Phonetic and Prosodic Characteristics of Disfluencies in French Spontaneous Speech

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

[1] University of Louvain, Belgium, [2] University Paris Diderot, France

george@mycontent.gr, mathieu.avanzi@gmail.com

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 production problems, 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) or indicate information status (Arnold et al. 2003). 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 **corpus** used in this study consists of 14 regional varieties of French, recorded in France, Switzerland and Belgium, with 112 native speakers (4 male and 4 female speakers per regional variety, aged between 20 and 80). We focus on semi-directed interviews, which are conversations where the interviewee produces many continuous stretches of speech and has little interaction with the interviewer. 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).

We propose a detailed annotation scheme for disfluencies, based on the typology of Shriberg (2001) and Brugos et al. (2012). **Simple** disfluencies include filled pauses (autonomous fillers, i.e. surrounded by silent pauses, as well as epenthetic vowels), mispronunciations, segment lengthening (drawls), and inter-word pauses. **Structured** disfluencies follow the pattern "reparandum – interruption point – optional editing terms – repair" and include repetitions, insertions, substitutions and deletions. **Complex** disfluencies (i.e. sequences of several simple and/or structured disfluencies) are annotated according to Heeman et al. (2006), leading to a fine-grained description of such events.

We developed software to facilitate the annotation and the extraction of **phonetic and prosodic features** of disfluencies and their context. The features under study are duration, pitch (average, register and contour), pitch and energy slopes, voice quality and vowel quality (F1 and F2 formants of fillers). We also take into consideration speaking rate and the morphosyntactic categories of disfluent tokens, as well as their left and right context.

Our findings are in line with previous studies on the phonetic and prosodic features of disfluencies (e.g. for English, Shriberg 1999, 2001; for French, Vasilescu et al. 2004; for Portuguese, Moniz et al. 2012). 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). The **vowel length distribution** of fillers is significantly different from the distribution of vowel length in fluent contexts, distinguishing disfluency-related lengthening from stress-related prominence. **Filled pauses** exhibit falling intonation contours, with a mean pitch close the speaker's mean pitch; however the formant values of vowels in filled pauses are significantly different from those of the same vowels in fluent contexts. Finally, pitch and energy increase in the **repair** region (compared to the reparandum) possibly signalling the boundaries of each region. We are currently exploring to what extent these observations are affected by **regional variation** and more detailed results will be presented at the conference.

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 / Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer, ver. 5.3.63, 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. / 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., 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. / Moniz, H., Batista, F., Trancoso, I., Mata da Silva, A.I. 2012. Prosodic context-based analysis of disfluencies. *Proc. Interspeech Portland, Oregon*. / 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.