Phonetic and Prosodic Characteristics of Disfluencies in French Spontaneous Speech

George Christodoulides [1] *Mathieu Avanzi* [1,2]

[1] Centre Valibel, Institute for Language & Communication, Université de Louvain, Belgium
[2] Swiss National Science Foundation & Université de Neuchâtel, Switzerland
george@mycontent.gr **mathieu.avanzi@gmail.com**

Ehm… Introduction

A key difference between **spontaneous speech** and controlled laboratory speech is the prevalence of **disfluencies** in the former (e.g. Shriberg 1994).

- Disfluencies typically signal "problems", such as interruptions in the production, **as the speaker incrementally constructs his message** (Levelt 1989).
- However, in specific contexts, these events may be used as **communicative devices**, e.g. in order to **manage dialogue interaction** (Moniz et al. 2009), **indicate information status** (Arnold et al. 2003), and may **coincide** with other prosodic **events** (e.g. major boundaries).
- Disfluencies have recently attracted the interest of the phonetic sciences and of computational linguists working on speech (e.g. Adda-Decker et al. 2004, Arbisi-Kelm 2010, Germesin et al. 2008, Stolcke & Shriberg 1996). The **automatic detection** of disfluencies in large spoken corpora is an important intermediate step that facilitates further analyses such as dependency parsing.

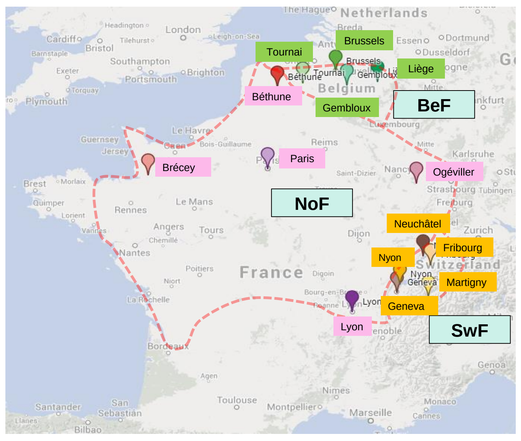
In this study, we focus on the **prosodic and phonetic characteristics** of disfluencies in spontaneous speech, in an effort to **detect regularities** that may improve automatic detection algorithms.

Corpus Data

Corpus CPROM-PFC (Avanzi 2014) created from the PFC corpus (Durand et al 2002, 2009)

- **14 regional varieties** of French recorded in France, Belgium and Switzerland (4 different cities per country)
- **112 speakers** (4 M, 4 F per city) aged 20-80
- Semi-directed sociolinguistic **interviews** (3 min per speaker) + **reading** of a text (398 words).
- In total, approximately 11 hours.

In this study **we focus only on the Spontaneous Speech** sub-corpus (approx. **7 h**).



Disfluency Annotation Scheme

The corpus has been orthographically transcribed, aligned to the phone and syllable level, and POS tagged. The corpus was annotated under *Praat* (Boersma & Weenink 2014). The annotations include the **3 levels of annotation proposed** by *DisMo* (Christodoulides et al. 2014). The part-of-speech annotations were manually corrected.

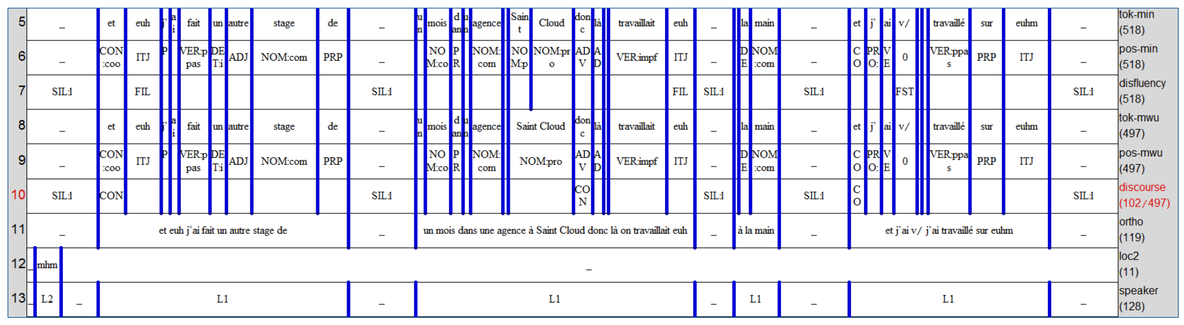


Figure 1: The three-level system of annotation as represented on a *Praat* text-grid. The level of minimal tokens includes tiers `tok-min`, `pos-min` and `disfluency`. Multi-word units may span several minimal tokens (tiers `tok-mwu` and `pos-mwu`); this allows to annotate MWEs having a different POS function than their constituent parts. Discourse markers (DMs) and related phenomena (tier `discourse`) may span multiple tokens; thus the POS tags of any token or expression functioning as a DM are preserved.

We propose a detailed **multi-level annotation scheme for disfluencies**, by combining aspects of the systems proposed in Shriberg (2001), Brugos et al. (2012) and Heeman et al. (2006).

Level 1: Simple disfluencies are those affecting only one token		
FIL	Filled pauses	`c' est pour ça que j' hésite euh un peu en parler` `FIL`
LEN	Hesitation-related lengthening	`au cercle d'oenologie de= Bruxelles` `LEN`
FST	Lexical false start	`comme infirmière so/ sociale` `FST`
WDP	Pause within word	`il m' a dit ça su+ _ +ffit` `WDP`
Level 2: Repetitions where a token or a series of tokens are repeated (exactly the same form)		

REP	Repetition	`les disques et et lancer les jingles`
		`REP* REP_`
		`il a il a il a dit que`
		`REP:1 REP:2 REP:1 REP*:2 REP_ REP_`
		`c' est pas c' est pas un système génial`
		`REP:1 REP:2 REP*:3 REP_ REP_ REP_`

Level 3: Structured editing disfluencies		
DEL	Deletion	`c' est vraiment un en tout cas la parole` `DEL DEL DEL DEL*`
SUB	Substitution	`cette personne était enfin c' est un ami de` `SUB* SUB:edt SUB_ SUB_`
INS	Insertion	`c' est vrai que Béthune euh vivre à Béthune ça aurait` `INS* INS+FIL INS_ INS_ INS_`
Level 4: Complex disfluencies are a combination of several structured ones: annotated separately using a backtracking table		
COM	Complex	`les ac/ les actions enfin les activités enfin professionnelles` `COM COM COM COM COM COM COM COM COM`

Simple (L1) disfluencies affect only one token. L1 annotation codes may combine with L2 and L3 codes. **Repetitions** are defined strictly (exact same form) and may combine with L3 codes; when several tokens are repeated, numbering describes the pattern. **Structured** disfluencies (L3) are those that follow the pattern:

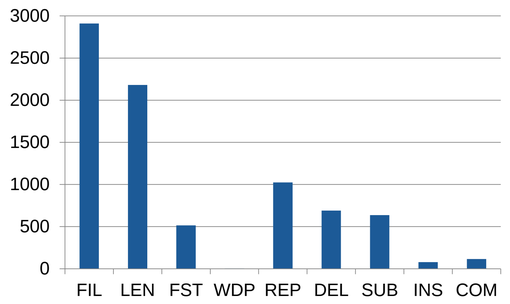
reparandum – interruption point – interregnum including optional editing terms – repair

The IP is indicated by appending an asterisk (*) to the annotation code; editing terms by appending ":edt" – these may be discourse markers or any other phrase used by the speaker to indicate his intention to change the utterance. The repair region is marked by appending an underscore (_) to the code. **Complex** disfluencies are annotated in detail using a **backtracking table**: codes such as COM(x,y) indicate the position of each token in this table allowing reconstruction of the analysis (see Heeman et al. 2006 for details). We have developed **software to facilitate human annotators**: in a *Disfluency Editor*, the user selects a region of tokens and the disfluency type. The system automatically infers the structure of the disfluency and adds appropriate annotation codes to each token.

Pattern Analysis

Regarding **disfluency types** observed in the corpus, we confirm previously identified tendencies:

- **Filled pauses** (both autonomous fillers and epenthetic vowels) are the most common single-token disfluencies, followed by **lengthening** (drawls)
- 82% of lexical **false starts** co-occur at the interruption point of structured disfluencies
- Among the structured disfluencies, **repetitions** (especially 1- and 2- token) are the most prevalent, followed by **deletions** and **substitutions**



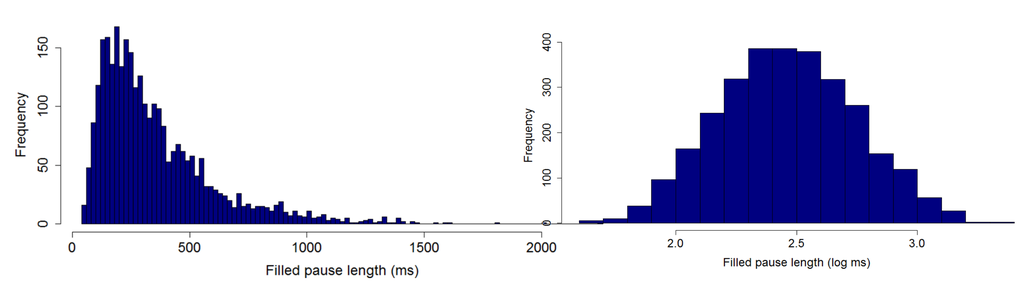
The most common **patterns** observed were the following:

Pattern	Description	Occurrences
FIL SIL:l	Filled pause followed by a long silent pause	446
REP* REP_	Repetition of one token	186
LEN SIL:l	Lengthening followed by a long silent pause	121
FIL SIL:b	Filled pause followed by a short silent pause	100
REP* REP+SIL REP_	Repetition of one token, with a silent pause in-between	90
LEN LEN	Consecutive lengthening of two tokens	74
LEN FIL	Lengthening followed by a filled pause	68
REP:1 REP*:2 REP_ REP_	Repetition of two tokens (A B A B)	58
REP:1 REP*:2 REP+SIL REP_ REP_	Repetition of two tokens (A B A B), with a silent pause in-between	39
DEL DEL*	Deletion of two tokens	38
REP* REP+FIL REP+SIL REP_	Repetition of one token, with a one filled and one silent pause in-between	34
SUB* SUB_	Substitution of one token with another one	33
REP:1 REP*:1 REP_	Double repetition of the same token (A A A)	33
DEL*	Deletion of one token	31
LEN SIL:b	Lengthening followed by a short silent pause	30
LEN FIL SIL:l	Lengthening followed by a filled pause and then a long silent pause	30
DEL DEL* SIL:l	Deletion of two tokens followed by a long silent pause	26

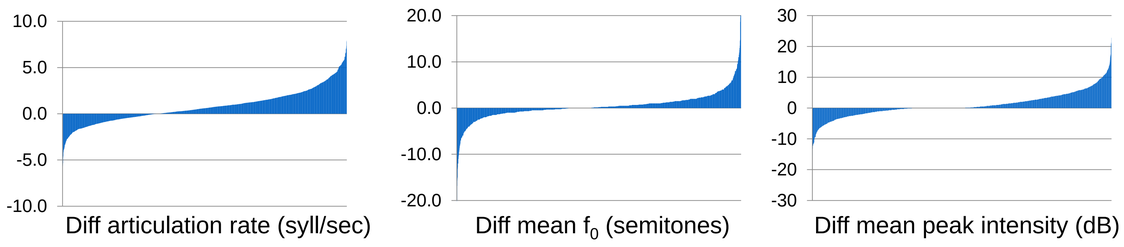
Results and Perspectives

The main findings of the phonetic/prosodic analysis of disfluencies in the corpus are the following:

1. In the **reparandum** region of structured disfluencies, the syllables affected are **those immediately preceding** the interruption point (**lengthening** or **trailing fillers** or **short silent pauses**).
2. The **duration of filled pauses** follows a unimodal log-normal distribution:



3. The vowel length distribution of fillers is significantly different from the distribution of vowel length in fluent contexts, distinguishing **disfluency-related lengthening** from **prosodic prominence**. Filled pauses exhibit **falling intonation contours**, with a mean pitch close the speaker's mean pitch.
4. Prosodic **cues** may signal the **interruption point** of a disfluency: **articulation rate, pitch and energy increase** in the **repair** region (compared to the reparandum) possibly signalling the boundaries of each region.



Three measurements (articulation rate, mean pitch and mean peak intensity) were taken 500 ms before the interruption point and 500 ms after the start of the repair region. The means of the differences (repair – reparandum) were **0.77 syll/s**, **0.38 ST** and **0.96 dB** respectively (in all cases: p<0.001; pairwise t-test). These cues are not always present, but may aid listeners, as well as automatic detection systems.

Perspectives

- Extend the study to other prosodic measures (e.g. voice quality, creakiness, slope)
- Based on the 7-hour annotated corpus, improve the automatic disfluency detection and annotation models used in *DisMo*.
- Apply this disfluency annotation scheme to major French spoken corpora, including different speaking styles, in order to study the effects of variation on the production of disfluencies.
- Disfluency-aware dependency parsing of French spoken corpora.

References

Adda-Decker, M., Habert, B., Barras, C., Boula de Mareüil, Ph., Paroubek, P. 2004. Une étude des disfluences pour la transcription automatique de la parole et l'amélioration des modèles de langage. *Actes des 25èmes JEP*.
Arbisi-Kelm, T. 2010. Intonation structure and disfluency detection in stuttering, In: Fougeron, C., Kühnert, B., D'Imperio, M., Vallé, N. (eds.) *Laboratory Phonology* 10. Berlin: De Gruyter Mouton, 405-432.
Arnold J. E., Fagnano M., Tanenhaus M. K. 2003. Disfluencies signal theee, um, new information, J. Psycholinguist Res, 32(1), 25-36
Avanzi, M. 2014. A Corpus-Based Approach to French Regional Prosodic Variation. *Nouveaux cahiers de linguistique française*, 31.
Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer, ver. 5.3, www.praat.org
Brugos, A., Shattuck-Hufnagel, S. 2012. A proposal for labelling prosodic disfluencies in ToBI. *Poster, Advancing Prosodic Transcription for Spoken Language Science and Technology*, July 31, 2012, Stuttgart, Germany.
Christodoules, G., Avanzi, M., Goldman, J.-Ph. 2014. DisMo: A Morphosyntactic, Disfluency and Multi-Word Unit Annotator, *Proc. of LREC*, 3902-3907.
Germesin, S., Becker, T., Poller P. 2008. Hybrid Multi-step Disfluency Detection, *Proc. 5th Intl Workshop on Machine Learning for Multimodal Interaction*.
Heeman, P., Mcmillin, A., Scott Yaruss, J. 2006. An annotation scheme for complex disfluencies. *Proc. 9th ICSLP*, 1081-1084.
Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press
Moniz, H., Batista, F., Trancoso, I., Mata da Silva, A.I. 2012. Prosodic context-based analysis of disfluencies. *Proc. Interspeech Portland, Oregon*.
Moniz, H., Trancoso, I., Mata da Silva, A.I., 2009. Classification of disfluent phenomena as fluent communicative devices in specific prosodic contexts, *Proc. Interspeech 2009, Brighton, UK*, 1719-1722.
Shriberg, E. 1994. Preliminaries to a Theory of Speech Disfluencies, Ph.D. thesis, University of California, Berkeley.
Shriberg, E. 1999. Phonetic Consequences of Speech Disfluency. *Proc. 14th ICPhS San Francisco*, 619-622.
Shriberg, E. 2001. To 'errrr' is human: ecology and acoustics of speech disfluencies, *J. Intl. Phonetic Assoc.* 31(1), 153-169.
Stolcke, A., Shriberg, E. 1996. Statistical language modeling for speech disfluencies. *Proc. ICASSP96*, 405-408.
Vasilescu, I., Candea, M., Adda-Decker, M. 2004. Hésitations autonomes dans 8 langues : une étude acoustique et perceptive. *Actes de MIDL 2004*, 25-30.

Presented at the 14th Conference on Laboratory Phonology, July 25-27 2014, Tokyo, Japan