong, weak or zero prominences. Second, all prosodic segments were annotated, including disfluencies and different kinds of pauses (silent pauses vs "uh" or syllabic hesitation in the proximity of pauses). Third, prototypical-stylized melodic contours were generated for units of different sizes and domains. The availability in the Rhapsodie Treebank of the contour of a large number of prosodic and syntactic units allow the user to build various lexicons of intonation shape in an extremely flexible way according to his/her research goals [25].

In more general terms, it should be highlighted that these annotation choices have allowed us to identify the primitives of prosodic structure independently from any reference to syntax or pragmatics, and to provide all the elements needed for a complete prosodic analysis of linguistic units.

2.2.2. Disfluencies protocol annotation

An element that breaks the flow phrase in the speech chain, like stumbling voice, is called a disfluency or hesitation (hes) in this study. It can take different types and often exhibits an excessive syllabic elongation. Moreover, it can often consist in a repetition of morphemes or in interruptions of words or sentences. Every syllable perceived as disfluent is marked with a specific tag, annotated H. The annotation of disfluencies is carried out manually under PRAAT [26]. Each encoding step performed sequentially on batches of data. The number of listening for annotating one batch is limited to 3.

The most classic disfluencies are:

- Interruptions: *it's not far* | ***you*** | ***I'm going***
- Repeated segments: it is not far **you you go**
- Hesitations: "Uh"
- Excessive syllabic lengthening (not corresponding to a boundary structure): ***we:::::::ll***

These phenomena are not mutually exclusive but can be combined. If disfluence is simple (one disfluent syllable), the disfluent syllable is annotated "H" in the dedicated tier. If it is combined, i.e. concerns several successive intervals, then all the corresponding syllables are tagged as disfluent.

3. Methodology and statistical analysis

In this study, we focus on the correlation of the number of DM with the number of disfluencies over all samples from the Rhapsodie corpus. To this purpose, we propose a statistical analysis focusing on two separate aspects.

A first analysis focuses on the average number of disfluencies and DM per minute (hes/min and DM/min). Studying these scatter-plot showing one vs the other across all samples, we performed a correlation study, to be developed in section 3.1.

Then, disfluencies and DM may simply be synchronized. If that would be the case, it would mean that the corresponding annotations were strongly influenced by each other. In section 3.2, we show through a synchronization analysis that it was actually not the case.

Finally, the average number of disfluencies and DM of the samples are used to identify those corresponding to planned or spontaneous speech.

3.1. Correlation of disfluencies and DM

A first remarkable fact is that there are approximately half DM compared to disfluencies in the whole Rhapsodie corpus, as can be seen in Table 1.

Table 1: *Basic numbers from the corpus*

Units	count
Syllables	45192
Words	33 182
Disfluencies	3460
DM	1818

However, having half as many DM than disfluencies does not mean they are uncorrelated. Indeed, when plotting hes/min vs DM/min for all samples clearly show a strong correlation between them.

To confirm this fact, we performed a linear regression whose results are displayed in figure 1. With an infinitesimal *p*-value of $p=1.8\times10^{-15}$, the null-hypothesis corresponding to decorrelation can safely be rejected. A correlation of 0.8 was observed between hes and DM.

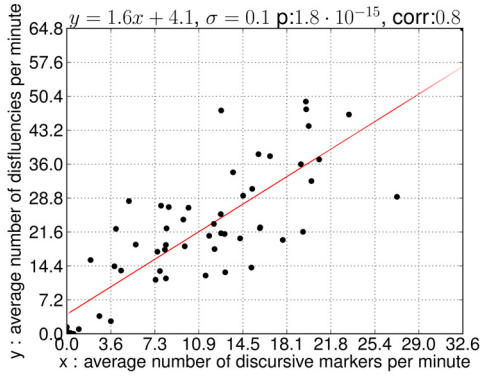

Figure 1: Average number *of disfluencies and discursive markers per minute for each sample (black dots)*

3.2. Independenceof DM and disfluency annotations

This strong correlation between DM/min and hes/min means that these phenomenon are related. Still, this may actually be due to a deterministic interaction such as a bias in the annotation.

In order to check for this hypothesis, we performed a synchronization analysis and display in Figure 2 the ratio of both hes and DM that have the same temporal support.

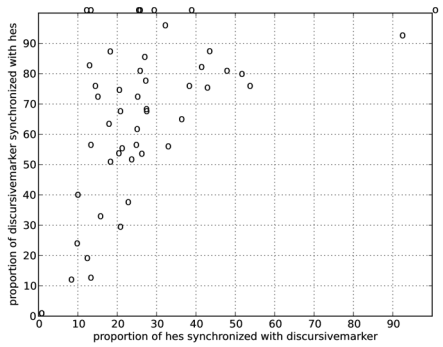

Figure 2 : *Display of synchronization between disfluencies and discursive markers.*

The low proportion (between 10% and 50%) of disfluencies which are synchronized with DM confirms that these units have been annotated independently from syntactic considerations, unlike DM.

On the contrary, there are a bit more (between 40% and 80%) of DM which are synchronized with disfluencies, proving that annotation of DM probably involved disfluencies in some way. However, this tendency looks rather weak considering the scatter-plot displayed in Figure 2, and may also simply be due to the higher number of disfluencies than DM, causing an increase in the probability of their synchronization.

3.3. When average number variation means planning

In this section we show how the joint presence of DM and disfluencies within a sample of the corpus is related to the corresponding type of discourse. To this purpose, we display for each sample its position along the regression line found in Figure 1. That way, a small value indicates a lack of both disfluencies and DM, whereas a high value indicates their joint important within the sample. The resulting barplot can be found in Figure 3.

Considering this figure, we can see that samples corresponding to planned speech contain very little disfluencies and DMs whereas the repartition of semi-spontaneous and spontaneous samples is more random. The same results pertain to public or private speech, the former including less DM and disfluencies than the latter. Hence, public speech may appear more planned, which seems natural. It should be emphasized that the two first samples are the corpus' shortest (less than 60 words) and don't contain any DM and disfluencies.

4. Conclusions

It is widely acknowledged in prosodic studies that there is a strong relation between disfluencies and speech planning. Similarly, many syntactical studies have established that discursive markers are typical of spontaneous speech.

However, there was no study we were aware of that performed a joint intonosyntactical analysis on a large-scale corpus to study how prosody and syntax agree on the same data. In this paper, we demonstrated that the density of disfluencies in a sample is indeed strongly correlated to the density of discursive markers, even if the two notions are showed not to be equivalent. It is hence our belief that a joint analysis of prosody and syntax may lead to a better understanding of spontaneous speech.

Since hesitations are the most frequent type of speech disfluence in many languages, it is possible that the majority of the synchronized cases falls within the class of hesitations. It would be and interesting perspective to examine more finely the degree of synchronization for sub-classes of disfluencies and discursive markers.

5. Acknowledgements

We warmly thank Antoine Liutkus for his English review and for his help in figures display as well as the Rhapsodie Consortium.

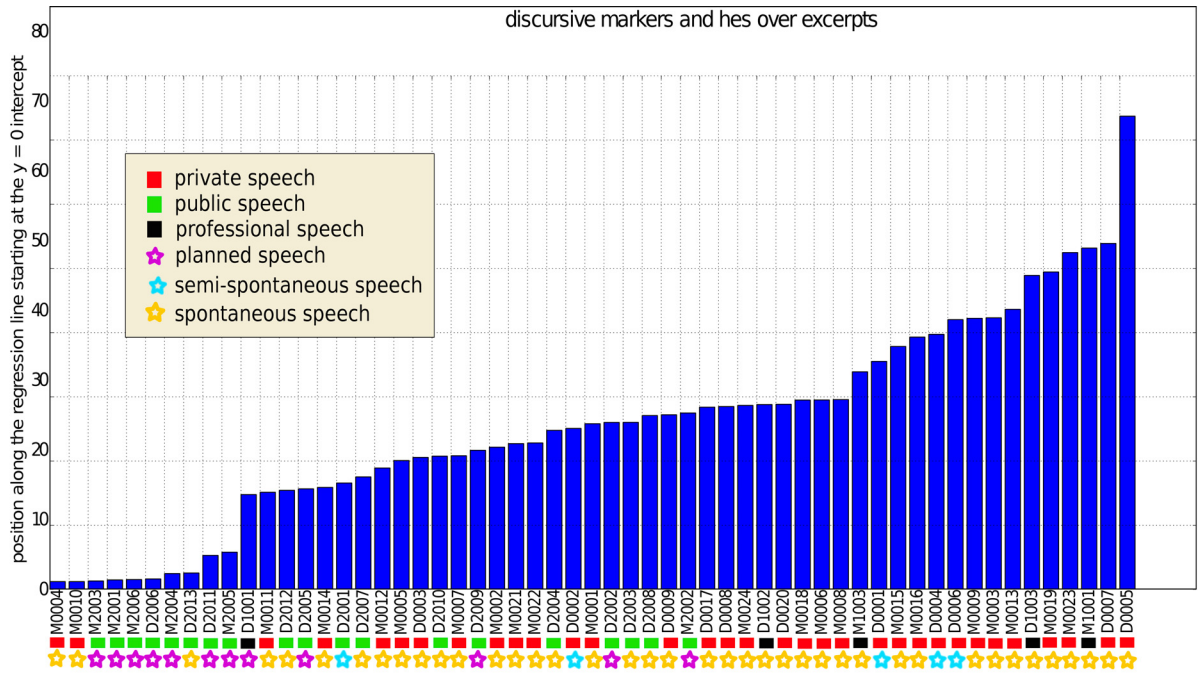

Figure 3: *Rate of disfluencies and discursive markers (Y-axis) for the 57 samples of the Rhapsodie corpus (X-axis), along with a description of the content of each sample in terms of public/private/professional/planned/spontaneous speech.*

6. References

[1] M. Corley, O.W. Stewart, "Hesitation disfluencies in spontaneous speech: The meaning of um", *Language and Linguistics Compass* 4, pp. 589–602, 2008.

[2] S. Schachter, N. Christenfeld, B. Ravina, F. Bilous, "Speech disfluency and the structure of knowledge", *Journal of Personality and Social Psychology* 60, pp. 362–267, 1991.

[3] S. Oviatt, "Predicting spoken disfluencies during human–computer interaction", *Computer Speech and Language* 9, pp. 19–35, 1995

[4] E. Shriberg, "Disfluencies in Switchboard", *Proceedings International Conference on Spoken Language Processing*, Addendum, pp.11–14, Philadelphia, 1996.

[5] H. Bortfeld, S.D. Leon, J.E. Bloom, M.F. Schober, S.E. Brennan, "Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender Language and Speech", *Language and Speech* 43(3), pp. 229–147, 2000.

[6] S. Merlo, L. Mansur, "Descriptive Discourse: Topic Familiarity and Disfluencies". *Journal of Communication Disorders* 37, pp. 489–503, 2004.

[7] G.S. Dell, "A spreading activation theory of retrieval in sentence production", *Psychological Review* 93, pp. 283–321, 1986.

[8] A. Lacheret, S. Kahane, P. Pietrandrea, "Rhapsodie: a Prosodic and Syntactic Treebank for Spoken French", *Studies in Corpus Linguistics*, Amsterdam, Benjamins, 2013.

[9] S. Branca-Rosoff, S. Fleury, F. Lefeuvre, M. Pires. *Discours sur la ville. Corpus de Français Parlé Parisien des années 2000 (CFPP2000)*, 2009.

[10] B. Laks, J. Durand, C. Lyche, Le projet PFC (Phonologie du Français Contemporain) : une source de données primaires structurées. In *Phonologie, variation et accents du français Hermès*, pp. 19–6, 2009.

[11] M. Avanzi, A.-C. Simon, J.-P. Goldman, A. Auchlin, "Un corpus de français parlé annoté pour l'étude des proéminences, c-prom. Actes des 23èmes journées d'étude sur la parole", Mons, Belgique, 2010.

[12] C. Blanche-Benveniste, "Un modèle d'analyse syntaxique 'en grilles' pour les productions orales", *Anuario de Psicologia* Liliane Tolchinsky (coord.), vol. 47, Barcelona, pp. 11–28, 1990.

[13] E. Cresti, *Corpus di italiano parlato*, Florence: Accademia della Crusca, 2000.

[14] E. Bonvino, F. Masini, P. Pietrandrea, "List Constructions: a semantic network", *Troisième Conférence Internationale de l'AFLiCo*, Nanterre, 2009.

[15] S. Kahane, P. Pietrandrea, "Les parenthétiques comme Unités illocutoires associées. Une perspective macrosyntaxique", *Linx* 61, pp. 49–70, 2012.

[16] C. Benveniste, "Problèmes de linguistique générale", volume 1 of Coll. TEL, Gallimard, 1966.

[17] J.R. Searle, "The classification of illocutionary acts", *Language in Society* 5, pp. 1–24, 1996.

[18] M. Avanzi, *L'interface prosodie/syntaxe en français. Dislocations, incises et asyndètes*, Bruxelles, Peter Lang, 2012.

[19] P. Mertens, *L'intonation du français : de la description linguistique à la reconnaissance automatique*, Thèse de Doctorat, Université de Louvain, 1987.

[20] J. Hart, R. Collier, A. Cohen, *A perceptual study of intonation, an experimental phonetic approach to speech melody*, Cambridge University Press, 2006.

[21] A. Lacheret, F. Beaugendre. *La prosodie du français*, Paris, CNRS, pp. 62, 1999.

[22] C.W. Wightman, "ToBI or not ToBI?", *Proceedings of Speech Prosody*, Aix-en-Provence, France, pp. 25–29, 2002.

[23] J. Buhmann, J. Caspers, V.J. van Heuven, H. Hoekstra, J.-P. Mertens, M. Swerts, "Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus", *Proceedings of LREC 2002*, Las Palmas, pp. 779–785, 2002.

[24] F. Tamburini, C. Caini, "An automatic System for Detecting Prosodic Prominence in American English Continuous Speech", *International Journal of Speech Technology* 8, pp. 33–44, 2005.

[25] A. Lacheret, N. Obin, M. Avanzi, "Design and evaluation of shared prosodic annotation for spontanesous French speech: from expert knowledge to non-expert annotation", *48th Annual Meeting of the Association for Computational Linguistics*, 4th Linguistic Annotation Workshop, Uppsala, Sweden, 2010.

[26] V. Aubergé, G. Bailly, "Generation of intonation: a global approach", In *Proceedings of the European Conference on Speech Communication and Technology*, Madrid, pp. 2065–2068, 1995.