

# Disfluences dans la parole spontanée conversationnelle : détection automatique utilisant des indices lexicaux et acoustiques

Camille Dutrey<sup>1, 2, 3, 4</sup> Sophie Rosset<sup>2</sup> Martine Adda-Decker<sup>2, 3</sup> Chloé Clavel<sup>5</sup>  
Ioana Vasilescu<sup>2</sup>

(1) EDF R&D, 92 141 Clamart, France

(2) LIMSI-CNRS, 91 403 Orsay, France

(3) LPP, Université Paris III – Sorbonne Nouvelle, 75 005 Paris, France

(4) Université Paris Sud, 91 405 Orsay, France

(5) Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, 75 014 Paris, France

{dutrey,madda,rosset,ioana}@limsi.fr, chloe.clavel@telecom-paristech.fr

## RÉSUMÉ

---

La parole spontanée est caractérisée par la présence de nombreux éléments disfluents qui peuvent s'avérer très informatifs quant au déroulement et à la compréhension du dialogue, tout en étant un obstacle à l'application de certains traitements automatiques. Nous proposons une méthode pour la détection automatique de disfluences dans des données conversationnelles en français. Nous développons et évaluons une méthode d'étiquetage séquentiel à base de CRF, reconnus pour donner de très bons résultats sur des tâches similaires. Notre système s'appuie sur des indices à la fois lexicaux et prosodiques : la combinaison de traits et de patrons construits à partir des transcriptions manuelles et du signal acoustique donne les meilleurs résultats pour de la détection de frontière, mais les indices acoustiques semblent dégrader les résultats pour une tâche conjointe d'identification de frontière et de structuration de la disfluence.

## ABSTRACT

---

**Disfluencies in conversational spontaneous speech : automatic detection using lexical and acoustic features**

*Spontaneous speech may be characterized by a relatively high rate of disfluencies. Keeping disfluencies in transcriptions may help to better understand the relative progress in dialogue, however their presence may harm a smooth application of various automatic processings. In this contribution, we propose a method to automatic disfluency detection in conversational speech. To this aim, we implemented a CRF (Conditional Random Field) approach to tag our data into disfluent/non-disfluent regions. CRFs are known to perform quite well on this kind of segmentation/labeling tasks. Implemented features include both lexical, morpho-syntactic and acoustic-prosodic information. Contrastive experiments show that the addition of features including information extracted from the speech signal beyond the mere transcription level give the best overall results in automatic disfluency region detection but seem to degrate the results for both region detection and structuration task.*

**MOTS-CLÉS** : champs aléatoires conditionnels, disfluences, parole conversationnelle, parole spontanée.

**KEYWORDS**: conditional random fields, disfluencies, conversational speech, spontaneous speech.

---

# 1 Introduction

La parole spontanée est caractérisée par la présence de nombreux éléments disfluents. Très variés du point de vue de leur forme et de leur fonction – hésitations vocaliques, marqueurs discursifs, auto-corrrections, faux-départs, *etc.* –, ils sont souvent considérés comme des signaux du message oral en construction, à la manière d’un brouillon précédent la version finale d’un texte écrit.

Si ces phénomènes ont été longtemps considérés comme préjudiciables à la qualité du discours et à sa compréhension, ils sont depuis quelques années de plus en plus étudiés ; leur lien avec le discours est plus spécifiquement analysé (voir par exemple Shriberg (1994); Beliao et Lacheret (2013); Vasilescu *et al.* (2010)). Malgré ces nombreuses études sur la richesse informationnelle des marques de l’oral spontané, le terme plutôt connoté négativement de « phénomène disfluent » perdure. Nous l’adoptons selon la définition proposée par Blanche-Benveniste *et al.* (1990) : une disflue intervient lorsque le déroulement syntagmatique de l’énoncé est brisé.

L’objectif de cette étude est de proposer une méthode permettant de détecter automatiquement les disfluences d’édition dans des données d’oral spontané conversationnel en français. Notre méthode est développée et évaluée sur des données conversationnelles issues de centres d’appels, fournies par EDF : l’information contenue dans ces conversations est très précieuse pour les industriels, car elle peut être utilisée pour améliorer la connaissance client et la gestion de la relation clients/entreprise. En particulier, elle peut contribuer à mettre en évidence les dysfonctionnements dans les pratiques relationnelles et les stratégies de communication clients/agents. Des données similaires ont fait l’objet de recherches approfondies en analyse de l’oral, particulièrement dans le projet DECODA (Béchet *et al.*, 2012).

Cependant, relativement peu de travaux ont été réalisés en détection automatique des disfluences sur des données orales en français : on peut citer le DISTAGGER, outil de détection des disfluences à base de patrons élaborés manuellement (Constant et Dister, 2010), ainsi que les travaux de Bouraoui et Vigouroux (2009) qui proposent dans une perspective phraséologique une méthode également à base de patrons pour la détection d’un certain nombre de phénomènes disfluents.

Notre contribution à la problématique de la détection automatique de disfluences sur des données orales spontanées est double : d’une part, concernant le phénomène à détecter, nous nous attachons au repérage de segments disfluents en vue de nettoyer les données pour des traitements avals mais surtout de les utiliser pour mieux comprendre et modéliser le discours ; d’autre part, nous faisons l’hypothèse qu’une approche hybride, basée à la fois sur le texte et sur des indices acoustiques, permet de mieux modéliser les disfluences, liées à ces deux dimensions.

Nous présentons en section 2 nos données d’expérimentation et le modèle de disfluences ayant servi à leur annotation ; la section 3 décrit la méthode de détection automatique de disfluences que nous avons développée dans cette article ; celle-ci est testée et évaluée en section 4 ; enfin, la section 5 présente nos conclusions et les perspectives de travail qui en découlent.

## 2 Corpus de dialogues : conversations en centres d’appels

Nous utilisons le corpus VOXFACTORY, développé au sein du projet éponyme (Clavel *et al.*, 2013), continuité du projet Infom@gic–Callsurf. L’entreprise d’électricité française EDF a mené une campagne d’enregistrement dans un centre d’appels, produisant le corpus de données VOXFACTORY.

Cet ensemble de données, composé de dialogues entre des agents EDF et des clients particuliers, couvre un grand nombre de sujets liés aux services proposés par l'entreprise, comme par exemple des ouvertures de contrats, des questions liées à la facturation, des problèmes techniques, *etc.*

Un sous-ensemble du corpus VOXFACTORY, le corpus VOXDISS (60 conversations entre des agents EDF et des clients individuels, pour environ 10 heures de parole), a été manuellement annoté en disfluences d'édition par l'entreprise Vecsys, partenaire du projet Vox Factory, à l'aide de l'outil d'annotation Transcriber (Barras *et al.*, 1998). Nous utilisons le corpus VOXDISS dans nos expérimentations pour évaluer notre méthode de détection automatique des disfluences d'édition en contexte de dialogue oral spontané homme/homme en français.

La stratégie suivie pour annoter ce corpus s'appuie sur le guide d'annotation de méta-données conçu par le Linguistic Data Consortium (Strassel, 2004). Quatre classes d'événements disfluents ont été annotées : les auto-corrrections (d'un mot simple ou d'un groupe de mots, accompagnée d'un changement de sens), les répétitions (d'un mot simple ou d'un groupe de mots), les faux-départ (abandon de l'énoncé) et les disfluences complexes (succession ou imbrication de différents types de disfluences d'édition). L'annotation des disfluences consiste en une annotation séquentielle des formes de surface, avec l'identification des trois parties distinctes structurant la zone disfluente, comme formalisé dans le modèle qui suit (en mettant en avant un exemple de répétition), selon la structuration de Levelt (1983) et la terminologie de Shriberg (1994) :

**(reparandum) \* ( phase d'édition ) [repair]**  
« je voulais payer (par) \* ( euh ben ) [par] carte »

FIGURE 1 – Structure interne des disfluences d'édition illustrée avec l'exemple d'une répétition avec phase d'édition issue du corpus VOXDISS.

Dans ce modèle, le *reparandum* réfère à l'énoncé que le locuteur souhaite modifier. La phase d'édition, qui est une phase optionnelle, fait référence au segment de préparation de la correction. Le symbole \* représente le point d'interruption qui initie la phase d'édition (le moment où le locuteur décide de modifier son énoncé). Le *repair* représente la phase de correction de l'énoncé.

### 3 Détection automatique des disfluences d'édition

Nous avons choisi de développer une méthode de détection automatique à l'aide des champs aléatoires conditionnels. En effet, non seulement cette méthode d'apprentissage automatique est connue pour donner d'excellents résultats sur des tâches similaires à la nôtre (par exemple pour la détection d'entités nommées) mais surtout les recherches récentes ayant abordé la détection de disfluences ont montré que les CRF donnaient les meilleurs résultats par rapport à d'autres modèles comme les modèles de Markov cachés ou l'entropie maximale (Liu *et al.*, 2006).

Nous présentons en section 3.1 les étiquettes définies pour les deux expériences de détection actuellement menées. La section 3.2 décrit notre système à base de CRF, son principe et sa mise en œuvre. Enfin, en 3.3 nous présentons les traits lexico-acoustiques et les patrons d'apprentissage que nous avons utilisés dans nos expériences.

### 3.1 Définition de la tâche et étiquettes de disfluences

La majorité des études existantes sur la détection automatique de disfluences a pour objectif la suppression des événements disfluents dans l'idée notamment de nettoyer les transcriptions. Le but principal de ces recherches est d'améliorer la lisibilité des données orales ou d'améliorer les modules aval de traitement automatique du langage naturel (voir par exemple Qian et Liu (2013)). La suppression se fait parfois même directement dans le flux d'une autre tâche, conjointe, par exemple d'analyse syntaxique (Rasooli et Tetreault, 2013). En fonction des objectifs de détection, l'accent est mis sur les événements disfluents de type pauses remplies (*filler words*) et/ou disfluences d'édition, et plus précisément sur une partie seulement des éléments composant leur structure (comme le *reparandum* ou le *repair*). En supprimant les zones disfluentes, la plupart des études se concentrent donc exclusivement sur la détection du *reparandum* (Georgila, 2009; Rasooli et Tetreault, 2013), éventuellement associé à la phase d'édition.

En complément de cet aspect de suppression de la zone disfluente à proprement parler, qui nous semble également important à des fins d'amélioration de la lisibilité des données, nous nous intéressons tout autant à la manière dont le locuteur modifie son énoncé qu'à la partie qu'il souhaite modifier, considérée comme la partie erronée. Ce segment de correction, le *repair*, peut être précieux dans la compréhension d'un dialogue : dans le cadre d'une auto-correction, le lexique choisi pour se corriger n'est pas fortuit (est-ce une correction visant à préciser le sens du propos ? un choix de synonyme visant à rendre le propos plus neutre ? etc.).

Nous avons ainsi choisi, pour initier nos travaux sur la détection automatique de disfluences d'édition, de réaliser deux séries d'expérience : la première (*XP1*) consiste à repérer les frontières des zones disfluentes, sans chercher à identifier les différents éléments composant la structure interne de cette zone. Notre seconde expérience (*XP2*) consiste à repérer un premier niveau de structure de la zone disfluente, en cherchant à séparer la partie « corrigée » (soit l'agrégat du *reparandum* et de la phase d'édition) de la zone de « réparation » de la disfluence (soit le *repair*).

**Étiquette définie pour l'*XP1*** Dans notre première expérience nous cherchons à identifier la totalité d'un segment disfluent (*reparandum*, phase d'édition et *repair*) de manière à pouvoir l'isoler du reste de l'énoncé. Nous avons donc défini une étiquette unique, déclinée selon si le token étiqueté débute un segment ou se trouve à l'intérieur de ce dernier (respectivement ***B-Dis*** et ***I-Dis***). Dans les autres cas, les tokens sont étiquetés « NULL ». Ainsi, dans l'exemple présenté en section 2, l'étiquetage de référence est le suivant : « je<sub>NULL</sub> voulais<sub>NULL</sub> payer<sub>NULL</sub> par<sub>B-Dis</sub> euh<sub>I-Dis</sub> ben<sub>I-Dis</sub> par<sub>I-Dis</sub> carte<sub>NULL</sub> ».

**Étiquettes définies pour l'*XP2*** Pour cette deuxième expérience nous avons défini des étiquettes différenciant la partie réellement disfluente (*reparandum* et phase d'édition le cas échéant) de sa réparation (*repair*). Cette structuration plus fine nous permet d'isoler d'une part la zone de la disfluence considérée pour le nettoyage (rapprochement de l'énoncé à une forme d'énonciation « finalisée ») et d'autre part les éléments lexicaux choisis par le locuteur pour « corriger » son énoncé. Dans cette perspective, les étiquettes suivantes ont été définies : pour la partie disfluente les labels ***B-Rpd-Edp*** (token initiant le segment) et ***I-Rpd-Edp*** (token à l'intérieur du segment) ; pour la partie *repair* les étiquettes ***B-Rpr*** (token initiant le segment) et ***I-Rpr*** (token à l'intérieur du segment) ; enfin, pour les tokens hors disfluence d'édition, l'étiquette « NULL ». Pour le même exemple que précédemment, l'étiquetage de référence devient : « je<sub>NULL</sub> voulais<sub>NULL</sub> payer<sub>NULL</sub> par<sub>B-Rpd-Edp</sub> euh<sub>I-Rpd-Edp</sub> ben<sub>I-Rpd-Edp</sub> par<sub>B-Rpr</sub> carte<sub>NULL</sub> ».

## 3.2 Choix des CRF et principes de mise en œuvre

Les champs aléatoires conditionnels (ou CRF, cf. Lafferty *et al.* (2001)) constituent un modèle d'apprentissage supervisé utilisé pour l'étiquetage séquentiel. Une de leur particularité est de pouvoir prendre en compte de grands ensembles de traits potentiellement redondants et d'intégrer des dépendances structurelles entre les classes que l'on cherche à identifier.

Nous avons utilisé la mise en œuvre des CRF fournie par Wapiti<sup>1</sup> (Lavergne *et al.*, 2010). Nous utilisons l'algorithme `rprop`, qui non seulement a déjà donné de bons résultats sur des tâches de détection structurées et complexes proches de la notre (voir par exemple Dinarelli et Rosset (2011)) mais surtout donne les meilleurs résultats sur notre corpus de développement. Le critère d'arrêt de l'algorithme a été empiriquement fixé à 500 itérations maximum, compte tenu des résultats obtenus sur le corpus de développement.

## 3.3 Traits et patrons lexico-acoustiques utilisés par le système

Nous avons développé une version minimal de notre système (*baseline*) sur le modèle de celle utilisée par Qian et Liu (2013). Les patrons sont simplement basés sur la forme fléchie des mots et sont appliqués sur une fenêtre de -2/+2 mots. N'utiliser que les informations lexicales au niveau local permet de mieux mesurer l'apport des traits et des patrons lexico-acoustiques utilisés pour enrichir le système.

Nous distinguerons trois versions du système : CRF\_L (basé sur des patrons lexicaux, cf. tableau 1), CRF\_A (basé des patrons acoustiques, cf. tableau 2) et CRF\_LA (basé sur des patrons lexico-acoustiques). Les patrons lexicaux considèrent des traits purement linguistiques, comme les parties du discours (générées automatiquement par une version adaptée au français du tagger Brill (Allauzen et Bonneau-Maynard, 2008)). Les patrons acoustiques sont des traits classiquement employés pour caractériser la parole d'un point de vue acoustique. Tous les attribus numériques continus ont été discrétisés à l'aide de l'outil `discretize4crf`<sup>2</sup>.

---

### Séquences lexicales : unigrammes/bigrammes, fenêtre [-2,+2]

---

- forme fléchie ;
- traits de partie du discours.

---

### Expressions rationnelles : unigrammes, fenêtre mot courant

---

- préfixes/suffixes aux positions 1 : 4 (soit 8 traits en tout) ;

---

### Patrons binaires (oui/non) en expressions rationnelles : unigrammes, fenêtre [-1,+1]

---

- est entièrement en capitales ? (fenêtre mot courant)
  - commence par une capitale ? (fenêtre mot courant)
  - comporte une ponctuation ? (fenêtre [-1,+1])
  - est une ponctuation ? (fenêtre mot courant)
  - comporte une ponctuation (premier/dernier caractère exclus) ? (fenêtre mot courant)
  - comporte un chiffre ? (fenêtre [-1,+1])
  - est un nombre ? (fenêtre mot courant)
- 

TABLE 1 – Patrons utilisés par le modèle de CRF reposant sur des traits lexicaux (CRF\_L).

---

1. <http://wapiti.limsi.fr/>.

2. <https://gforge.inria.fr/projects/discretize4crf/>.

---

**Séquences acoustiques : unigrammes/bigrammes, fenêtre [-2,+2]**

---

- Prononciation du mot ;
  - Durée du mot ;
  - Nombre de phonèmes du mot ;
  - Durée moyenne des phonèmes du mot.
- 

TABLE 2 – Patrons utilisés par le modèle de CRF reposant sur des traits acoustiques (CRF\_A).

## 4 Expériences

Nous avons mené deux expériences pour la détection de disfluences, correspondant aux tâches et étiquettes définies précédemment (XP1 et XP2). Nous détaillons en section 4.1 le corpus d'expérimentation VOXDISS. Enfin, nous décrivons et discutons les résultats en sections 4.2 (XP1) et 4.3 (XP2).

Les mesures utilisées dans nos expériences correspondent aux mesures d'évaluation classiques de Précision, Rappel, F-mesure et SER ou *Slot Error Rate* (Makhoul *et al.*, 1999)<sup>3</sup>.

### 4.1 Données d'expérimentation

Le tableau 3 présente la subdivision du corpus VOXDISS opérée pour l'entraînement, le développement et le test de notre système de détection des disfluences d'édition. Nous avons sélectionné aléatoirement (au sein des quartiles du corpus) des conversations suivant une distribution homogène sur la durée des dialogues.

Corpus	Nb. appels	Dur. tot.	Dur. moy.	Nb. moy. loc.	Nb. moy. dis.	Densité lex. dis.
Train	48	9h19	12m04	2.1	62.54	10.91%
Dev	5	1h01	12m20	2.2	54.20	8.6%
Test	7	1h13	11m29	3	37.14	7.4%

TABLE 3 – Subdivision du corpus VOXDISS en données d'entraînement, de développement et de test. Densité lex. dis. : pourcentage des mots du corpus dans une zone disfluente.

### 4.2 Identification des frontières des segments disfluents (XP1)

Nous cherchons ici à identifier la frontière des disfluences d'édition, *reparandum*, phase d'édition et *repair* sans structuration interne. L'objectif de cette première tâche est d'identifier, localiser et quantifier les disfluences d'édition dans des données orales spontanées. Le tableau 4 présente les résultats obtenus par les différentes versions du système (Baseline, CRF\_L, CRF\_A et CRF\_LA) sur les données de test.

---

3. Le SER mesure les erreurs en terme d'insertions, de substitutions (frontières et types) et de suppressions. Cette mesure s'apparente au WER (ou *Word Error Rate*), utilisé en Reconnaissance Automatique de la Parole.

Modèle	Précision	Rappel	F-mesure	SER
Baseline	0.490	0.187	0.270	0.823
CRF_L	0.523	0.254	0.342	0.776
CRF_A	0.422	0.183	0.255	0.810
CRF_LA	0.523	0.257	0.345	0.765

TABLE 4 – Performances de détection de disfluences par CRF sans isolement du *repair*.

L'observation la plus frappante est la dégradation des résultats dès lors que l'on se base sur des indices acoustiques uniquement. Les résultats obtenus par CRF\_A sont en effet moins bons que ceux obtenus par la baseline, à la fois en Précision, Rappel, F-mesure et Slot Error Rate.

Toutefois, les meilleurs résultats sont obtenus par CRF\_LA, lorsque l'on combine des patrons lexicaux et acoustiques, ce qui laisse présager de bonnes perspectives d'amélioration sur cette voie. En effet, cette configuration utilisant des indices mixtes obtient un gain de 7,5 points en F-mesure (F=27% pour la baseline *versus* 34,5% pour CRF\_LA) et de 5,8 points pour le SER (SER=82,3% pour la baseline *versus* 76,5% pour CRF\_LA).

Il est également intéressant de voir que la précision reste stable entre CRF\_L et CRF\_LA : l'ajout des indices acoustiques, qui utilisés seuls donnent les moins bons résultats, permet un gain sur toutes les autres mesures sans pour autant dégrader la précision.

### 4.3 Identification et structuration des segments disfluent (XP2)

Nous cherchons ici à identifier la frontière des disfluences d'édition, *reparandum*, phase d'édition et *repair* avec structuration interne. L'objectif de cette seconde tâche est similaire à celui poursuivi précédemment (XP1), mais cette fois le système cherche à isoler la partie réellement disfluente (*reparandum* + phase d'édition) de sa réparation (*repair*). Les résultats obtenus par les différentes versions du système sur les données de test sont présentés dans le tableau 5.

Modèle	Précision			Rappel			F-mesure			SER
	Rpd-Edp	Rpr	All	Rpd-Edp	Rpr	Rpd-Edp	All	Rpr	All	
Baseline	0.596	0.723	0.654	0.225	0.283	0.251	0.327	0.407	0.363	0.790
CRF_L	0.550	0.626	0.586	0.233	0.292	0.259	0.327	0.399	0.360	0.794
CRF_A	0.482	0.557	0.517	0.206	0.255	0.228	0.289	0.350	0.316	0.816
CRF_LA	0.586	0.688	0.632	0.260	0.311	0.283	0.360	0.429	0.391	0.752

TABLE 5 – Performances de détection de disfluences par CRF avec isolement du *repair*. Rpd-Edp : *reparandum* + phase d'édition. Rpr : *repair*.

Considérant les résultats obtenus pour cette seconde tâche, notons la supériorité du modèle combinant des patrons lexicaux et acoustiques (CRF\_LA). Excepté en précision, CRF\_LA obtient en effet les meilleurs résultats pour toutes les mesures d'évaluation, avec un gain de 2,8 points en F-mesure pour l'ensemble des étiquettes (F=36,3% pour la baseline *versus* 39,1% pour CRF\_LA) et de 3,8 points pour le SER (SER=79% pour la baseline *versus* 75,2% pour CRF\_LA).

En prenant en compte les résultats de l'XP1, il est clair que cette deuxième tâche est davantage per-

formante. Ces résultats sont encourageants quant à la poursuite de nos travaux sur l'identification de la structure des disfluences d'édition.

Enfin, le *repair* semble plus évident à localiser, son identification donnant de meilleurs résultats dans toutes les configurations du système. Cela peut être expliqué par le fait que la zone actuellement étiquetée *Rpd-Edp* nécessite également une structuration plus fine, dans la mesure où elle combine actuellement *reparandum* et phase d'édition, deux zones aux propriétés lexicales relativement différentes.

## 5 Conclusion et Perspectives

Nous avons présenté dans cette étude deux expériences de détection automatique de disfluences d'édition sur des données orales conversationnelles en français. Dans l'objectif de nettoyer les données orales de ces zones disfluentes et d'utiliser ces phénomènes en compréhension du discours, nous avons développé une méthode à base de CRF utilisant à la fois des indices lexicaux et acoustiques. Nous avons développé un premier jeu de patrons pour notre système, que nous avons testés individuellement puis combinés. Nous avons envisagé une détection des zones disfluentes avec isolement du *repair* et sans structuration interne.

Les résultats préliminaires obtenus sont prometteurs, et nous confortent dans l'idée d'utiliser conjointement des indices lexicaux et acoustiques. Actuellement, les patrons acoustiques sont moins développés que les patrons lexicaux, ce qui explique la faiblesse des résultats obtenus avec cette méthode : nous envisageons une utilisation plus fine des traits déjà identifiés, et allons enrichir cette méthode de nouveaux traits et patrons acoustiques (fréquence fondamentale, formants, silences, *etc.*). De plus, à ce stade de nos recherches, nous travaillons exclusivement sur les transcriptions manuelles des conversations. Cela explique également la plus grande efficacité des patrons lexicaux. Nous projetons d'évaluer notre système de détection des disfluences en conditions réelles sur des sorties de reconnaissance automatique de la parole, et faisons l'hypothèse que dans cette configuration le modèle basé sur des patrons acoustiques gagne en performance.

Par ailleurs, les résultats obtenus en structurant les disfluences d'édition (isolement du *repair*) sont meilleurs et nous confortent également dans l'objectif de réaliser une identification structurée des disfluences d'édition. Ainsi, nous envisageons d'isoler la phase d'édition du *reparandum* dans la zone disfluente, et de tester l'utilisation d'étiquettes différenciant les quatre classes de disfluences présentes dans nos données (répétitions, auto-corrections, faux-départ, disfluences complexes).

## Remerciements

Ce travail a été partiellement financé par la convention CIFRE 2011/0916.



# Références

- ALLAUZEN, A. et BONNEAU-MAYNARD, H. (2008). Training and Evaluation of POS Taggers on the French MULTITAG Corpus. In *Proceedings of LREC*.
- BARRAS, C., GEOFFROIS, E., WU, Z. et LIBERMAN, M. (1998). Transcriber : a Free Tool for Segmenting, Labeling and Transcribing Speech. In *Proceedings of LREC*, pages 1373–1376.
- BELIAO, J. et LACHERET, A. (2013). Disfluency and Discursive Markers : When Prosody and Syntax Plan Discourse. In *Proceedings of DiSS*, pages 5–8.
- BLANCHE-BENVENISTE, C., BILGER, M., ROUGET, C. et VAN DEN EYNDE, K. (1990). *Le français parlé : Études grammaticales*. Éditions du CNRS.
- BOURAOU, J.-L. et VIGOUROUX, N. (2009). Traitement automatique de disfluences dans un corpus linguistiquement contraint. In *Actes de TALN*.
- BÉCHET, F., MAZA, B., BIGOUROUX, N., EL-BÈZE, M., DE MORI, R. et ARBILLOT, E. (2012). DECODA : a Call-center Human-human Spoken Conversation Corpus. In *Proceedings of LREC*.
- CLAVEL, C., ADDA, G., CAILLIAU, F., GARNIER-RIZET, M., CAVET, A., CHAPUIS, G., COURCINOUS, S., DANESI, C., DAQUO, A.-L., DELDOSSI, M., GUILLEMIN-LANNE, S., SEIZOU, M. et SUIGNARD, P. (2013). Spontaneous Speech and Opinion Detection : Mining Call-centre Transcripts. *Language Resources & Evaluation*, 1:40.
- CONSTANT, M. et DISTER, A. (2010). *Spoken Communication*, chapitre Automatic Detection of Disfluencies in Speech Transcriptions, pages 259–272. Cambridge Scholars Publishing.
- DINARELLI, M. et ROSSET, S. (2011). Models Cascade for Tree-Structured Named Entity Detection. In *Proceedings of IJCNLP*, pages 1269–1278.
- GEORGILA, K. (2009). Using Integer Linear Programming for Detecting Speech Disfluencies. In *Proceedings of NAACL HLT*, pages 109–112.
- LAFFERTY, J. D., MCCALLUM, A. et PEREIRA., F. C. N. (2001). Conditional Random Fields : Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of ICML*.
- LAVERGNE, T., CAPPÉ, O. et YVON, F. (2010). Practical Very Large Scale CRFs. In *Proceedings of ACL*, pages 504–513.
- LEVELT, W. J. M. (1983). Monitoring and Self-repair in Speech. *Cognition*, 14:41–104.
- LIU, Y., SHRIBERG, E. E., STOLCKE, A., HILLARD, D., OSTENDORF, M. et HARPER, M. (2006). Enriching Speech Recognition with Automatic Detection of Sentence Boundaries and Disfluencies. *IEEE Transactions on Audio, Speech, and Language Processing*, 14:1526 – 1540.
- MAKHOUL, J., KUBALA, F., SCHWARTZ, R. et WEISCHEDEL, R. (1999). Performance Measures For Information Extraction. In *Proceedings of DARPA Broadcast News Workshop*, pages 249–252.
- QIAN, X. et LIU, Y. (2013). Disfluency Detection Using Multi-step Stacked Learning. In *Proceedings of NAACL HLT*, pages 820–825.
- RASOOLI, M. S. et TETREULT, J. (2013). Joint Parsing and Disfluency Detection in Linear Time. In *Proceedings of EMNLP*, pages 124–129.
- SHRIBERG, E. E. (1994). *Preliminaries to a Theory of Speech Disfluencies*. Thèse de doctorat, Berkeley University of California.
- STRASSEL, S. (2004). *Simple Metadata Annotation Specification*. Linguistic Data Consortium.
- VASILESCU, I., ROSSET, S. et ADDA-DECKER, M. (2010). On the Functions of the Vocalic Hesitation euh in Interactive Man-machine Question Answering Dialogs in French. In *Proceedings of DiSS/LPSS*.