TIPA. Travaux interdisciplinaires sur la parole et le langage

29 | 2013
Le français parlé

Une étude quantitative des marqueurs discursifs, disfluences et chevauchements de parole dans des interviews politiques

A quantitative study of discourse markers, disfluencies and overlapping speech in political interviews

Philippe Boula de Mareüil, Gilles Adda, Martine Adda-Decker, Claude Barras, Benoît Habert et Patrick Paroubek

Édition électronique
URL : http://journals.openedition.org/tipa/830
DOI : 10.4000/tipa.830
ISSN : 2264-7082

Éditeur
Laboratoire Parole et Langage

Référence électronique
Philippe Boula de Mareüil, Gilles Adda, Martine Adda-Decker, Claude Barras, Benoît Habert et Patrick Paroubek, « Une étude quantitative des marqueurs discursifs, disfluences et chevauchements de parole dans des interviews politiques », *TIPA. Travaux interdisciplinaires sur la parole et le langage* [En ligne], 29 | 2013, mis en ligne le 19 décembre 2013, consulté le 19 avril 2019. URL : http://journals.openedition.org/tipa/830 ; DOI : 10.4000/tipa.830

Ce document a été généré automatiquement le 19 avril 2019.

La revue *TIPA. Travaux interdisciplinaires sur la parole et le langage* est mise à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Une étude quantitative des marqueurs discursifs, disfluences et chevauchements de parole dans des interviews politiques

A quantitative study of discourse markers, disfluencies and overlapping speech in political interviews

Philippe Boula de Mareüil, Gilles Adda, Martine Adda-Decker, Claude Barras, Benoît Habert et Patrick Paroubek

Nous sommes très reconnaissants envers l'INA pour le corpus audiovisuel et sa transcription. Merci aussi à Frédérique Bénard et à Marguerite Lenhardt pour leur aide précieuse à cette recherche.

1. Introduction

1 Le français parlé, comme toute langue parlée, avec son cortège d'hésitations, de lapsus, de faux départs, de répétitions, pose d'importants problèmes à la linguistique et au traitement automatique (Schegloff *et al.*, 1977 ; Levelt, 1989 ; Blanche-Benveniste, 1990 ; Fernandez-Vest, 1994 ; Gadet, 1997 ; Shriberg, 2001 ; Duez, 2001 ; Chanet, 2004 ; Dister, 2007 ; Bertrand *et al.*, 2009 ; Bazillon, 2011). Pour étudier ces phénomènes, pour quantifier la variation diamésique (*i.e.* de médium) entre langue parlée et langue écrite, les corpus « apparentés » où le « même » contenu est exprimé à la fois à l'oral et sous forme textuelle sont des ressources privilégiées. Ce type de documents, avec des données de parole et leur transcription plus ou moins fidèle ou débarrassée des spécificités de l'oral, se trouve aujourd'hui relativement facilement (débats parlementaires, archives d'émissions de télévision...).

2 Dans l'étude rapportée dans cet article, nous sommes partis d'interviews télévisées et nous avons utilisé les transcriptions *bona fide* réalisées à destination de la presse afin de

fournir une transcription *verbatim* plus fidèle à l'original. Les transcriptions *bona fide* ne sont pas des transcriptions au mot près de ce qui a été dit. En particulier, les hésitations, reformulations et amorces de mots abandonnés sont (presque) systématiquement gommées. Néanmoins, ces transcriptions restent suffisamment proches de ce qui a été dit pour pouvoir être utilisées à des fins de citation.

3 Alors qu'un laps de temps sépare la lecture de l'écriture, la simultanéité de la production et de la perception dans la communication parlée rend patents les bredouillages, bafouillages et cafouillages, achoppements et piétinements dans le flux de parole. La chaîne écrite traditionnelle comme la chaîne parlée est linéaire, mais elle permet dans sa phase de conception (traçable dans le brouillon) des ratures et des effacements normalement impossibles à l'oral. L'écrit a vocation à être persistant, contrairement aux mots parlés qui jaillissent de façon éphémère et s'envolent — à moins bien sûr d'être enregistrés. Il n'est pas astreint aux mêmes contraintes physiologiques que la phonation, qui utilise les mêmes organes que la respiration : beaucoup de fautes de frappe, au clavier, n'ont aucune base phonétique ou phonologique. On peut distinguer l'écriture (processus dynamique), l'écrit (validé par le scripteur) et la lecture (étape de décodage), alors que dans la conversation informelle de tous les jours, typiquement face à face avec des tours de parole dont l'ordre n'est pas préétabli, l'idéation (la recherche des mots, le travail de formulation), le fait de parler et d'écouter sont synchrones. D'où également des chevauchements de parole délicats à linéariser. La transcription de l'oral implique des choix théoriques (voire idéologiques) qui sont devenus un objet de recherche linguistique en soi (Blanche-Benveniste, 1990 ; Morel & Danon-Boileau, 1998 ; Delais-Roussarie & Durand, 2003). Elle est toujours un début d'analyse, ne serait-ce qu'à travers l'espace entre les mots (Catach, 1992), puisqu'elle discrétise des phénomènes qui peuvent présenter des ambiguïtés (ex. *on (n')a pas*). Elle est le résultat d'un long travail de standardisation qui peut faire paraître l'écrit pour plus stable, plus homogène que l'oral. Elle suppose des conventions et des jugements subjectifs qui dépendent des domaines linguistiques étudiés. La ponctuation, en particulier, rend plus lisible les transcriptions et facilite le traitement automatique, mais n'est qu'un équivalent approximatif de la prosodie (Catach, 1994 ; Dister *et al.*, à paraître).

4 Il a été montré, dans Adda-Decker *et al.* (2003) ainsi que dans Barras *et al.* (2004), qu'il est relativement simple d'aligner automatiquement les transcriptions automatiques d'un corpus de parole avec des transcriptions *bona fide* en s'appuyant sur les îlots de confiance constitués de sous-séquences similaires.[1] Nous pouvons ainsi fournir aux transcripteurs humains, à un coût relativement faible, une transcription approximative où les portions douteuses sont mises en évidence : les transcripteurs peuvent ainsi concentrer leurs efforts sur les parties de la transcription qui nécessitent le plus leur attention.

5 À partir d'un corpus d'une dizaine d'émissions télévisées, nous avons, dans un premier temps, porté notre attention plus particulièrement sur le traitement des marqueurs discursifs — ces « "petits mots" qui habitent l'oral » (Morel & Danon-Boileau, 1998) — et des phénomènes d'hésitations, répétitions et faux départs, que le terme de « disfluence » se propose de recouvrir. Ce terme que nous préférons à celui de *scories* (Gadet, 1997) évoque toujours, malheureusement, quelque chose de gênant en soi — toutefois, c'est la graphie *dysfluence* qui est réservée aux pathologies du langage. Les « marqueurs discursifs », quant à eux, recouvrent ce que la littérature pragmatique — ou sa variante francophone, la théorie de l'énonciation — appelle « connecteurs » (ex. *mais, alors, donc*) ainsi que les « ponctuants » comme *ben, voilà* et ce que l'analyse conversationnelle nomme

les « phatiques » (ex. *tu sais/vois, écoutez, hein*), mais exclut les « régulateurs » du type *hmm* (Fernandez-Vest, 1994 ; Chanet, 2004). Cependant, il n'est pas toujours évident de distinguer nettement entre ces catégories (Dister, 2007). Et les mots cités en exemples ci-dessus ne sont pas systématiquement des marqueurs discursifs. Le terme *bon*, par exemple, peut être une interjection servant à marquer une certaine transition dans le discours (Tanguy, 2009), mais il peut aussi être un adjectif remplissant une véritable fonction syntaxique. De même que les disfluences, les marqueurs discursifs ont fait l'objet d'un important travail d'annotation manuelle, qui est relaté dans cette étude.

6 Nous nous sommes concentrés dans un second temps sur les chevauchements de parole : la transcription, l'annotation ainsi que les interactions avec les disfluences (répétitions, faux départs, hésitations) et les marqueurs discursifs. Ces chevauchements de parole, impliquant plus d'un locuteur parlant simultanément, constituent de sérieux problèmes pour la transcription de dialogues (Blanche-Benveniste, 1990 ; Delais-Roussarie & Durand, 2003 ; Dister *et al*., à paraître) et sont autant d'écueils pour la reconnaissance de la parole, qui les négligeait encore récemment. Ils sont pourtant naturels dans la communication orale et essentiels pour l'analyse conversationnelle (Schegloff, 2000).

7 Même des interviews télédiffusées, matériel relativement contrôlé sur lequel nous travaillons, produisent des chevauchements — en moins grand nombre, certes, que dans des conversations téléphoniques ou des réunions. Une interview est un type d'interaction asymétrique entre des locuteurs de statuts différents, ayant des rôles complémentaires. Elle diffère également de la conversation entre pairs sur l'axe compétitif/coopératif (Grice, 1975 ; Schegloff, 2000). La symétrie et la compétitivité dans l'échange accroissent le degré d'interactivité, et partant les alternances de tours de parole ainsi que le taux de chevauchements. La coopérativité peut consister à aider l'interlocuteur à parler, ce qui d'habitude n'est pas nécessaire dans des interviews politiques. Dans les interviews politiques, les journalistes ont souvent un point de vue à défendre : ils s'inscrivent dans une double énonciation dans laquelle ils s'adressent également à un public. Ils peuvent contredire et interrompre les invités, favorisant par là même les chevauchements. Le but de l'étude de ce genre de matériau est d'arriver à prendre en compte des informations sur la synchronisation et la régulation des tours de parole en interaction. La question est alors : comment distinguer différents types de chevauchements et quelle est leur influence sur les disfluences ?

8 La section suivante (section 2) présente le corpus et la méthode suivie, avec les directives de transcription et le schéma d'annotation. Nous développons ensuite l'analyse des disfluences et marqueurs discursifs (section 3) et des chevauchements de parole (section 4). Enfin, la section 5 résume les principaux résultats et ouvre quelques perspectives.

2. Corpus, transcription et annotation

2.1. Corpus

9 Nos études reposent sur un corpus d'interviews politiques : 9 émissions de *L'heure de vérité* enregistrées de 1992 à 1994 et fournies par l'Institut National de l'Audiovisuel (INA). Pendant chaque émission, un homme politique ou une personnalité publique est interviewé par plusieurs journalistes. Ces derniers ont préparé leurs questions, et les réponses ne sont pas entièrement spontanées : on a plutôt affaire à un discours surveillé

— *careful speech* (Labov, 1976). En revanche, un modérateur, qui dirige les débats et veille au temps consacré aux différents thèmes, peut souvent interrompre la personne interviewée ou poser des questions. Cette configuration favorise donc les chevauchements de parole. Une partie seulement des nombreuses disfluences fournit de l'information quant au travail de formulation du locuteur ; une autre partie correspond à une lutte pour la parole entre interlocuteurs, même si les journalistes ne « coupent » pas la parole n'importe quand.

10 Pour chaque émission, nous disposons à la fois de l'audio et de la transcription *bona fide* destinée à la presse. Sans reproduire les paroles exactes, cette transcription se veut assez proche de l'audio, tout en suivant des conventions implicites. De fait, la plupart des disfluences ont été ignorées. Aussi une transcription « exacte » (*verbatim*) a-t-elle été produite manuellement, avec tous les événements audibles contenant notamment les phénomènes de disfluences. Pour ce faire, une transcription automatique a été générée, à l'aide d'un système de reconnaissance de la parole : le système du LIMSI (Gauvain *et al.*, 2005). Nous avons d'abord aligné les transcriptions automatiques et *bona fide* avec le signal de parole. Puis le corpus a été corrigé à la main (88 000 mots) pour fournir la transcription *verbatim,* avec une version modifiée du logiciel TRANSCRIBER (Barras *et al.*, 2001). Ce travail a été facilité par la sélection automatisée de zones de discordance entre transcriptions automatique et *bona fide*, zones qui contiennent *a priori* des disfluences non reproduites dans la transcription *bona fide* (Adda-Decker *et al.*, 2003 ; Barras *et al.*, 2004). Il en résulte une représentation écrite de l'oral qui est un compromis entre fidélité et lisibilité (en orthographe standard, sans trucages tels que *y'a* ou *ch'sais pas*), alignée avec les données de parole (*cf.* figure 1).

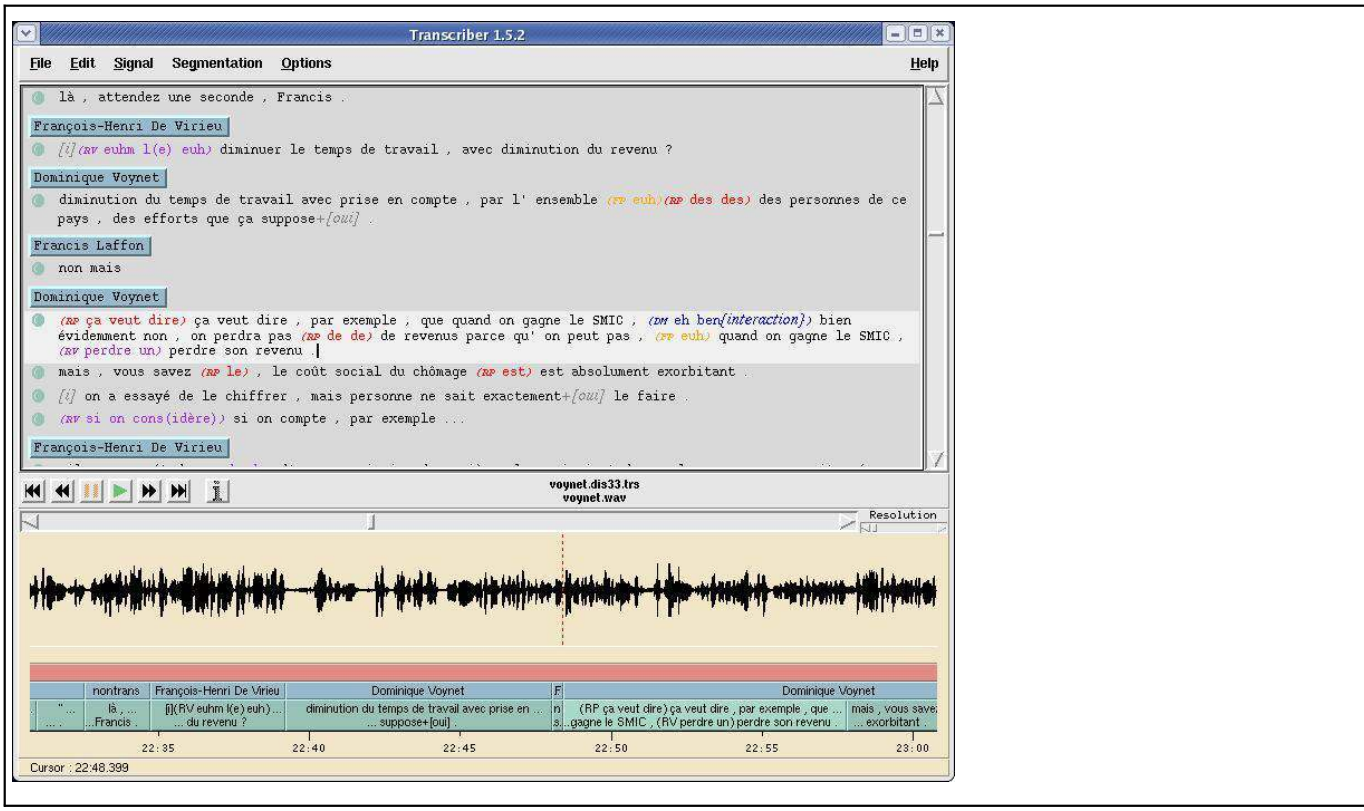
2.2. Annotation des disfluences et marqueurs discursifs

11 Dans ce travail émargeant aux deux champs que sont la linguistique de corpus et le traitement automatique, nous avons globalement suivi les consignes d'annotation développées par le Linguistic Data Consortium (LDC[2]). Nous avons annoté les marqueurs discursifs (DM pour D*iscourse Markers*) et les phénomènes de disfluences, que nous avons classés en trois types principaux : FP (*Filled Pauses*) pour les hésitations du type *euh*, RP (*RePetitions*) pour les répétitions et RV (*ReVisions*) pour les faux départs.

12 Les hésitations, transcrites *euh*, ont été annotées automatiquement. Les répétitions couvrent des mots (éventuellement tronqués ou interrompus par un autre locuteur), des emphases renforçant un propos, ainsi que des reprises à distance après incises. Les faux départs impliquent des fragments de mots, des mots ou groupes de mots qui sont abandonnés, sans nécessairement être corrigés. Les phrases inachevées, cependant, n'entrent pas dans cette catégorie. Quant aux marqueurs discursifs, ils peuvent avoir un simple rôle de remplisseur (une sorte de respiration au sein du flux de parole) ou une vraie fonction de structuration du discours (Auchlin, 1981 ; Fernandez-Vest, 1994 ; Chanet, 2004). Suivant la terminologie de l'école genevoise d'analyse du discours (Roulet *et al.*, 1991), ils peuvent être consécutifs (ex. *alors, donc*), contre-argumentatifs (ex. *mais*) ou réévaluatifs (ex. *enfin*). Mais il n'est pas toujours aisé d'interpréter leur rôle précis, puisqu'on ne connaît pas les intentions exactes des intervenants. La conjonction *et*, en particulier, peut être utilisée pour structurer le dialogue, pour commencer un tour de parole, pour éviter un *euh* stigmatisé, pour lier deux énoncés ou pour éviter d'être interrompu. Il en va de même des expressions comme *je crois que*, qui peuvent être de

simples tics de langage pour certains locuteurs, et qui sont difficiles à annoter de façon cohérente. Souvent, de telles expressions de réserve ou d'atténuation sont là pour adoucir le message, le rendre moins abrupt, le relativiser, temporiser et laisser la place à des opinions autres. En dernier lieu, des cas complexes existent, où de faux départs peuvent inclure des marqueurs discursifs, qui à leur tour peuvent inclure des répétitions et des faux départs. Ils ont également été prévus dans l'annotation, même s'ils sont rares.

Figure 1. Vue d'écran du logiciel TRANSCRIBER avec un extrait de transcription *verbatim* vérifiée manuellement et complétée d'annotations pour les événements propres à l'oral spontané



2.3. Transcription et annotation de la parole superposée

13 Sur 8 émissions de *L'heure de vérité,* une segmentation manuelle en portions de parole superposée a été faite en relâchant les contraintes d'ancrage temporel. Même la localisation précise du début et de la fin des chevauchements, en effet, n'est pas triviale. Le logiciel TRANSCRIBER a été adapté en conséquence pour faciliter cette tâche. Deux situations ont été distinguées : (1) le chevauchement ne conduit pas à un changement de locuteur primaire, qui reste le même au terme du chevauchement ; (2) le locuteur primaire s'arrête et le locuteur secondaire devient le locuteur primaire d'un nouveau tour de parole. Ces deux situations sont présentées dans la figure 2.

Figure 2. Exemples de segmentation et de transcription orthographique dans deux situations

Alain Duhamel
O /vous avez/ dû souvent souffrir, alors ?

Rony Brauman (1)
O non, non, non, /au contraire/ /*Duhamel :* non/ non non, je me suis vraiment épanoui, j'adore ce que je fais, je l'ai toujours fait avec passion et j'ai pas vraiment souffert.
O écoutez il y a des /moments durs vous pouvez pas.../
Alain Duhamel (2)
O /alors c'était quoi ?/ la volonté de transformation ? déjà la lutte, ou la /lutte contre/ /*Brauman :* non/ l'injustice ?
O c'était quoi ?

(1) le chevauchement ne conduit pas à un changement de locuteur primaire, qui reste le même au terme du chevauchement ; (2) le locuteur primaire s'arrête et le locuteur secondaire devient le locuteur primaire d'un nouveau tour de parole.

14 Les chevauchements de parole ont également été annotés en utilisant quatre étiquettes : ***bck*** (***b****a****ck****channel*) pour les régulateurs du type *hmm*, ***tst*** (***t****urn* ***st****ealing*) pour les prises de parole, ***att*** (***a****nticipated* ***t****urn* ***t****aking*) pour les anticipations de tour de parole et ***cmp*** (« complementary ») pour les chevauchements complémentaires. Les régulateurs indiquent que l'on suit l'interlocuteur, qu'on le comprend, qu'on est d'accord ou en empathie avec lui ; ils ne dérangent pas ou que peu le locuteur principal (Yngve, 1970 ; Goodwin, 1986 ; Clancy *et al.*, 1996 ; Cerrato & D'Imperio, 2003). À l'opposé, les prises de parole que nous avons étiquetées comme ***tst*** interrompent clairement le locuteur primaire, même si la tentative peut échouer comme pour tout autre acte de langage. L'anticipation de tour de parole correspond au cas où le locuteur qui intervient semble percevoir que le locuteur primaire a terminé (indices de fin de syntagme ou de proposition, mélodie descendante, etc.). Enfin, l'étiquette complémentaire a été introduite pour les chevauchements qui visent à compléter le propos du locuteur principal : une répétition éventuellement paraphrasée de ce que dit le locuteur primaire, un accord ou un désaccord explicite, une brève réponse anticipée, une précision avancée ou requise non seulement sur le contenu mais aussi sur la forme de l'échange (temps imparti, thème abordé), un bon mot ou la poursuite de l'énoncé en cours. Cette étiquette complémentaire, contrairement à celle de prise de parole (***tst***), est attribuée à des commentaires ou des énoncés qui se suffisent à eux-mêmes : le locuteur qui intervient ne prend pas la parole pour développer une argumentation. Ce type de chevauchement peut être favorisé par le contexte situationnel : au-delà des locuteurs directement impliqués dans l'émission, un acteur peut souhaiter apporter des informations supplémentaires à l'auditoire. Le tableau 1 donne un exemple de chaque type de chevauchement.

15 Les différences entre étiquettes peuvent s'avérer subtiles et donner lieu à des interprétations divergentes. L'assignation d'une étiquette unique n'est pas toujours aisée : même les *hmm* peuvent avoir différentes fonctions communicatives, comme indiquer à son interlocuteur une intention d'entrer en scène. D'un extrême à l'autre dans les chevauchements, du régulateur à la prise de parole ***tst***, les transitions progressives sont courantes dans les tours de parole les plus longs.

16 Pour les régulateurs, des catégorisations plus fines ont été proposées (Bertrand *et al.*, 2009), distinguant les fonctions d'accusé-réception (*continuer*), celles consistant à exprimer la compréhension (*understanding*), à montrer l'accord avec l'interlocuteur (*assessment*), etc. Mais de l'aveu même des auteurs, identifier ces différentes catégories est une tâche très délicate, et bien que la nécessité de recourir à plusieurs annotateurs soit

mentionnée, aucune évaluation de l'accord inter-annotateur n'est fournie. Un pré-test sur *L'heure de vérité* nous a conduits à renoncer à une telle annotation en différents sous-types de régulateurs, en raison du trop grand désaccord entre annotateurs (Adda-Decker *et al.*, 2008).

Tableau 1. Types de chevauchements retenus (les locuteurs sont indiqués par les lettres A et B)

Type	Chevauchement de parole
bck (régulateur)	A : c'est simplement /le fait/ /B : hmm/ que...
cmp (complémentaire)	A : j'ai une dernière question /sur/ /B : très courte/ sur votre...
tst (prise de parole)	A : et dans /ce cas.../ B : /je veux/ revenir...
att (anticipation)	A : et cela mène à l'action /humanitaire ?/ B : /eh bien je/ pense

17 À partir de notre corpus, deux émissions ont été traitées par 5 annotateurs. La référence résultait d'une harmonisation à travers une négociation entre les annotateurs, d'abord, une adjudication, ensuite, pour les étiquettes disputées. Le tableau 2 présente pour l'une des émissions la distribution des étiquettes entre les différents annotateurs. Il confirme le caractère intermédiaire de l'étiquette complémentaire, et montre un pourcentage de confusion assez élevé (24 %) entre prises de parole anticipées ou non (***att*** contre ***tst***). Pour les 5 annotations comparées à la référence, la mesure Kappa (de Cohen) de l'accord inter-annotation se situe entre 0,7 et 0,8. Chacune des six émissions suivantes a été traitée par un seul annotateur et passée à un collègue pour vérification. Les corrections impliquaient entre 3 % et 6 % des étiquettes, ce qui peut être vu comme une estimation du taux de désaccords résiduels. L'annotation manuelle du corpus fondée sur les quatre types de chevauchement retenus donne donc un bon accord inter-annotateur. Ce premier résultat permet d'étudier la distribution des chevauchements ainsi que leur lien avec les disfluences et les marqueurs discursifs. Nous y reviendrons en section 4.

Tableau 2. Recouvrement des distributions des étiquettes produites par les 5 annotateurs relativement à la distribution finale des étiquettes pour une émission (de 50 minutes)

Chevauchement de parole		étiquettes des annotateurs (%)			
étiquettes finales	total (#)	***bck***	***cmp***	***tst***	***att***
bck (régulateur)	63	**91,1**	8,0	1,0	0,0
cmp (complémentaire)	50	9,2	**75,8**	15,0	0,0
tst (prise de parole)	107	0,4	3,6	**89,2**	6, 8
att (anticipation)	26	0,0	0,0	24,0	**76,0**

3. Analyse des disfluences et marqueurs discursifs

3.1. Distribution globale

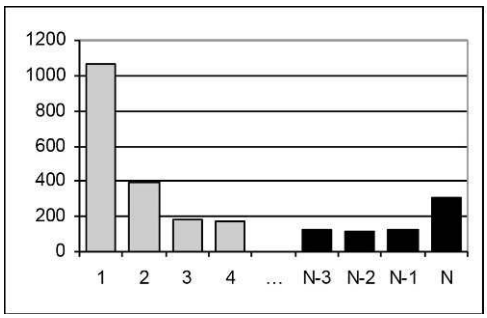
18 Dans une première étude sur la variation diamésique, nous avons analysé les disfluences et marqueurs discursifs dans 9 émissions, écartant 24 minutes de parole superposée. Notre schéma d'annotation nous a permis de classifier les mots impliqués dans les hésitations (1,9 %), les répétitions (2,3 %), les faux départs (2,2 %) et les marqueurs discursifs (2,5 %). Les proportions, calculées par rapport au nombre total de mots dans le corpus, sont relativement bien équilibrées malgré une certaine variabilité inter-locuteur confirmée sur un autre corpus par Bazillon (2011) (*cf.* tableau 3). Les journalistes font davantage d'hésitations et de répétitions, tandis que les invités produisent davantage de faux départs et de marqueurs discursifs. Un test de comparaison de deux proportions indique que chaque différence est significative[3] avec $\alpha = 0,05$. Cette différence peut être due aux difficultés que rencontrent les journalistes quand ils essaient d'interroger leur interlocuteur, alors que les invités essaient de construire une véritable argumentation.

Tableau 3. Disfluences (FP, RP, RV), et marqueurs discursifs (DM) des invités et des journalistes.

Locuteur	**#mots**	**%FP** (hésitations)	**%RP** (répétitions)	**%RV** (faux départs)	**%DM** (marqueurs)	**%dis.+DM**
Brauman	8174	1,2	2,2	3,6	1,5	8,4
de Robien	7589	1,8	1,0	1,9	4,7	9,4
Delors	7462	0,6	3,1	3,4	3,2	10,4
Voynet	7177	2,5	1,8	1,7	4,0	10,0
Pasqua	5385	0,9	1,6	1,5	1,4	5,4
Diouf	4809	0,8	1,9	2,2	0,4	5,6
Brittan	4806	4,3	5,8	3,5	4,4	18,0
Pinay	4006	0,7	3,3	3,3	1,6	8,8
Chevènement	3842	2,8	1,4	0,9	3,8	8,9
Lamassourre	2729	1,1	0,9	0,6	0,6	3,2
Total invités	*55979*	*1,6*	*2,1*	*2,2*	*2,8*	*9,1*
de Virieu	10184	2,3	1,8	1,2	1,6	7,0

Duhamel	7175	2,2	2,8	1,2	1,8	8,0
Colombani	4818	1,3	2,3	2,9	2,2	8,7
du Roy	3706	4,3	2,5	2,7	1,3	10,8
Diop	1904	2,0	1,9	1,4	1,8	7,5
Tesson	1270	2,0	2,5	5,8	2,4	12,7
Giesbert	886	2,2	4,6	3,1	5,6	16,0
Laffon	809	2,6	2,2	0,6	1,9	7,3
d'Orcival	743	1,1	2,2	2,4	3,2	8,9
English	622	3,5	4,2	2,3	3,7	13,7
Total journalistes	*32117*	*2,4*	*2,4*	*1,9*	*2*	*8,7*
Total	**88056**	**1,9**	**2,3**	**2,2**	**2,5**	**8,9**

19 Comme établi par Shriberg (2001), les disfluences et marqueurs discursifs apparaissent dans notre corpus au début plutôt qu'à la fin des énoncés. Ce résultat apparaît dans la figure 3 et se retrouve dans la figure 4, où sont affichés les taux moyens pour les énoncés de moins de 12 mots et de plus de 16 mots. Nous interprétons ce fait par une plus grande difficulté à démarrer plutôt qu'à continuer de parler.



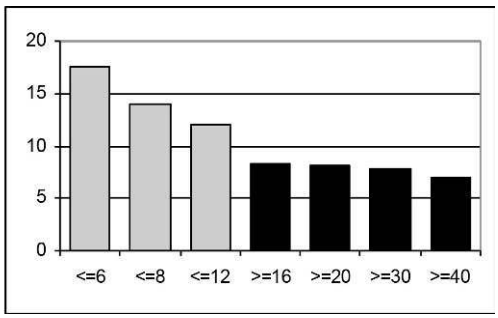


Figure 3 : distribution des disfluences et marqueurs discursifs en fonction de la position dans l'énoncé, de 1 (1er mot) à *N* (dernier mot).

Figure 4 : pourcentage de mots étiquetés comme disfluences ou marqueurs discursifs en fonction de la longueur de l'énoncé (en nombre de mots).

20 Outre le contexte d'apparition, le contenu des hésitations, répétitions, faux départs et marqueurs discursifs a été étudié. Les quelque 1500 hésitations correspondent toutes à *euh* ou ses variantes. Quant aux répétitions, faux départs et marqueurs discursifs, leur distribution en termes d'items lexicaux ou d'expressions figées semble suivre la loi de Zipf.[4]

3.2. Hésitations, répétitions et faux départs

21 Concernant les hésitations, l'analyse montre que plus du tiers d'entre elles apparaissent à une frontière de phrase indiquée par un point final (14 %) ou à une frontière majeure de syntagme indiquée par une virgule (21 %) dans la transcription *bona fide*. Pour les 65 % d'hésitations restantes, le tableau 4 donne la distribution des contextes gauche et droit les plus fréquents, pris indépendamment. Même en milieu de phrase, les hésitations précèdent fréquemment un déterminant ou une préposition, et elles auraient plutôt tendance à suivre une conjonction ou une préposition. Cette asymétrie suggère que l'on évite les hésitations dans les syntagmes nominaux, en particulier entre un déterminant et un nom ; dans ces cas, d'autres mécanismes sont préférés, comme l'allongement de la syllabe finale ou des répétitions. En situation de dialogue, les *euh* d'hésitation ont pu être interprétés comme une façon d'occuper le terrain pour ne pas être interrompu. Ils constituent un signe d'inachèvement par lequel le locuteur souligne son intention de continuer (Morel & Danon-Boileau, 1998 ; Candea, 2000). Nos résultats n'étayent pas entièrement cette interprétation, car il est peu probable qu'une phrase s'arrête après des mots comme *que* ou *et* (parmi les plus fréquents qui apparaissent en contexte gauche d'hésitation). Ils laissent plutôt penser, au-delà de la redondance nécessaire dans les interactions, que c'est le travail de formulation qui joue le rôle de premier plan, rejoignant la présence plus massive des *euh* en début plutôt qu'en fin d'énoncé. Ils demandent cependant à être mis en relation avec les chevauchements de parole (*cf.* § 4).

Tableau 4. Distribution des contextes les plus fréquents, considérés indépendamment, pour les hésitations et les faux départs (nombres d'occurrences et pourcentages)

Ctxt gauche d'hésitation			**Ctxt droit d'hésitation**			**Ctxt droit de faux départ**		
mot	#	%	mot	#	%	mot	#	%
que	40	4	*de*	53	5	*d'*	34	5
et	27	3	*la*	41	4	*l'*	30	4
pour	26	3	*des*	38	4	*la*	29	4
de	21	2	*les*	33	3	*vous*	25	3
avec	19	2	*l'*	26	3	*de*	23	3
à	13	1	*le*	23	2	*on*	21	3
qui	12	1	*un*	21	2	*le*	19	3

22 Les répétitions et les faux départs partagent des traits en commun : d'abord, ils concernent 1 ou 2 mots en moyenne, et il existe une forte corrélation (0,8) entre les décomptes de répétitions et de faux départs des locuteurs. Les locuteurs qui produisent beaucoup de répétitions ont aussi tendance à faire beaucoup de faux départs — alors que pour Shriberg (1994) il y a en anglais les *repeaters* et les *deleters*. Ensuite, les répétitions et

les faux départs les plus fréquents ont tendance à concerner des mots outils monosyllabiques : *de* (72 répétitions et 45 faux départs), *le* (40 répétitions et 39 faux départs), etc. Pour tous les locuteurs, aux deux premières places et dans le même ordre, nous trouvons des mots très fréquents en français. La forme *le* est beaucoup plus souvent un déterminant qu'un pronom, même si rien n'empêche un pronom sujet comme *je* d'être l'un des mots les plus fréquemment répétés ou révisés (Henry & Pallaud, 2003).

23 Dans le tableau 4, la plupart des mots sont communs aux répétitions et aux faux départs. Ce n'est pas surprenant si l'on accepte l'interprétation suivante : dans le processus de sélection lexicale, un mot amorce comme l'article masculin singulier *le* du français peut être répété s'il s'accorde grammaticalement avec ce qui suit et corrigé sinon. Ceci pose des questions intéressantes pour des études syntaxiques comparatives. Le fait qu'il y ait plus de noms masculins que de noms féminins en français — 11 000 contre 9000 dans ILPho (Boula de Mareüil *et al*., 2000), 16 000 contre 12 000 dans le dictionnaire BDLEX (Pérennou & Calmès, 1987) — ne semble pas être un argument suffisant pour expliquer pourquoi *le* surpasse de beaucoup en fréquence *la,* aussi bien pour les répétitions que pour les faux départs. La position de repos que constitue le schwa de *le* est une autre interprétation possible. Cependant, le phénomène reste vrai pour l'indéfini *un* qui est plus souvent repris en *une* que l'inverse. Aussi le masculin *le* (de même, *un*) a-t-il pu être interprété comme forme neutre, non-marquée, du déterminant (Dister, 2007) — à l'instar du *bonhomme* qui, dans les passages pour piétons ou des dessins d'enfants, peut représenter un homme ou une femme. En revanche, la conjonction *et* se prête mal aux faux départs, et se trouve essentiellement présente parmi les répétitions.

24 Un examen de la partie droite du tableau 4 montre que les mots qui suivent le plus fréquemment des mots étiquetés comme faux départs sont « d' » et « l' » —marginaux dans les répétitions. Il s'agit justement des formes abrégées des mots les plus fréquemment impliqués dans des faux départs (*de* et *le*). Cela signifie que les schémas de faux départ les plus fréquents sont de la forme « de d' » devant un mot commençant par une voyelle. Nous avons ensuite *la* (plus fréquent que *le*), ce qui est en accord avec Dister (2007) et avec ce que nous venons de voir. Ensuite, la présence de *vous* ou *on* est frappante puisque ces pronoms personnels sont absents du tableau 5, où sont consignés les mots les plus souvent impliqués dans un faux départ, une répétition ou un marqueur discursif. Ils représentent véritablement des ruptures syntaxiques après un syntagme qui a été abandonné.

Tableau 5. Mots les plus souvent impliqués dans un faux départ, une répétition ou un marqueur discursif (nombres d'occurrences et pourcentages)

Faux départ			**Répétition**			**Marqueur discursif**		
mot	#	%	mot	#	%	mot	#	%
de	45	2	*de*	72	4	*et*	214	10
le	39	2	*le*	40	2	*alors*	141	6
à	15	1	*et*	33	2	*je crois que*	50	2

que	14	1	*je*	29	2	*mais*	44	2
la	13	1	*un*	23	1	*donc*	36	2
les	11	1	*à*	23	1	*eh bien*	33	1
je	11	1	*les*	23	1	*hein*	32	1

25 Les mots pleins (noms et verbes en particulier) peuvent également être impliqués dans les répétitions et les faux départs, et sont davantage affectés par les phénomènes de troncation que ne le sont les mots outils. Ceci n'est pas étonnant, puisqu'ils sont bien plus souvent polysyllabiques. Dans notre schéma d'annotation, les phénomènes de troncation sont éclatés entre répétitions et faux départs, mais ils ne représentent que 0,4 % de notre corpus.

3.3. Marqueurs discursifs

26 Au sein des marqueurs discursifs, on trouve comme attendu des conjonctions, des adverbes et des interjections (*cf.* tableau 5) : *et, alors*, etc. Soit l'un soit l'autre de ces mots est le marqueur discursif le plus fréquent ou le deuxième le plus fréquent chez chaque locuteur, mais l'autre marqueur du doublé de tête est très variable (*cf.* tableau 6). Par exemple, la conjonction *mais* est le quatrième marqueur discursif le plus fréquent, alors qu'il n'apparaît dans le doublé de tête d'aucun locuteur ; et inversement, l'expression *je crois (que)* n'est pas très fréquente, mais elle apparaît dans le doublé de tête de deux locuteurs (*cf.* tableaux 5 et 6). De façon intéressante, ces deux locuteurs sont des invités.

27 Les journalistes sont davantage enclins à utiliser des remplisseurs impersonnels (par exemple l'interjection *hein* pour le modérateur de Virieu). Quant aux invités qui produisent de nombreux marqueurs discursifs, ils recourent à un large éventail d'expressions différentes. Mais est absente, pour tous les locuteurs, la particule *quoi* qui ponctue et régule souvent la conversation ordinaire. Nous avons distingué trois grands types de marqueurs discursifs : structuration (ex. *alors*), prise de position (ex. *je crois que*) et interaction (ex. *hein*). Chacun représente environ un tiers de tous les marqueurs discursifs, même si le dernier sous-type est un peu moins fréquent.

Tableau 6. Les deux marqueurs discursifs les plus fréquents pour chaque locuteur, avec le pourcentage qu'ils représentent au sein de la classe des marqueurs discursifs pour ce locuteur

Invités	mots	%	Journalistes	mots	%
Brauman	*ben, et*	37	Colombani	*alors, et*	33
Brittan	*et, je crois que*	66	de Virieu	*alors, hein*	39
Chevènement	*et, hein*	29	Diop	*alors, donc*	64
Delors	*et, je pense que*	24	Duhamel	*et, alors*	47

de Robien	*et, eh bien*	49	du Roy	*alors, et*	61
Diouf	*et, moi*	50	Tesson	*et, moi*	37
Lamassourre	*et, je crois*	38	English	*alors, et*	33
Pasqua	*et, alors*	33	Giesbert	*alors, bon*	43
Pinay	*et, moi*	35	Langelier	*et, alors*	62
Voynet	*je crois que, lors*	45	d'Orcival	*alors, et*	50

28 Enfin, le patron marqueur discursif + disfluence apparaît deux fois plus que le patron disfluence + marqueur discursif (98 *vs* 50 occurrences). En raison de leur rôle structurant et remplisseur, les marqueurs discursifs sont souvent utilisés en début de message. Nous aurons dans ce qui suit l'occasion d'y revenir.

4. Analyse des chevauchements de parole

29 Une seconde étude de la variation diamésique s'est focalisée sur les chevauchements de parole, absents ou linéarisés dans les transcriptions *bona fide,* mais fréquents dans les échanges parlés (3–4 par minute, en moyenne) même s'ils sont de courte durée (2,5 mots pour des tours de parole de 30 mots en moyenne, moins de 5 % des données en durée cumulée). Ces chevauchements peuvent être regroupés en deux catégories : intrusifs pour les prises de parole (***tst***) et les chevauchements complémentaires (***cmp***), non-intrusifs pour les régulateurs (***bck***) et les anticipations (***att***). Les chevauchements non-intrusifs comme les régulateurs, qui encouragent une interaction fluide, sont particulièrement brefs (*cf.* tableau 7), en accord avec une règle de l'analyse conversationnelle selon laquelle on cherche à minimiser les blancs et les chevauchements entre interlocuteurs. Les chiffres sont comparables pour les locuteurs actifs et passifs (c'est-à-dire entre ceux qui par leur intervention produisent la situation de chevauchement et ceux qui, interrompus, gardent la parole). Mais les locuteurs actifs en situation de prise de parole tendent à parler plus vite ; ils produisent davantage de mots que les locuteurs passifs par rapport à la situation de chevauchement (ceux qui sont interrompus).

Tableau 7. Décompte des segments de parole superposée (nombre de segments d'un type donné, fréquence), nombre de mots du segment et longueur moyenne en mots, pour les rôles P (locuteur passif par rapport à la situation de chevauchement) et A (locuteur actif), pour les types de segments *bck, cmp, tst* et *att*

catégorie		nombre de segments	fréquence (/ min)	#mots	%mots	longueur moyenne (en mots)
bck	P A	461	1,2	719 550	0,8 0,6	1,6 1,2
att	P A	168	0,4	345 391	0,4 0,5	2,1 2,3

cmp	P A	278	0,7	955 974	1,1 1,1	3,4 3,5
tst	P A	438	1,1	1447 1658	1,7 1,9	3,3 3,8

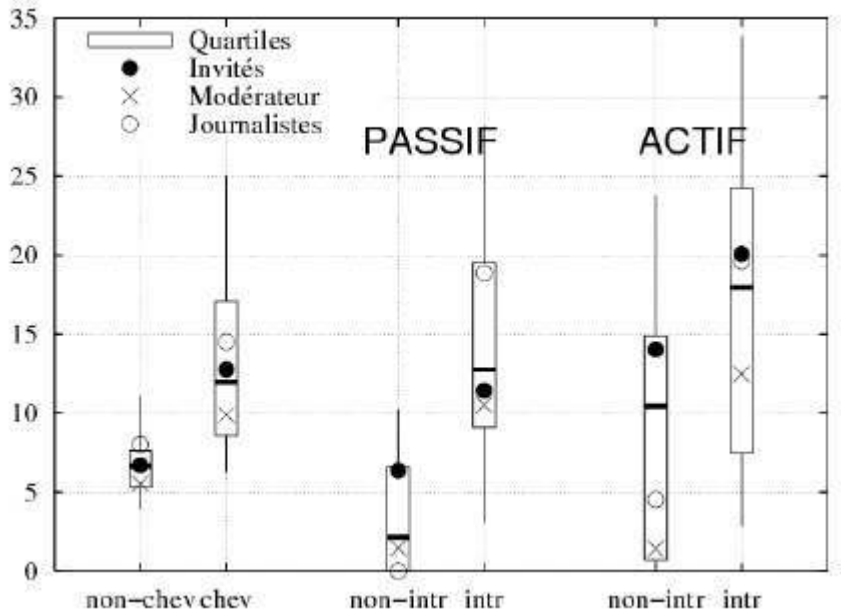
30 Les chevauchements engendrent deux fois plus de disfluences que les portions de parole non-superposée (*cf.* tableau 8). L'augmentation du taux de disfluences touche surtout les répétitions et vaut en particulier pour les locuteurs actifs dans les situations de chevauchements intrusifs comme les prises de parole (***tst***). On mesure davantage de répétitions et de marqueurs discursifs chez les locuteurs actifs que chez les locuteurs passifs par rapport à la situation de chevauchement, ce qui s'explique aussi par la position en début de tour de parole (*cf.* § 3). En même temps, les locuteurs primaires (passifs vis-à-vis de la situation de chevauchement) deviennent spectaculairement disfluents lors de commentaires complémentaires apportés par leurs interlocuteurs. Ceci corrobore le caractère intrusif de ces chevauchements complémentaires, qui ne visent pas un changement de locuteur, mais qui, à la fois par leur longueur et leur contenu informationnel, peuvent déranger le locuteur primaire.

Tableau 8. Taux de marqueurs discursifs et de disfluence (hésitation, répétition ou faux départs) pour la parole non-superposée et superposée — pour des rôles de locuteur passif (P) et actif (A) —, pour chacun des types de chevauchement (intrusif comme les prises de parole ou les chevauchements complémentaires ou non-intrusif comme les régulateurs et les anticipations)

catégorie		marqueur discursif	hésitation	répétition	faux départ	tous types
non-superposé		2,4	2,0	2,5	2,5	6,9
superposé	P	2,1	1,6	7,2	2,3	11,1
	A	5,9	0,5	11,0	3,0	14,5
non intrusif	P	2,4	1,6	1,3	2,0	4,9
	A	7,2	0,6	5,2	0,9	6,7
intrusif	P	2,0	1,6	9,5	2,5	13,6
	A	5,4	0,4	13,0	3,8	17,2

31 En revanche, les régulateurs n'accroissent pas le taux de disfluences des locuteurs passifs. Ce taux est même plus bas que dans la parole non-superposée. On observe enfin des différences intéressantes entre journalistes et invités, dont les rôles sont asymétriques. Même si dans l'ensemble les taux de disfluences sont comparables, la représentation de la figure 5, à base de boîtes à moustaches,[5] montre que les journalistes ont des taux de disfluences plus élevés lorsqu'ils sont locuteurs passifs en situation de chevauchement intrusif (***tst*** ou ***cmp***). Dans ce cas, on semble assister à un échange des rôles standards (actifs dans l'interruption pour les journalistes et passifs pour les invités).

Figure 5. Taux de disfluences pour les différents types de segments de parole, superposée (chev) ou non (non-chev), pour les locuteurs passifs ou actifs par rapport à la situation de chevauchement, et pour les catégories int(intrusive)/non-intr(usive)



5. Conclusion

32 Ainsi, l'alignement de transcriptions *bona fide* et automatique d'une dizaine d'émissions d'interviews télévisées a facilité et accéléré le processus de transcription *verbatim* de ces données audio. Dans ce travail, l'annotation de la transcription *verbatim* a permis de quantifier les proportions de mots impliqués dans les hésitations, les répétitions, les faux départs et les marqueurs discursifs (environ 2 % pour chaque type, hors chevauchements de parole), ainsi que la proportion en durée cumulée des chevauchements de parole (près de 5 % des données). Les occurrences de disfluences, marqueurs discursifs et chevauchements de parole ont été étudiées en fonction de la position dans le discours, du statut du locuteur (rôle, autorité, etc.) et du contexte des mots voisins (selon que certaines séquences sont davantage enclines à apparaître ensemble, dans un ordre donné). D'après les chiffres obtenus, les différences entre français parlé et écrit, au moins dans le style bien policé de l'interview politique, ne sont pas telles qu'il faille poser deux systèmes autonomes. Cependant, bien des phénomènes n'ont pas été traités dans cet article : l'usage de *car* et de temps synthétiques comme le passé simple, l'inversion du verbe et du pronom sujet, à l'écrit ; la substitution de *nous* par *on*, le redoublement du sujet par un pronom anaphorique, la réduction de *cela* en *ça*, les dislocations et la chute du *ne* de négation, à l'oral. Ils pourraient très bien l'être dans le cadre que nous nous sommes fixé, d'analyse comparative de la variation diamésique, comme d'autres dimensions de la variation.

33 Que ce soit pour des études pragmatiques (Mondada, 2001) ou une application à la reconnaissance de la parole, un travail important de modélisation reste à faire. Nous pensons que dresser un inventaire descriptif des différents types de disfluences, marqueurs discursifs et chevauchements de parole y contribue, et peut aider à améliorer

la transcription automatique de la parole conversationnelle, les performances dans l'état de l'art actuel étant encore loin de celles que l'on obtient sur de la parole préparée (Barras, 2008).

34 Une particularité des conversations, plus encore que des interviews politiques, est la compétition relativement fréquente pour la prise de parole entre les intervenants. Les disfluences et les segments de parole superposée y sont donc très communs. Notre corpus de *L'heure de vérité* n'échappe pas non plus à cette règle. Malgré sa taille, les conclusions que l'on peut tirer dépendent du genre de discours, ici l'interview télévisée, et profiteraient d'une comparaison avec diverses formes de parole conversationnelle. En gardant cette fin à l'esprit, nous avons comparé les probabilités des marqueurs discursifs comme *je crois que* ou *je pense que* avec ce que nous trouvons dans d'autres corpus de transcriptions *verbatim* en français — journaux télévisés (3,6 millions de mots) et conversations téléphoniques (1 million de mots). Nous remarquons que pour les interviewés étudiés dans ce travail, les valeurs que nous mesurons sont proches de celles qui peuvent être estimées sur la conversation téléphonique, alors que pour les journalistes de *L'heure de vérité* nous sommes sous le seuil estimé sur les journaux télévisés.

35 Une analyse prosodique a été ébauchée sur le débit de parole de ces expressions, sur l'allongement final des mots outils et sur la distribution des pauses. Les transcriptions *bona fide* et *verbatim* ont également été enrichies d'étiquettes morphosyntaxiques. Cette information sera utilisée dans un avenir proche ; et nous prévoyons d'étudier les relations qui existent entre les disfluences et les tours de parole, ainsi que l'influence que la lutte pour la prise de parole a sur l'apparition des disfluences. Enfin, l'étude que nous avons entreprise ne pourrait que profiter d'une mise en relation avec l'étude des mouvements oculaires et des gestes corporels présents dans la vidéo, puisqu'il s'agit d'enregistrements télévisés.

BIBLIOGRAPHIE

Adda-Decker, M., Habert, B., Barras, C., Adda, G., Boula de Mareüil, P., Paroubek, P. (2003) A disfluency study for cleaning spontaneous speech automatic transcripts and improving speech language models, *3rd ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*, Göteborg, p. 67–70.

Adda-Decker, M., Barras, C., Adda, G., Paroubek, P., Boula de Mareüil, P., Habert, B. (2008) Annotation and analysis of overlapping speech in political interviews, *6th International Conference on Language Resources and Evaluation*, Marrakech, p. 3105–3111.

Auchlin, A. (1981) *Mais heu, pis bon, ben alors voilà quoi !* : Marqueurs de structuration de la conversation et complétude, *Cahiers de linguistique française*, 4, p. 141–159.

Barras, C. (2008) Étude de grands corpus de parole : apport de la transcription automatique, in Bilger, M. (éd.), *Les enjeux de la transcription de la langue parlée*, Perpignan : Presses Universitaires de Perpignan, p. 258–268.

Barras, C., Geoffrois, É., Wu, Z., Liberman, M. (2001) Transcriber: development and use of a tool for assisting speech corpora production, *Speech Communication*, 33, 1–2, p. 5–22.

Barras, C., Adda, G., Adda-Decker, M., Habert, B., Boula de Mareüil, P., Paroubek, P. (2004) Automatic audio and manual transcripts alignment, time-code transfer and selection of exact transcripts, *4th International Conference on Language Resources and Evaluation*, Lisbonne, p. 877–880.

Bazillon, T. (2011) *Transcription et traitement manuel de la parole spontanée en vue de sa reconnaissance automatique*, Thèse de doctorat, Université du Maine, Le Mans.

Bertrand, R., Ader, M., Blache, P., Ferré, G., Espesser, R., Rauzy, S. (2009) Représentation, edition et exploitation de données multimodales : le cas des backchannels du corpus *CID*, *Cahiers de Linguistique*, 33, 2, p. 183–212.

Blanche-Benveniste, C. (1990) *Le français parlé : études grammaticales*, Paris : Éditions du CNRS.

Boula de Mareüil, P., Yvon, F., d'Alessandro, C., Aubergé, V., Vaissière, J., Amelot, A. (2000) A French phonetic lexicon with variants for speech and language processing, *2nd International Conference on Language Resources and Evaluation*, Athènes, p. 273–276.

Boula de Mareüil, P., Habert, B., Bénard, F., Adda-Decker, M., Barras, C., Adda, G., Paroubek, P. (2005) A quantitative study of disfluencies in French broadcast interviews, *4th ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*, Aix-en-Provence, p. 27–32.

Candea, M. (2000) *Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation » en français oral spontané*, Thèse de doctorat, Université Sorbonne Nouvelle, Paris 3.

Catach, N. (1992) *L'orthographe*, Paris : Presses Universitaires de France.

Catach, N. (1994) *La ponctuation*, Paris : Presses Universitaires de France.

Cerrato, L. & D'Imperio, M. (2003) Duration and tonal characteristics of short expressions in Italian, *15th International Congress of Phonetic Sciences,* Barcelone, p. 1213–1216.

Chanet, C. (2004) Fréquence des marqueurs discursifs en français parlé : quelques problèmes de méthodologie, *Recherches sur le français parlé*, 18, p. 83–106.

Clancy, P. M., Thompson, S. A., Suzuki, R., Tao, H. (1996) The conversational use of reactive tokens in English, Japanese, and Mandarin: Conversational structuring across languages, *Journal of pragmatics*, 26, 3, p. 355–387.

Delais-Roussarie, É. & Durand, J. (2003) *Corpus et variation en phonologie du français : méthodes et analyses*, Toulouse : Presses Universitaires du Mirail.

Dister, A. (2007) *De la transcription à l'étiquetage morphosyntaxique. Le cas de la banque de données V ALIBEL*, Thèse de doctorat, Université catholique de Louvain, Louvain-la-Neuve.

Dister, A., Goldman, J.-P., Marlet, R. (à paraître) The orthographic and phonetic transcription, in Lacheret, A., Kahane, S., Pietrandrea, P. (Eds), *Rhapsodie: a prosodic-syntactic treebank for spoken French*, Amsterdam: John Benjamins.

Duez, D. (2001) Caractéristiques acoustiques et phonétiques des pauses remplies dans la conversation en français, *Travaux Interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence,* 20, p. 31–48.

Fernandez-Vest, J. (1994) *Les particules énonciatives*, Paris : Presses Universitaires de France.

Gadet, F. (1997) *Le français ordinaire*, Paris : Armand Colin.

Gauvain, J.-L., Adda, G., Adda-Decker, M., Allauzen, A., Gendner, V., Lamel, L., Schwenk, H. (2005) Where are we in transcribing French broadcast news?, *9th European Conference on Speech Communication and Technology*, Lisbonne, p. 1665–1668.

Goodwin, C. (1986) Between and within. Alternative sequential treatments of continuers, *Human Sstudies*, 9, 2–3, p. 205–217.

Grice, H. P. (1975) Logic and conversation, *Syntax and Semantics*, 3, p. 41–58.

Henry, S. & Pallaud, B. (2003) Word fragments and repeats in spontaneous spoken French, *3rd ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*, Göteborg, p. 77–80.

Labov, W. (1976) *Sociolinguistique*, Paris : Éditions de Minuit.

Levelt, W. J. M. (1989) *Speaking. From intention to articulation*, Cambridge: The MIT Press.

Mondada, L. (2001) Pour une linguistique interactionnelle, *Marges linguistiques*, 1, p. 142–162.

Morel, M.-A. & Danon-Boileau, L. (1998) *Grammaire de l'intonation. L'exemple du français*, Paris : Ophrys.

Pérennou, G. & Calmès, M. de (1987) *BDLEX, base de données lexicales du français écrit et parlé*, Toulouse : Travaux du laboratoire CERFIA.

Roulet, E., Auchlin, A., Moeschler, J., Rubattel, C., Schelling, M. (1991) *L'articulation du discours en français contemporain*, Berne : Peter Lang.

Schegloff, E. A., Jefferson, G., Sacks, H. (1977) The preference for self-correction in the organization of repair in conversation, *Language*, 53, p. 361–382.

Schegloff, E. A. (2000) Overlapping talk and the organization of turn-taking for conversation, *Language in Society*, 29, p. 1–63.

Shriberg, E. (1994) *Preliminaries to a theory of speech disfluencies*, Thèse de doctorat, University of California, Berkeley.

Shriberg, E. (2001) To 'errrr' is human: ecology and acoustics of speech disfluencies, *Journal of the International Phonetic Association*, 31, 1, p. 153–169.

Tanguy, N. (2009) *Les segments averbaux, unités syntaxiques de l'oral*, Thèse de doctorat, Université Sorbonne Nouvelle, Paris 3.

Yngve, V. (1970) On getting a word in edgewise, *6th Regional Meeting of the Chicago Linguistic Society*, Chicago, p. 567–577.

NOTES

1. Ce travail a été mené dans le cadre d'une action transversale du département Communication Homme-Machine du LIMSI, cherchant à combiner des compétences en traitement automatique de l'oral comme de l'écrit.

2. http://www.ldc.upenn.edu/Projects/MDE/

3. (Voir formule) $z = (p_1 - p_2) / \sqrt{p(1-p)(1/n_1 + 1/n_2)} > 1{,}96$ en valeur absolue

où n_1 est le nombre de mots pronouncés par les journalistes,
n_2 est le nombre de mots pronouncés par les invités,
p_1 est la proportion de disfluences ou de marqueurs discursifs produits par les journalistes,
p_2 est la proportion de disfluences ou de marqueurs discursifs produits par les invités,
$p = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$.

4. De fait, la distribution est quasi-linéaire sur une échelle logarithmique. Pour diverses représentations graphiques, voir Boula de Mareüil *et al.* (2005) ainsi qu'Adda-Decker *et al.* (2003, 2008).

5. La visualisation a été obtenue en utilisant le logiciel R (http://www.r-project.org). Dans une boîte à moustaches (*boxplot* en anglais), la barre horizontale représente la médiane, la boîte représente les 2e et 3e quartiles, tandis que les barres verticales au-dessus et au-dessous de la boîte représentent les 1er et 4e quartiles.

RÉSUMÉS

À l'interface entre linguistique de corpus et traitement automatique, le travail présenté ici vise à éclairer les différences relevant de la variation diamésique, entre oral et transcriptions écrites à différents niveaux d'élaboration. Un corpus d'interviews télévisées a été utilisé (une dizaine d'émissions de *L'heure de vérité*), impliquant des journalistes et des hommes politiques ou des personnalités représentant la société civile. Les transcriptions *bona fide* réalisées à destination de la presse (où la plupart des disfluences et marqueurs discursifs ainsi que les chevauchements de parole ont été éliminés) ont été alignées avec les transcriptions fournies par un système de reconnaissance de la parole, facilitant la production de transcriptions *verbatim* où tous les événements audibles (y compris la parole superposée) ont été transcrits manuellement.

Trois types de disfluences (les hésitations, les répétitions et les faux départs) ont été distingués et annotés, de même que les marqueurs discursifs. Ces derniers, ainsi que chaque type de disfluences, représentent environ 2 % des mots du corpus hors chevauchements de parole (8 % au total). Ils ont été analysés par type d'énoncé, de locuteur (selon le statut journaliste/invité) et de patrons les plus fréquents.

Les chevauchements de parole ont également été annotés manuellement au moyen de quatre étiquettes, pour les régulateurs ou *backchannels* du type *hmm*, pour les prises de parole (avec interruption claire d'un locuteur par un autre), pour les anticipations de tour de parole (quand le locuteur qui intervient semble percevoir que son interlocuteur a terminé) et pour les chevauchements complémentaires (commentaires ou énoncés qui se suffisent à eux-mêmes, où le locuteur qui intervient ne prend pas la parole pour développer une argumentation). L'interaction entre chevauchements de parole, disfluences et marqueurs discursifs a ensuite été étudiée, et les taux d'occurrences observés ont été mis en relation avec le rôle des locuteurs. Les chevauchements de parole sont assez fréquents (en moyenne 3–4 par minute), même s'ils sont de courte durée (5 % des données), les chevauchements non-intrusifs comme les régulateurs de type *hmm* étant plus courts que les chevauchements intrusifs comme les prises de parole. Les disfluences sont deux fois plus nombreuses en parole superposée qu'en parole non-superposée : les répétitions, en particulier, sont concernées au premier chef. Enfin, des différences intéressantes ont été relevées entre les comportements actifs/passifs (de celui qui prend ou qui a la parole) des journalistes et des invités.

At the interface between corpus linguistics and automatic speech processing, this study aims at increasing our understanding of spontaneous speech-related phenomena, based on 8 hours of television shows (*L'heure de vérité*) of French political interviews recorded in the early nineties. During each show, a political figure or a representative of civil society is interviewed by several journalists. The reported work focuses on the transcription, annotation and analysis of discourse

markers, disfluencies and speech overlaps. Press-oriented (*bona fide*) transcripts available for these shows and the output of a speech recognition system were used and aligned to speed up the transcription process, in order to provide a fine-grained (*verbatim*) transcription of the audio data, including all audible speech events. Sibling corpora are very useful resources to facilitate hand corrections. A segmentation into multi-speaker speech portions was also performed manually by relaxing temporal anchoring constraints in the case of overlaps, because even the precise localisation of the beginning and the end of such events is not straightforward. The Transcriber software (trans.sourceforge.net/en/presentation.php) was customised accordingly to facilitate this task. Two situations were distinguished: (1) the overlap does not entail a speaker change (the primary speaker remains the same at the end of the overlap); (2) the primary speaker stops and the secondary speaker becomes the primary speaker of a new turn.

Three types of disfluencies were distinguished: filled pauses, repetitions and false starts. Together with discourse markers, they were analysed by utterance, speaker and pattern types. Silent pauses and lengthening phenomena were also measured, but they are not addressed in this paper. Speech overlaps were annotated by using 4 tags: back-channel, turn stealing, anticipated turn taking, and complementary. Back-channels like "`hmm"s indicate that we follow our interlocutor, understand him/her, agree with him/her; they barely disturb the main speaker. On the opposite, turn stealings clearly interrupt the main speaker, even though the attempt may fail as any other speech act. Anticipated turn taking corresponds to the case where the incoming speaker seems to perceive cues indicating that the main speaker has finished (phrase or clause boundary, falling pitch, etc.). Finally, the complementary label was introduced for overlaps which aim at complementing the main speaker's utterance: a possibly paraphrased repetition of the primary speaker's statement, an explicit agreement or disagreement, a short anticipated answer, a precision forwarded or required, not only on the content but also on the form of the exchange (schedule, approached topic), a witty remark or the continuation of the utterance. This complementary label, contrary to the turn stealing one, is assigned to self-sufficient comments or utterances: the entering speaker does not take the floor to develop an argument. This type of overlap may be favoured by the situational context: beyond the speakers actively involved in the show, an actor may wish to provide additional information to the audience.

Differences between overlap tags may happen to be subtle and give rise to diverging interpretations. A unique label assignment is not always straightforward. Even "hmm"s can have different communicative functions such as signalling that one is eager to jump in. From one extreme to the other, progressive transitions are common during long-lasting turns. Two shows were annotated by 5 annotators, and the reference resulted from harmonising the different annotations through first negotiation, then adjudication, for the disputed labels. The label distribution for the different annotators confirmed the intermediate nature of the complementary label, and showed a rather high confusion percentage (24%) between anticipated turn takings and turn stealings. Yet, the manual annotation of the corpus based on the four overlap types gave a good inter-annotator agreement (Kappa measures around 0.7). This first result allowed us to study the distribution of overlaps and their interplay with disfluencies and discourse markers.

In non-overlapping speech, each disfluency type (as well as discourse markers) accounts for about 2% of the corpus. Among disfluencies, hesitations (transcribed as *euh* in French) can be found almost anywhere. More precisely, 35% of filled pauses occur at a sentence boundary indicated by a full stop (14%) or at a major phrase boundary indicated by a comma (21%) in the *bona fide* transcription. In the middle of a sentence, hesitations frequently precede a determiner or a preposition and they rather follow a conjunction or a preposition. This asymmetry suggests that hesitations are avoided within noun phrases, especially between a determiner and a noun. In this situation, other mechanisms such as final lengthening or repetitions are preferred. Repetitions and false starts exhibit some features in common: first, they both involve 1 or 2

words on average, and there is a high correlation (0.8) among speakers between their numbers of repetitions and false starts: speakers who produce many repetitions also tend to make many revisions. Second, most frequent repetitions and false starts tend to be monosyllabic function words: *de* 'of', *le* (corresponding to the determiner 'the' far more often than the pronoun 'him'), etc. Interestingly, *le* outweighs *la* in both repetitions and false starts: it may be considered not only as the masculine form but also as the neutral form of the determiner. By contrast, the conjunction *et* 'and' hardly lends itself to revisions, and it is only found among repetitions. It may also be considered as a discourse marker: as such, it is even more frequent than *alors* 'so' in the corpus.

Our study then focused on overlaps, which are frequent (3-4 per minute on average) even if they are short (2.5 words compared to 30-word speaker turns on average). Their cumulative duration represents less than 5% of the data. Non-intrusive overlaps such as back-channels, which encourage a fluid interaction, are particularly short. Figures are comparable for active and passive speakers (i.e. incoming speakers who produce the overlap situation and floor holding speakers who are interrupted). However, active speakers in the turn stealing situation tend to speak faster (they produce more words) than their passive competitors.

Overlaps generate twice as many disfluencies as non-overlapping speech portions. The disfluency rate increase mainly concerns repetitions, in particular for active speakers in intrusive overlap situations such as turn stealings. More repetitions and discourse markers are observed for active speakers than for passive speakers, which can also be explained by the turn-start position. Our study showed that disfluencies and discourse markers occur at the beginning rather than at the end of utterances.

Passive (primary) speakers become dramatically disfluent within complementary comments brought by their interlocutors. This corroborates the intrusive nature of these complementary overlaps which do not aim at a speaker change but may disturb the main speaker due to their length and informational content. By contrast, back-channels do not increase the disfluency rate of passive speakers. This rate is even lower than it is in non-overlapping speech.

Finally, interesting differences are observed between journalists and interviewees, whose roles are asymmetric. Even though their disfluency rates are on the whole comparable, journalists show higher disfluency rates when they are passive speakers in intrusive (turn stealing or complementary) overlap situations. In this case, there seems to be an exchange of standard roles (active interruption for journalists and passive overlaps for interviewees).

Enriched and more accurate models are necessary for both talk-in-interaction analysis and speech recognition. We think that drawing up a descriptive inventory of discourse markers, disfluencies and speech overlaps may contribute to the design of a pragmatics model and may be profitable to improve automatic conversational speech transcription, whose performance is still poor as compared to prepared speech recognition.

INDEX

Mots-clés : français parlé, parole spontanée, disfluences
Keywords : spoken French, spontaneous speech, disfluencies

AUTEURS

PHILIPPE BOULA DE MAREÜIL

LIMSI-CNRS, Orsay
Philippe.Boula.de.Mareuil@limsi.fr

GILLES ADDA

LIMSI-CNRS, Orsay
Gilles.Adda@limsi.fr

MARTINE ADDA-DECKER

LIMSI-CNRS, Orsay
LPP, Université Sorbonne Nouvelle, Paris
Martine.Adda-Decker@limsi.fr

CLAUDE BARRAS

LIMSI-CNRS, Orsay
Université Paris Sud, Orsay
Claude.Barras@limsi.fr

BENOÎT HABERT

LIMSI-CNRS, Orsay
ENS, Lyon
benoit.habert@ens-lyon.fr

PATRICK PAROUBEK

LIMSI-CNRS, Orsay
Paroubek@limsi.fr