Automatic Detection and Annotation of Disfluencies in Spoken French Corpora

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

[1] Centre Valibel, Institute for Language & Communication, Université de Louvain, Belgium
[2] DTAL, Faculty of Modern & Medieval Languages, University of Cambridge, UK

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

Abstract

In this paper we propose a multi-step system for the semi-automatic detection and annotation of disfluencies in spoken corpora. A set of rules, statistical models and machine learning techniques are applied to the input, which is a transcription aligned to the speech signal. The system uses the results of an automatic estimation of prosodic, part-of-speech and shallow syntactic features. We present a detailed coding scheme for simple disfluencies (filled pauses, mispronunciations, false starts, drawls and intra-word pauses), structured disfluencies (repetitions, deletions, substitutions, insertions) and complex disfluencies. The system is trained and evaluated on a transcribed corpus of spontaneous French speech, consisting of 112 different speakers and balanced for speaker age and sex, covering 14 different varieties of French spoken in Belgium, France and Switzerland.

Index Terms: disfluencies, automatic annotation, corpus processing, spoken French

1. Introduction

An important characteristic of spoken language is the prevalence of a class of phenomena called disfluencies, such as filled pauses, repetitions and false starts. Disfluencies are an interesting phenomenon to study as such, especially in the domain of psycholinguistics; at the same time, it is necessary to take into account this class of phenomena when applying natural language processing techniques to spoken language corpora. In this paper we present a detailed annotation scheme and a modular automatic detection system for disfluencies, targeting the **semi-automatic annotation** of these phenomena in **manually transcribed** data.

Disfluencies can be considered as disruptions of the ideal delivery of speech, as "cases in which a contiguous stretch of linguistic material must be deleted to arrive at the sequence the speaker 'intended', likely the one that would be uttered upon a request for repetition" [40]. However, speech is the most natural way to communicate, and an "ideal" delivery is often elusive, even in speaking styles that permit extensive preparation and rehearsal [37]. Another approach views disfluencies as linguistic devices used to manage the flux of time, facilitating cognitive processes both for the speaker and the listener, while incrementally constructing a message through a series of steps including planning and self-monitoring [25]. Such 'dis'fluencies may draw the listeners' attention to new or complex information ([3], [7]) or they can be used as fluent communicative devices to manage dialogue interaction ([12], [32]). Studies on French include [1], [7] on the use of disfluencies during broadcast political interviews to manage interaction, and their relationship with speech overlaps; [35] presents a qualitative and quantitative analysis of false-starts and repetitions; [35] correlates self-interruptions with dialogue events in the *CID* corpus; and [44] examines the acoustic features of filled pauses in 8 languages including French.

Work on the automatic detection of disfluencies has focused on modelling to improve the performance of ASR systems (e.g. [2]) or to improve parsing performance on transcriptions of spoken language (e.g. [23], for French [34]). Various feature sets and machine learning algorithms have been used: CART decision trees with exclusively lexical features ([33], Spanish); or exclusively prosodic features ([39], English). In [26] it is shown that "the detection of disfluency interruption points is best achieved by a combination of prosodic cues, word-based cues and POS-based cues" while "the onset of a disfluency is best found using knowledge-based rules" and "specific disfluency types can be aided by modelling word patterns". [27] and [28] compared the performance of Hidden Markov Models (HMM), maximum entropy models and Conditional Random Fields (CRF) models in detecting disfluencies using both lexical and prosodic features and found that "discriminative models generally outperform generative models". Working on the English Switchboard corpus, [17] proposed a system for disfluency detection based on CRF models combined with Integer Linear Programming (ILP) rules. In [18] CRF models are compared with ILP, showing that ILP performed better when there was relatively few domain-specific data for training, while [30] incorporate dialogue interaction data into their system. For French, [16] reported a series of experiments on a CRF-based approach to automatic disfluency detection, working on a corpus of recordings of customer service interactions; the objective is to render the manually-transcribed data more readable in order to improve automated opinion mining. [9] have worked on automatic disfluency detection in French air traffic control conversations. [36] demonstrate that it may be more efficient to perform the tasks of syntactic chunking and disfluency detection simultaneously. Finally, [19] suggests using hybrid systems for disfluency detection, i.e. "different detection techniques where each of these techniques is specialised within its own disfluency domain"; we have adopted this approach in the system presented in this paper.

The article is structured as follows: section 2 describes a detailed, language-independent disfluency annotation scheme. A 7-hour corpus of spontaneous French speech, presented in section 3, has been annotated with this protocol. Section 4 describes a multi-step hybrid system for the automatic annotation of disfluencies in spoken French that was trained and evaluated on these data. Section 5 reports the results of the evaluation, and we conclude with perspectives for future work in Section 6.

2. Disfluency Annotation Scheme

The annotation scheme is based on the output of the *DisMo* annotator [10], which is specifically designed for spoken language corpora (we use the version for French in this paper). The annotations are produced on three levels: minimal tokens, multi-word expressions and discourse structuring devices. At the minimal token level, the tokenisation algorithm splits an

input utterance as much as possible, stopping only when further separation would lead to out-of-dictionary forms (e.g. *parce que, aujourd'hui*); however "*tout à fait*" would be split into three minimal tokens. Each minimal token is associated with a part-of-speech tag. These tokens are grouped into multi-word units (MWUs), with a corresponding part-of-speech tag. The combination of these two levels allows us to represent the fact that a series of minimal tokens may have a different morphosyntactic function than its constituent parts. MWUs include numerals, adverbial phrases, named entities and typical multi-word expressions. A third, independent annotation level is dedicated to strings of tokens functioning as discourse structuring devices, including conjunctions and discourse markers. Figure 1 shows how these three levels of annotation are represented on a horizontal timeline under *Praat* [6]. The alignment boundaries of intervals expresses the relationships of containment between the three levels (for example, in the case of "Saint Cloud" in Figure 1, two minimal-level tokens are grouped into an MWU).

The unit of annotation for disfluencies under our system it the minimal token, with optional extensions that add annotations to the syllable level. We propose an annotation scheme for disfluencies that combines aspects from the systems described in [40], [5] and [21]. We have also taken into consideration previous work done on annotation schemes for disfluencies: in English, notably, by the LDC in the context of the MDE project for the annotation of the Switchboard corpus [29] and for French the work of the VoxForge project [13]. We adopt the following taxonomy for disfluencies:

Simple disfluencies are those affecting only one minimal token, and include: **filled pauses**, i.e. autonomous fillers and epenthetic vowels; hesitation-related syllable **lengthening** (in corpora containing a phonetic transcription and syllabification, it is possible to indicate precisely which syllable(s) are lengthened); **lexical false starts**, i.e. fragments of words whose articulation was interrupted by the speaker; **intra-word pauses**; and **mispronounced** words.

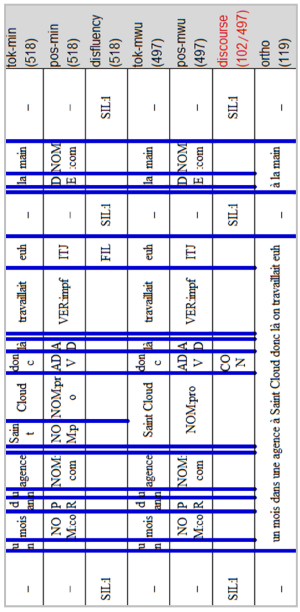
Repetitions are defined as strings of one or more tokens that are repeated by the speaker in exactly the same form, possibly interspersed with filled or silent pauses, as long as this repetition does not serve a grammatical or emphatic purpose. For example, in the utterance "*le le le chien*", we consider that the definite article "le" is repeated; however in the utterance "*il est très très joli*", the repetition of the adjective "très" is emphatic and thus not annotated as a disfluency (cf. [14]).

Structured disfluencies include **substitutions** (the speaker backtracks and modifies some tokens already uttered, using the same syntactic structure; e.g. *normalement je louais enfin je loue toujours un appart*), **insertions** (the speaker backtracks and adds tokens, using the same syntactic structure, e.g. *c'est vrai que Béthune vivre à Béthune…*) and **deletions** (the speaker abandons a series of tokens and continues, or starts a new syntactic structure, e.g. *c'est vraiment en tout cas la parole…*). To be annotated as such, structured disfluencies must follow the reparandum – interregnum – repair pattern; otherwise we use the annotation scheme for complex disfluencies.

Complex disfluencies are the result of combining several structured disfluencies. Unlike [40] we do not limit the annotation scheme to nested complex disfluencies, but adopt the "backtracking table" notation proposed by [21].

Repetitions and structured disfluencies can be described as a sequence of three contiguous regions, following [40]: *(reparandum) * interruption point (interregnum, including optional editing terms) (repair)*. The **reparandum** is the part of the utterance that is repeated or that will be corrected, edited, or deleted. The **interruption point** is the point between the reparandum and the interregnum; this instance in time does not necessarily coincide with the moment the speaker detected the trouble or his intention to alter the utterance. The **interregnum** is the part between the reparandum and the repair. It may optionally include **explicit editing terms**, i.e. words or phrases used by the speaker to signal the correction (e.g. "enfin"). The **repair** is the continuation of the message that follows the disfluency, so that if the first two regions are

Figure 1: *An example of the multi-level annotation system of DisMo is presented in the form of a Praat TextGrid on the left. The table on the right summarises the disfluency annotation scheme with examples from our corpus.*



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 one or more tokens repeated in 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 a été` `INS* INS+FIL INS_ INS_ INS_`
Level 4: Complex disfluencies are a combination of several structured ones		
COM	Complex	`les ac/ les actions enfin les activités enfin professionnelles` `COM COM COM COM COM COM COM COM COM`

removed the remainder is lexically fluent [37]. We have chosen *not* to include discourse markers as "disfluencies". Figure 1 summarises the various disfluency types and provides examples from our corpus, along with the annotation codes associated with each minimal token.

Whenever more than one token are repeated, we use numbering to indicate the repetition pattern: the first token in the repeated sequence is annotated as REP:1, the second as REP:2 etc. Given the series of tokens affected by a repetition (from the reparandum to the repair), it is possible to use pattern matching to find the interruption point and assign the numbering. For repetitions and structured disfluencies (codes REP, SUB, INS and DEL), we use extension to the tags to indicate these regions: an asterisk (*) is appended at the interruption point, i.e. at the end of the tag of the last token of the reparandum; explicit editing terms are indicated by appending ":edt" to the main tag; and the repair region is signalled by appending an underscore (_) to the main tag. Note that for deletions the repair is not annotated, since anything that follows the material deleted could be considered as the repair. The annotation scheme is **hierarchical**, in the sense that Level 1 annotation codes (simple disfluencies) may combine with Level 2 and Level 3 codes; and Level 2 codes (repetitions) may combine with Level 3 codes (structured disfluencies). In the example of an insertion given above, a filled pause produced within the interregnum region of the insertion is annotated as INS+FIL. This allows us to model the co-occurrence of simple disfluencies within the interregnum part or near the interruption point of structured disfluencies.

3. Data

The corpus used in this study consists of recordings extracted from the database "Phonologie du Français Contemporain" [15] and is presented in detail in [4]. The corpus consists of samples of 14 regional varieties of French, recorded in three different countries of Europe: 5 varieties spoken in the Northern part of Metropolitan France (Béthune, Brécey, Lyon, Paris and Ogéviller); 5 varieties spoken in Switzerland (Fribourg, Geneva, Martigny, Neuchâtel and Nyon) and 4 varieties spoken in Belgium (Brussels, Gembloux, Liège and Tournai). For each of the 14 locales, 4 female and 4 male speakers were selected; they were born and raised in the city in which they were recorded. The age of the speakers ranges between 19 and 82 years, and is controlled for each of the 14 groups of speakers, between male and female speakers, and between male and female speakers across the 14 groups. Each speaker was recorded in a reading text task (Reading sub-corpus) and in semi-directed sociolinguistic interviews, in which the informant has minimal interaction with the interviewer, from which 3 minutes of speech per speaker were extracted (Spontaneous Speech sub-corpus). In this study we focus only on the Spontaneous Speech sub-corpus.

4. Method

The automatic disfluency detection system is organised in different modules: each module is responsible for annotating specific types of disfluencies, using the most appropriate method for each type. We have chosen a modular design, similar to the one in [19] because the analysis of our corpus shows that some disfluencies are more common than others.

The detection of filled pauses (FIL) and lexical false starts (FST) from the raw output of an automatic speech recognition system is outside the scope of this paper: we assume that the input already contains these phenomena and that transcription conventions are sufficient to identify them: for example, filled pauses are transcribed as either "euh" or "euhm" (generally, a fixed list of tokens), and lexical false starts are followed by a slash character, as in "mar/" (generally, a transcription convention that can be identified using regular expressions). Although this limits the applications of our system, it is a reasonable assumption given that our initial objective is to facilitate the annotation of existing large French spoken language corpora. Similarly, intra-word pauses (WDP) are identified only on the basis of string matching.

The detection of hesitation-related **lengthening** of syllables is optional, used for corpora that include a reliable automatic or manual syllabification. It is based on a Support Vector Machine (SVM) classifier that takes into account the following prosodic features: the length of the syllable relative to windows of ±3 neighbouring syllables (in order to normalise for articulation rate), its structure (consonants – vowels), its position within the token, and relative pitch over the same windows. The objective is to distinguish increases in syllable length that will be perceived as a hesitation, rather that acoustic cues of prosodic prominence.

The following step in the cascade is the detection of **repetitions** using a Conditional Random Fields (CRF) model, with the following features: word form, part-of-speech (DisMo level-1 tag, i.e. one out of 12 main POS categories, and level-2 tag, i.e. one out of the 64 specific POS tags), whether the token is equal to the following 1, 2… 7 tokens. Combinations of such features are included in the model, over windows of 1 to 3 tokens. The model labels sequences as belonging or not belonging to a repetition cluster (R or 0). The repetition clusters identified by the CRF model are post-processed to find patterns and assign the detailed REP annotation codes.

Using the results of the previous steps (annotation of simple disfluencies and repetitions), the next processing step is to identify **editing** disfluencies. Both lexical and prosodic information is input to a CRF model in order to predict the onset of the disfluency, its reparandum region and the interruption point. The lexical features include token, POS tag (as above), and edit distance with the following 1...7 tokens. An SVM classifier produces hypotheses regarding possible interruption points, based on the following prosodic features: difference in articulation rate, mean pitch and mean intensity 500 ms before and after each possible interruption point (defined as the end of the last syllable of each token); these hypotheses are also input into the CRF model as features of the corresponding token. The output of the model is post-processed to assign the detailed INS, SUB or DEL codes. The user may choose to discard sequences that cannot fit into these patterns, or to leave them annotated with a generic DIS code (depending on whether manual intervention is envisaged or not).

Having separate modules deal with each class of phenomena allows us to fine-tune their parameters to the specificities of each class. We can also use the predictions of one module into the next one: in our system the automatic detection is performed in three steps: (1) repetitions (based on lexical features, SIL and FIL codes); (2) interruption point prediction (taking into account lexical features, POS, and the results of the repetitions module – the fact that a token forms part of a repeated structure is added as a feature to the CRF model) and (3) editing type disfluencies (taking into account

the interruption point predictions, which are entered as a feature to the CRF model). The LEN module is optional.

5. Results and Discussion

We evaluate the performance of each module separately. Table 1 summarises the accuracy, precision and recall measures (as applicable) of the different modules. As expected, the types of disfluencies occurring more frequently in our corpus are better detected by a system based on probabilistic models. In all cases we have used 5-fold cross validation: the 63k corpus is divided in 5 "folds", each containing an approximately equal number of sequences to annotate (sequences are segmented at silent pauses over 500 ms in length). Four folds are used as a training corpus and the resulting model is used to annotate the fifth one that functions as an evaluation; the reported results are the averages of applying this process with all 5 possible combinations.

Table 1: *Overall evaluation of the automatic disfluency detection system*

Disfluency type / Method	Prec	Recall	F-meas
LEN – SVN classifier	78.2%	87.4%	82.5%
REP – CRF model	84.3%	75.8%	79.8%
IP – Interruption point hypotheses	76.7%	52.0%	62.0%
SUB, INS, DEL – CRF models	(see Table 2)		

In Table 1, the measures are calculated on the token level (i.e. number of tokens correctly/incorrectly classified as being part of this specific type of disfluency): this is because LEN affects a single token, while IP is added to the single token that is an interruption point. For repetitions, the CRF model just outputs one code (REP) indicating that the token in question is part of a repeated sequence. The system then uses the iterative *Diff* algorithm to calculate the exact repetition codes.

The detection and annotation of editing-type disfluencies has proven to be much more challenging. Inspired by [16] we have conducted experiments to evaluate four possible BIO (begin-inside-outside) schemes that may be the desired output of the CRF model of this module. Table 2 presents these options: the editing terms are annotated as a separate region (method 1, 3), or included in the reparandum region (method 2, 4); it may be desirable to also predict the repair region (method 1 and 2) or not (method 3 and 4).

We have then evaluated the performance of the alternative BIO systems, in two experiments: in the first one, the correct interruption point was always given to the CRF model (setting an upper limit of the performance); in the second one, the predicted interruption points were used (actual performance). Table 2 summarises the results in terms of precision and recall.

Table 2*: Evaluation of the editing disfluency CRF models*

	Reparandum / Editing terms	Repair region	Prec	Recall	F-meas
Gold standard Interruption Points (upper limit)					
1	Separate	Predict	77.6%	51.4%	61.9%
2	Merged		74.7%	44.7%	55.9%
3	Separate	Ignore	82.4%	62.8%	71.3%
4	Merged		76.9%	53.2%	62.9%
Predicted Interruption Points (actual performance)					
1	Separate	Predict	54.3%	36.5%	43.7%
2	Merged		48.6%	31.3%	38.0%
3	Separate	Ignore	62.6%	42.1%	50.3%
4	Merged		59.2%	36.2%	44.9%

The strategy of considering the reparandum and the editing terms as one contiguous region, contrasted with the repair region, yields marginally better F-measure results.

6. Conclusion and Perspectives

The aim of this paper was threefold: we presented a protocol to annotate disfluencies in spoken French corpora, an annotation tool to facilitate human annotators, and the results of training and evaluating different machine learning algorithms for the detection and annotation of disfluencies. The evaluation of the system was performed on a 63k token corpus of French spontaneous speech. Our results indicate that an automatic detection of the majority of speech disfluencies in a transcribed corpus is feasible, even if, as expected, some phenomena are more easily recognizable automatically than others. Currently there are no publicly-available, large-scale corpora of spoken French containing a gold-standard, detailed annotation of disfluencies. Our intention is to apply our system to several large spoken French corpora, that already are (or will soon become) publicly available, using an iterative re-training methodology to improve the accuracy of the automatic annotation. This annotation campaign will open up perspectives for further research, especially in the domain of disfluency-aware syntactic parsing of spoken French.

7. Acknowledgments

The second author's research was funded by the Swiss National Science Foundation (SNF grant Advanced Postdoc.Mobility n° P300P1-147781).

8. References

[1] Adda G., Adda-Decker M., Barras C., Boula de Mareüil P., Habert B., Paroubek P., "Speech Overlap and Interplay with Disfluencies in Political Interviews", *Proc. ParaLing '07*, pp. 41-46, 2007.

[2] Adda-Decker M., Habert B., Barras C., Boula de Mareüil Ph., Paroubek P., "Une étude des disfluences pour la transcription automatique de la parole et l'amélioration des modèles de langage". *Actes des 25èmes journées d'étude sur la parole*, 2004.

[3] Arnold J. E., Fagnano M., Tanenhaus M. K., "Disfluencies signal theee, um, new information", *Journal of Psycholinguist Research*, vol. 32, no. 1, pp. 25-36, 2003.

[4] Avanzi M., "A Corpus-Based Approach to French Regional Prosodic Variation", *Nouveaux cahiers de linguistique française*, vol. 31, pp. 309-323, 2014.

[5] Brugos A., Shattuck-Hufnagel S., "A proposal for labelling prosodic disfluencies in ToBI", Poster presentation, *Advancing Prosodic Transcription for Spoken Language Science and Technology*, Stuttgart, Germany, 2012.

[6] Boersma P., Weenink D., *Praat: doing phonetics by computer*, ver. 5.3.63, www.praat.org

[7] Bosker H. R., Quené H., Sanders T., De Jong N. H. "The processing of disfluencies in native and non-native speech", Talk presented at the *19th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP 2013)*, Marseille, France, 2-4 September, 2013.

[8] Boula de Mareüil P., Habert B., Bénard F., Adda-Decker M., Barras C., Adda G., Paroubek P., "A quantitative study of disfluencies in French broadcast interviews", *Proceedings of DiSS 2005, Disfluency in Spontaneous Speech Workshop*, Aix-en-Provence, France, 10-12 September 2005.

[9] Bouraoui J.-L., Vigouroux N., "Traitement automatique de disfluences dans un corpus linguistiquement contraint", *Actes de TALN*, 2009.

[10] Christodoulides G., Avanzi M., Goldman J.-Ph., "DisMo: A Morphosyntactic, Disfluency and Multi-Word Unit Annotator", *Proc. of 9th International Conference on Language Resources and Evaluation (LREC)*, pp. 3902-3907, 2014.

[11] Christodoulides G., "Praaline: Integrating tools for speech corpus research", *Proc. of the 9th International Conference on Language Resources and Evaluation (LREC)*, pp. 31-34, 2014.

[12] Clark H., Fox Tree J., "Using uh and um in spontaneous speaking", *Cognition*, vol. 84, pp. 73-111, 2002.

[13] Clavel C., Adda G., Cailliau F., Garnier-Rizet M., Cavet A., Chapuis G., Cournicous S., Danesi C., Daquo A.-L., Deldossi M., Guillemin-Lanne S., Seizou M., Suignard P., "Spontaneous Speech and Opinion Detection: Mining Call-centre Transcripts", *Language Resources and Evaluation*, vol. 1, pp. 40, 2013.

[14] Dister A., "'parler sans accent pour moi c'est sans sans sans bafouiller' Quelles répétitions de formes en français parlé?", *Actes du 4e Congrès mondial de linguistique française*, 2014.

[15] Durand J., Laks B., Lyche C., (eds.), *Phonologie, variation et accents du français*, Paris, Hermès, 2009.

[16] Dutrey C., Clavel C., Rosset S., Vasilescu I., Adda-Decker M., "A CRF-based Approach to Automatic Disfluency Detection in a French Call-Centre Corpus", *Proceedings of Interspeech*, pp. 2897-2901, 2014.

[17] Georgila K., "Using integer linear programming for detecting speech disfluencies", *Proceedings of Human Language Technologies, Companion Volume: Short Papers*, Boulder, Colorado, 31 May – 5 June, 2009.

[18] Georgila K., Wang N., Gratch J., "Cross-domain speech disfluency detection", *Proc. of SIGDIAL*, pp. 237-240, 2010.

[19] Germesin S., Becker T., Poller P., "Hybrid Multi-step Disfluency Detection", *Proceedings of the 5th International Workshop on Machine Learning for Multimodal Interaction*, 2008.

[20] Goldman J.-Ph., "EasyAlign: an automatic phonetic alignment tool under Praat", *Proc. Interspeech*, Italy, 2011.

[21] Heeman P., McMillin A., Scott Yaruss J., "An annotation scheme for complex disfluencies". *Proceedings of the 9th ICSLP*, pp. 1081-1084, 2006.

[22] Henry S., Pallaud B., "Word fragments and repeats in spontaneous spoken French", *Proceedings of the Disfluency in Spontaneous Speech Workshop DiSS'03*, pp. 77-80, 2003.

[23] Jorgensen F., "The Effects of Disfluency Detection in Parsing Spoken Language", *Proc. of NODALIDA*, pp. 240-244, 2007.

[24] Lafferty J., McCallum A., Pereira F., "Conditional Random Fields: Probabilistic Models for Segmenting and Labelling Sequence Data", *Proceedings of the International Conference on Machine Learning*, pp. 282-289, 2001.

[25] Levelt W. J. M., *Speaking: From Intention to Articulation. Cambridge*, MA: MIT Press, 1989.

[26] Liu Y., Shriberg E., Stolcke A., "Automatic Disfluency Identification in Conversational Speech Using Multiple Knowledge Sources", *Proc. 8th Eurospeech*, Geneva, 2003.

[27] Liu Y., Shriberg E., Stolcke A., Harper M., "Comparing HMM, maximum-entropy and conditional random fields for disfluency detection". *Proceedings of Interspeech*, Lisbon, Portugal, pp. 3313-3316, 2005.

[28] Liu Y., Shriberg E., Stolcke A., Hillard D., Ostendorf M., Harper M., "Enriching speech recognition with automatic detection of sentence boundaries and disfluencies". *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1526-1540, 2006.

[29] Mateer M., Taylor A., Disfluency annotation stylebook for the Switchboard corpus, Manuscript, Department of Computer and Information Science, University of Pennsylvania, 1995.

[30] Mieskes M., Strube M., "A Three-stage Disfluency Classifier for Multi Party Dialogues", *Proc. of the 6th International Language Resources and Evaluation (LREC)*, pp. 2681-2686, 2008.

[31] Moniz H., Batista F., Trancoso I., Mata da Silva A.I., "Prosodic context-based analysis of disfluencies". *Proceedings of Interspeech*, Portland, Oregon, 2012.

[32] Moniz H., Trancoso I., Mata da Silva A.I., "Classification of disfluent phenomena as fluent communicative devices in specific prosodic contexts", *Proc. Interspeech*, UK, pp. 1719-1722, 2009.

[33] Moreno I., Pineda L., "Speech repairs in the DIME corpus", *Research in Computing Science*, vol. 20, pp. 63-74, 2006.

[34] Nasr A., Bechet F., Favre B., Bazillon T., Deulofeu J., Valli A., "Automatically enriching spoken corpora with syntactic information for linguistic studies", *Proceedings of 9th International Conference on Language Resources and Evaluation (LREC)*, 2014, p. 854-858.

[35] Pallaud B., Rauzy S., Blanche P., "Auto-interruptions et disfluences en français parlé dans quatre corpus du CID", *TIPA. Travaux interdisciplinaires sur la parole et le langage*, vol. 29, 2013, Available online: http://tipa.revues.org/995

[36] Peshkov K., Prévot L., Rauzy S., Pallaud B., "Categorising Syntactic Chunks for Marking Disfluent Speech in French Language", *Proceedings of DiSS*, pp. 59-62, 2013.

[37] Shriberg E. "To 'errrr' is human: ecology and acoustics of speech disfluencies", *Journal of the International Phonetic Association*, vol. 31, no. 1, pp. 153-169, 2001.

[38] Shriberg E., "Phonetic Consequences of Speech Disfluency", *Proc. of the 14th ICPhS*, San Francisco, pp. 619-622, 1999.

[39] Shriberg E., Bates R., Stolcke A., "A prosody-only decision-tree model for disfluency detection", *Proceedings of Eurospeech,* pp. 2383-2386, 1997.

[40] Shriberg E., *Preliminaries to a Theory of Speech Disfluencies*, Ph.D. thesis, University of California, Berkeley, 1994.

[41] Snover M., Dorr B., Schwartz R., "A lexically-driven algorithm for disfluency detection". In Dumais S., Marcu D., Roukos S. (eds.) *Proceedings of the HLT-NAACL*, Boston, USA, Association for Computational Linguistics, 2004.

[42] Stolcke A., Shriberg E. "Statistical language modelling for speech disfluencies", *Proc. of ICASSP*, pp. 405-408, 1996.

[43] Strassel S., *Simple Metadata Annotation Specification*, Linguistic Data Consortium, 2004.

[44] Vasilescu I., Candea M., Adda-Decker M., "Hésitations autonomes dans 8 langues : une étude acoustique et perceptive", *Actes de MIDL*, pp. 25-30, 2004.