

Traitement automatique de disfluences dans un corpus linguistiquement contraint

Jean-Leon Bouraoui (1), Nadine Vigouroux (2)

(1)IRIT – Paul Sabatier, 118, route de Narbonne, 31062 Toulouse, France
{bouraoui,vigourou}@irit.fr

Résumé Cet article présente un travail de modélisation et de détection des phénomènes de disfluence. Une des spécificité de ce travail est le cadre dans lequel il se situe: le contrôle de la navigation aérienne. Nous montrons ce que ce cadre particulier implique certains choix concernant la modélisation et l'implémentation. Ainsi, nous constatons que la modélisation fondée sur la syntaxe, souvent utilisée dans le traitement des langues naturelles, n'est pas la plus appropriée ici. Nous expliquons la façon dont l'implémentation a été réalisée. Dans une dernière partie, nous présentons la validation de ce dispositif, effectuée sur 400 énoncés.

Abstract This article presents a work of modeling and detection of phenomena disfluences. One of the specificity of this work is its framework: the air traffic control. We show that this particular framework implies certain choices about modeling and implementation. Thus, we find that modeling based on the syntax, often used in natural language processing, is not the most appropriate here. We explain how the implementation has been completed. In a final section, we present the validation of this device, made of 400 utterances.

Mots-clés : Dialogue oral spontané, Analyse linguistique de corpus, Compréhension robuste, Contrôle Aérien, Phraséologie, Disfluences, Modèles de langage, Traitement Automatique du Langage Naturel

Keywords: Spontaneous speech dialog, corpus linguistic analysis, robust understanding, Air Traffic Control, phraseology, disfluencies, language models, Natural Language Processing

1 Introduction

Les disfluences sont un phénomène apparaissant fréquemment dans toute production orale spontanée. Elles ont donné lieu à de nombreuses études, que ce soit dans le domaine du Traitement Automatique de la Parole, ou celui du Traitement Automatique du Langage Naturel. En effet, leur étude et leur identification précise est primordiale. Sur le plan théorique, pour mieux comprendre et modéliser les problèmes pouvant advenir lors de toute communication orale. Sur le plan applicatif, pour améliorer la robustesse des systèmes automatiques de dialogue oral.

Or, la prise en compte des disfluences dans le cadre du TALN est particulièrement difficile. En effet, il s'agit de phénomènes se manifestant, en apparence, de manière totalement désordonnée ; de plus leur apparition provoque généralement des perturbations dans la régularité de la production orale. Malgré tout, leur traitement automatisé fait l'objet de plusieurs travaux, cependant encore assez rares.

Cet article présente lui aussi un travail de ce type. Une de ses spécificités est qu'il est mené dans le cadre d'une tâche particulièrement contrainte, notamment d'un point de vue linguistique, par l'utilisation d'une phraséologie spécifique : le Contrôle de Trafic Aérien. Comme cet article le montre, cela implique un certain nombre de choix théoriques et pratiques.

La présentation de notre travail se fait en quatre temps. D'abord, la description du corpus et de ses caractéristiques. Ensuite, la présentation des différents types de disfluences relevées dans le corpus¹, et la manière dont ils ont été modélisés. Enfin, nous donnons les principes de l'implémentation obtenue, avant de présenter la validation que nous avons mis en œuvre, ainsi que les résultats obtenus.

2 Présentation du corpus d'étude

2.1 Les communications entre contrôleurs et pseudo-pilotes

Le corpus utilisé notre étude est constitué d'enregistrements de dialogues oraux spontanés entre des contrôleurs aériens en formation et des « pseudos-pilotes ». Ces derniers sont des instructeurs simulant le rôle de pilotes en exercice. Deux langues sont utilisées : le français (majoritaire) et l'anglais. Le but des exercices enregistrés est d'entraîner les apprentis contrôleurs, et ensuite de les évaluer. Il s'agit pour eux de gérer plusieurs avions situés dans la zone contrôlée, par exemple en leur assignant une vitesse et/ou une position données. Le canal audio ne peut être « occupé » que par un seul locuteur à la fois, ce qui empêche tout recouvrement de parole.

Les productions orales des contrôleurs et des pilotes sont gouvernées par une stricte phraséologie¹. Celle-ci décrit, par exemple, la manière dont les locuteurs doivent prononcer les identifiants des avions, ou bien l'ordre que doivent observer les différentes parties d'un

¹ Décrite dans l'arrêté du 27 juin 2000 relatif aux procédures de radiotéléphonie à l'usage de la circulation aérienne générale (J.O n° 171 du 26 juillet 2000, p. 11501)

message. Durant la formation, ainsi d'ailleurs que dans des conditions réelles de travail, la phraséologie n'est pas toujours systématiquement appliquée. Le cadre général qu'elle fixe est cependant respecté.

Il est également important de noter que les dialogues enregistrés appartiennent bien à la catégorie du discours oral spontané. Nous tenons à le préciser car le rôle prépondérant tenu par la phraséologie pourrait laisser à penser que tous les énoncés produits sont déjà planifiés à l'avance. Or, ce n'est pas le cas : ni les contrôleurs, ni les pilotes ne savent à l'avance ce qu'ils vont avoir à dire. La phraséologie définit seulement le cadre général de production des énoncés ; ce qui est dit repose sur l'interaction dynamique entre un contrôleur et un pilote ou pseudo-pilote donnés, en fonction d'une situation variable.

2.2 Caractéristiques du corpus d'étude

Nous avons procédé à la transcription et l'annotation des dialogues selon les spécifications de (Coullon *et al.*, 2000).

Ces spécifications ont pour but de déterminer les éléments à transcrire, d'obtenir l'homogénéité des transcriptions dans le cas où plusieurs annotateurs se succèdent.

Les enregistrements ont été effectués avec un DAT (Digital Audio Tape), et échantillonnés à 16 kHz (16 bits). Pour des raisons d'enregistrement, la qualité sonore souffre parfois de problèmes résultants de la saturation ou de bruits tels que les interférences ; cependant, les dialogues sont intelligibles. Le corpus dure 36h50mn, et comprend 76 306 mots ; 16 locuteurs différents ont participé à sa production.

3 Modélisation proposée

Pour prendre en compte les disfluences, nous avons procédé en deux grandes étapes. Dans un premier temps, il a fallu identifier précisément les phénomènes à traiter. Une fois cela fait, nous avons ensuite pu procéder à la modélisation de leurs patterns d'apparition. Ce processus est décrit dans les deux sous sections suivantes.

3.1 Identification des phénomènes de disfluences

La tâche n'est pas aussi triviale qu'il n'y paraît, surtout lorsqu'il s'agit des disfluences. En effet, dans la littérature consacrée à ce phénomène, les termes utilisés pour désigner un phénomène donné varient souvent d'un auteur à l'autre (cf. par exemple (Candéa, 2000), (Henry *et al.* 2004), (Kurdi, 2003) ou encore (Schriberg, 1994)).

Nous avons défini 7 catégories de disfluences. Lorsque cela est nécessaire, nous donnons des exemples pour illustrer notre propos (en mettant la disfluence en gras). Pour les comparaisons avec d'autres terminologies, ainsi que les distributions de chaque catégorie, nous renvoyons à (Bouraoui *et al.*, 2006).

- Hésitation : désigne l'interjection « euh ». Exemple:

maintenons niveau 1 0 0 Poitiers Amboise euh Lacan

- Répétition : un mot (ou un groupe de mots) apparaît au moins deux fois à la suite. Nous n'avons pas pris en compte la répétition de disfluences. Exemple:

*station **station** calling euh repeat your callsign*

- Amorce : l'arrêt de la production d'un mot avant la fin normale de celui-ci. Dans notre terminologie, une amorce correspond toujours à un fragment de mot que l'on peut identifier (grâce à la connaissance de la phraséologie). Exemple (l'amorce est entre crochets) :

*speed euh 200 Kts [**mak**] euh minimum*

Le contexte et la phraséologie aident à comprendre que le contrôleur commence à prononcer « maximum ». Il se rend compte que cela ne convient pas et s'interrompt (« mak »). Enfin, il dit le mot correct : « minimum ».

- Auto-correction : le locuteur fait une ou plusieurs erreurs, et se corrige dans le même énoncé. La différence avec les amorces est que le mot (ou les mots) erroné est prononcé complètement.

*France Air Force **900** 893*

- Fragment de mot : un ou plusieurs phonèmes inidentifiables (par opposition aux amorces). Exemple (le fragment est entre crochets) :

*due to [**ou**] due traffic euh descend level 9 0*

- Allongement : l'allongement d'une unité phonétique d'un mot, supérieur à 0,5 seconde ; peut être combiné aux hésitations.
- Pause longue : toute pause supérieure à 0,5 seconde et comprise à l'intérieur un tour de parole.

3.2 Modélisation des phénomènes

Dans la majorité des études sur les disfluences, la modélisation, lorsqu'elle est faite, repose sur la représentation syntaxique et/ou sémantique des phénomènes. Or, nous considérons que, compte tenu de la nature linguistique très particulière induite par la phraséologie et ses restrictions, il est inutile, voire contre-productif, de nous lancer dans une modélisation basée sur les relations lexico-syntaxiques. En effet, les contraintes posées par la phraséologie permettent en grande partie de décrire le système linguistique formé par celle-ci, indépendamment ou presque des « parties du discours » auxquelles appartiennent les éléments.

Plus concrètement, nous cherchons un langage de représentation qui présente au moins les caractéristiques suivantes : adapté à l'objet d'étude (disfluences), grande capacité d'expression et de représentation, simplicité des patrons obtenus, et enfin modularisation selon les besoins de l'utilisateur. Nous avons trouvé réunies ces propriétés dans le schéma d'annotation proposé par (Bear *et al.* 1993), utilisé dans un contexte relativement identique au nôtre par (Kurdi, 2003, notamment p.72-75).

Le but de ce schéma est de pouvoir rendre compte de certains phénomènes de l'oral spontané, par exemple dans des transcriptions, en les étiquetant. Les auteurs désignent les phénomènes en question par le terme générique de repairs. Ce terme correspond, dans les grandes lignes, à ce que nous avons appelé répétitions, amorces, et auto-corrections. Mais le schéma peut également rendre compte d'autres disfluences.

Les auteurs proposent tout un ensemble de symboles génériques permettant de représenter la majorité des disfluences. Par exemple, FP (Filled Pauses) pour les hésitations, C (Cues) pour des mots tels que « désolé », etc. Chaque occurrence d'un symbole est associée à un index numérique qui permet le cas échéant de la distinguer dans un même énoncé.

Le schéma permet également de rendre compte des différentes zones impliquées dans les disfluences plus complexes telles que les répétitions ou les auto-corrections. Pour cela, les auteurs utilisent la barre verticale (« | ») pour indiquer la position du point d'édition. Il s'agit du « lieu » suivant l'apparition de la disfluence, à partir duquel le locuteur se rend compte qu'il y a un problème dans sa production orale, et tente d'y remédier. Soit par exemple l'énoncé suivant, comportant une auto-correction (en gras) :

*montez niveau **maintenez** niveau*

Sa représentation selon le schéma de Bear *et al.* sera :

R1 M1 | R1 M1

où le symbole R correspond au remplacement d'un mot par un autre de part et d'autre du point d'édition (respectivement, l'erreur et la correction) et M (Matching Words) indique que le même mot est utilisé des côtés du point d'édition.

A l'aide de ce schéma d'annotation, que nous avons augmenté de certains symboles plus spécifiques à la tâche de contrôle aérien, nous avons systématiquement modélisé toutes les occurrences de disfluences de notre corpus. A partir des représentations obtenues nous avons dégagé des patrons plus généraux, qui nous ont été utiles pour la dernière partie de ce travail : l'implémentation.

4 Implémentation

4.1 Cadre d'implémentation

Le système dans lequel nous intégrons notre modèle a été développé dans le cadre du projet VOICE, et est décrit dans (Truillet *et al.*, 2005).

Plus précisément, le but de ce système est de gérer les entrées vocales pouvant être produites par les pseudo-pilotes, en extraire le sens, et l'interpréter de manière à ce que les ordres reçus soient exécutés. Il a déjà fonctionné dans des conditions de simulation du contrôle aérien, même si un certain nombre de contraintes restent à prendre en compte, dont la gestion des disfluences !

Le système VOICE dispose d'une architecture modulaire de composants logiciels communiquant entre eux. Ces derniers sont dédiés chacun à une sous-tâche. Notamment, l'un d'eux gère la reconnaissance de la parole. A partir des résultats fournis par celui-ci, un autre

choisit la solution la plus probable, en fonction du contexte du dialogue (historique, connaissances de la tâche, etc.). Enfin, la décision prise est ajoutée à l'historique. Nous renvoyons à (Truillet *et al.*, 2005) pour une description détaillée des autres composants, et du rôle de chacun.

Le système VOICE fonctionne à partir d'une grammaire de type *CFG* (*Context Free Grammar*), modifiée périodiquement en fonction du contexte d'interaction. Elle se présente sous la forme de règles de réécriture, non probabilisées. Ce formalisme est particulièrement bien adapté à la phraséologie du contrôle aérien, compte tenu de la standardisation des productions et du faible nombre de schémas lexico-syntaxiques utilisés². Par contre, ces règles ne prenaient pas encore en compte la modélisation des disfluences que nous avons décrit ci-dessus; de plus, certaines disfluences complexes ne peuvent être traitées uniquement par des règles de CFG. De ce fait, un traitement supplémentaire est nécessaire, que nous décrivons dans la section suivante.

4.2 Modélisation computationnelle et traitement informatique

Nous distinguons deux types de phénomènes à traiter. D'une part les disfluences « ponctuelles ». Nous rangeons dans cette catégorie les hésitations, et les fragments de mots. Nous les appelons « ponctuelles » car elles ne remettent pas en cause, à elles seules, l'organisation de la suite de mots dans laquelle elles apparaissent³. Nous considérons donc que seules l'identification de leurs occurrences est nécessaire et suffisante pour un traitement automatique (quelle que soit l'utilisation faite par la suite de leur identification).

Ce n'est pas la même chose avec les disfluences plus complexes que sont les amorces, les répétitions, et les auto-corrections. Elles affectent en effet un empan plus important en termes de taille que la seule position dans laquelle elles apparaissent. Les concernant, il faut non seulement identifier leurs occurrences, mais également distinguer les éléments répétés (dans le cas des amorces, distinguer l'élément corrigé de la correction).

Que nous apprend l'analyse des différents patrons de ces disfluences ? Dans la majorité des cas, nous constatons une symétrie des composants situés de part et d'autre du point d'édition. Cette observation va nous servir de point d'achoppement pour la mise en œuvre pour l'implémentation. Nous ne décrivons pas ici la totalité des algorithmes, mais les principes généraux qui les sous-tendent et dont ils ne sont que le développement :

- Dans un premier temps, associer à chaque élément lexical de l'énoncé à traiter une classe de mot correspondant. Cela est assuré par la CFG utilisée dans VOICE, moyennant quelques modifications (par exemple la définition de classes de mot, afin d'ajouter un niveau « sémantique ». Ainsi, au verbe « descendez » serait associé la classe « verbe_niveau »). C'est à ce niveau que se situe le traitement des disfluences « ponctuelles ». Si on a affaire à une disfluence plus complexe, un traitement plus approfondi s'impose ;

² Pour plus de détails sur les caractéristiques linguistiques de la phraséologie, voir notamment (Falzon, 1989).

³ On pourrait objecter ici que l'apparition de ce type de disfluences est souvent concomitante avec des perturbations plus larges du contexte où elles apparaissent. Cependant, ces relations sont-elles de cause ou de conséquence ? La question reste à trancher... En attendant, nous adoptons comme hypothèse de travail l'indépendance de causalité des phénomènes entre eux.

- « Parser » l'énoncé avec une fenêtre coulissante. Celle-ci, d'une taille maximale de 10 éléments lexicaux, se déplace de gauche à droite de l'énoncé. Le principe est de détecter, dans l'intervalle de la fenêtre, l'occurrence d'une répétition d'un patron (classe de mots ou mot du lexique, correspondant au schéma de). Lorsque cela arrive, le déplacement de la fenêtre stoppe, le temps de procéder à plusieurs traitements à l'intérieur de celle-ci :
 1. T
trouver la position du point d'édition. Pour cela, on lance une analyse par règle de CFG à partir du début de la fenêtre, sur l'ensemble du contenu de celle-ci, comme s'il s'agissait d'un énoncé à part entière. Lorsque l'analyse se termine avant que l'un des éléments terminaux de la CFG n'ait été atteint, cela indique avec certitude la présence du point d'édition
 2. L
le point d'édition trouvé, nous avons à distinguer les éléments répétés entre eux (ou les éléments corrigés de leur correction). On fait la comparaison entre les éléments situés à gauche du point d'édition et ceux de droite ; les éléments répétés (d'un point de vue lexical) correspondent... à des répétitions, tandis que les éléments répétés (au niveau de la classe de mot) correspondent respectivement à une erreur et à sa correction.
- Enfin, le déplacement de la fenêtre est réamorcé, à partir de l'élément situé le plus à droite lors de sa dernière instance.

Nous avons donné les grandes lignes du traitement informatique proposé. Ne figurent pas par exemple, les processus nécessaires pour distinguer, par exemple, une répétition disfluente d'une répétition « normale », ainsi que d'autres problèmes de ce genre.

5 Evaluation

5.1 Méthodologie d'évaluation

Nous retenons les paramètres suivants comme pertinents pour la validation du module que nous proposons :

- Temps d'exécution : ce paramètre permet de déterminer si le dispositif que nous proposons est raisonnablement intégrable dans un système visant à fonctionner en temps réel. En effet, au-delà d'un certain seuil, la question du rapport entre le gain apporté par le dispositif et la perte qu'il occasionne en temps de calcul se poserait ;
- Taux de précision : il permet d'établir le nombre de détections correctes par rapport au nombre total de détections (y compris les « faux positifs ») ;
- Taux de rappel : il détermine le nombre de disfluences correctement détectées par rapport au nombre total d'occurrences ;
- Délimitation : la délimitation des frontières de début et de fin des disfluences est une des tâches accomplies par notre module. Il est par conséquent important de

déterminer les performances dans l'accomplissement de celle-ci. Pour cette mesure également, nous utilisons les taux de précision (ici, nombre de délimitations correctes par rapport au nombre total de délimitations effectuées) et de rappel (nombre de délimitations correctes par rapport au nombre total de disfluences) ;

- Identification de la bonne catégorie : il s'agit de déterminer si une disfluence, même détectée et délimitée correctement, est également correctement catégorisée. Il peut, par exemple, arriver qu'une répétition soit étiquetée comme étant une auto-correction, ou réciproquement. Ce risque est d'autant plus important que notre algorithme de détection est basé sur l'occurrence de répétitions, que ce soit au niveau lexical ou sémantique. La performance pour ce paramètre est importante : détecter et étiqueter correctement un lexème donné de la zone disfluente est le premier pas pour pouvoir ensuite corriger celle-ci (par exemple, supprimer les répétitions surnuméraires, ou remplacer un mot corrigé par sa correction).

5.2 Constitution du corpus d'évaluation

Pour l'évaluation, il nous a été nécessaire de nous procurer un deuxième corpus, indépendant de celui, présenté plus haut, à partir duquel a été effectuée la modélisation. Il est très difficile de se procurer des corpus oraux de dialogues de contrôle aérien, pour des raisons liées au caractère sensible de ce travail. De ce fait, nous avons dû produire « artificiellement » un corpus de disfluences apparaissant dans ce contexte. Pour cela, nous nous sommes inspirée de la campagne « par défi » du GDR I3⁴ (Antoine *et al.* 2002). Elle consiste à sélectionner un ensemble d'énoncés, considérés comme représentatifs du domaine de tâche étudié (Antoine *et al.* 2002, p. 563). A partir de ces énoncés de base, des énoncés dérivés sont construits. Le processus de dérivation consiste à transformer chaque énoncé initial de manière à ce qu'il contienne un phénomène pouvant poser problème à un système de compréhension donné.

Nous avons été contraints d'opérer quelques modifications à cette méthodologie afin de l'adapter à notre contexte d'étude. Dans un premier temps, nous avons sélectionné un ensemble d'énoncés de notre corpus d'étude, choisis parce qu'ils ne comportent aucune disfluence. A partir de chacun de ces énoncés, nous avons construit manuellement 6 énoncés comportant chacun une disfluence (hésitation, répétition, ou auto-correction). Nous avons généré de cette manière 306 énoncés. Nous avons également ajouté à ce corpus principal deux autres sous-ensembles. D'une part, les 51 énoncés de base ne comportant aucune disfluence. Il nous a en effet paru intéressant de les faire traiter par notre module, afin de déterminer si celui-ci ne provoque pas de « faux positifs ». D'autre part, 43 énoncés, comportant au moins une disfluence, provenant de transcriptions de dialogues de contrôle aérien enregistrés en situation « réelle » (et non en situation d'apprentissage, contrairement à notre corpus d'étude). Au total, notre corpus de validation comporte donc 400 énoncés différents, répartis en quatre sous-ensembles différents (chaque sous-ensemble correspondant à la vérification d'une ou plusieurs hypothèses)⁵. Les énoncés sont présentés au module sous forme écrite (et non orale) : on évite ainsi tout biais lié à des problèmes de reconnaissance vocale, pour se concentrer sur la validation des capacités d'identification de notre module.

⁴ Groupement De Recherche « Information - Interaction - Intelligence » ; <http://www.irit.fr/GDR-I3/>.

⁵ Faute de place, nous ne pouvons décrire ici en détail les hypothèses en question. Nous renvoyons pour cela à (Bouraoui, 2008).

5.3 Résultats

Par manque de place, nous ne pouvons présenter ici l'intégralité des résultats que nous avons obtenus. Nous n'indiquons donc que les plus représentatifs, et renvoyons à nouveau à (Bouraoui, 2008) pour plus de détails.

Globalement, les résultats obtenus sont satisfaisants et valident le module. Ainsi, le temps d'exécution est très rapide: le temps maximum d'exécution est inférieur à 5 millisecondes. En termes de détection des disfluences, les performances sont elles aussi bonnes voire très bonnes selon le sous-ensemble considéré du corpus : le taux de précision varie en effet entre 84% et 95%, et le taux de rappel entre 56% et 93%. Concernant ce dernier résultat, le taux de 56% observé correspond à une difficulté du module pour la tâche de délimitation des auto-corrections, sur des disfluences particulièrement complexes.

Enfin, la comparaison avec les performances d'autres dispositifs de détection des disfluences, décrits dans la littérature, permet de déterminer que nos résultats se situent dans la « moyenne supérieure ».

6 Conclusions et perspectives

Nous avons décrit dans cet article le processus complet de traitement automatique que nous proposons pour prendre en compte les disfluences dans des dialogues oraux spontanés de contrôle aérien.

Il a consisté dans un premier temps en une analyse fine du contexte, et des disfluences observées. Nous avons ensuite procédé à l'implémentation du dit modèle, dans le cadre d'un système existant déjà, mais ne permettant pas de traiter ce type de phénomènes. Enfin, nous avons validé le module d'identification des disfluences que nous avons conçu. Il apparaît que notre module obtient de bonnes, voire très bonnes performances, et est robuste. Notre hypothèse sur la spécificité du traitement automatique des disfluences dans la phraséologie est confirmée elle aussi. L'évaluation montre également que le dispositif que nous proposons est très bien adapté, aussi bien à la modélisation qu'au traitement automatique des disfluences dans un langage contraint.

Le traitement proposé n'est pas exhaustif puisque certains cas très complexes ne sont pas pris en compte ; néanmoins, il couvre théoriquement une grande part des disfluences. A plus long terme, nous envisageons d'étudier dans quel mesure le modèle proposé peut être appliqué à d'autres types de tâches que le seul contrôle aérien. Nous sommes en effet persuadés qu'au moins une partie des traitements que nous avons proposés sont applicables également dans d'autres contextes, même moins restreints que celui de la phraséologie.

Références

ANTOINE J-Y., BOUSQUET C., GOULIAN J., ZAKARIA KURDI, M., ROSSET S., VIGOUROUX N., VILLANEAU J., (2002) "Predictive and objective evaluation of speech understanding: the 'challenge' evaluation campaign of the I3 workgroup of the French CNRS", *Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Espagne, Vol. II, p. 529-536, 2002.

- BEAR J., DOWDING J., SHRIBERG E. (1993) "A System for Labeling Self-Repairs in Speech." *Technical Note 522*, Stanford Research International.
- BOURAOUI J.-L., VIGOUROUX N. (2006) « Etude de disfluences dans un corpus linguistiquement contraint ». *Journées d'Etudes sur la Parole (JEP 2006)*, Dinard, p. 429-432.
- BOURAOUI J.-L. (2008) *Analyse, modélisation, et détection automatique des disfluences dans le dialogue oral spontané contraint : le cas du Contrôle Aérien*, Thèse de doctorat, Université Paul Sabatier.
- CANDEA M. (2000) *Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané*. Thèse d'État, Université Paris III (Sorbonne Nouvelle).
- COULLON I., GRAGLIA L. (2000) *Spécifications de la base de données pour l'analyse des communications VHF en route*. CENA internal report.
- FALZON P. (1989) *Ergonomie cognitive du dialogue*. Presses Universitaires de Grenoble.
- HENRY S., Campione E., Véronis J. (2004) « Répétitions et pauses (silencieuses et remplies) en français spontané », *JEP'04*, p.261-264, Fès, Maroc.
- KURDI M.- Z. (2003) *Contribution à l'analyse du langage oral spontané*. Thèse de doctorat, Université J. Fourier, Grenoble, France.
- SHRIBERG E. (1994) *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of Berkeley, California.
- TRUILLET P., BOTHOREL G. (2005) « VOICE : une plate-forme pour la formation au contrôle aérien », *IHM 2005*, p. 67-80, Toulouse.