



HAL
open science

La focalisation prosodique dans la parole interprétée en français

Rémi Godement-Berline

► **To cite this version:**

Rémi Godement-Berline. La focalisation prosodique dans la parole interprétée en français. Linguistique. Université Sorbonne Paris Cité, 2018. Français. NNT : 2018USPCC005 . tel-02011350

HAL Id: tel-02011350

<https://theses.hal.science/tel-02011350>

Submitted on 7 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat
de l'Université Sorbonne Paris Cité
Préparée à l'Université Paris Diderot
Ecole doctorale « Sciences du Langage » (ED 132)
Laboratoire de Linguistique Formelle (LLF)

La focalisation prosodique dans la parole interprétée en français

Par Rémi Godement-Berline

Thèse de doctorat de Linguistique théorique, descriptive et
automatique

Dirigée par Philippe Martin

Présentée et soutenue publiquement à Paris le vendredi 23 février 2018

Présidente du jury : Barbara Hemforth, Directrice de recherche, CNRS, Université Paris Diderot

Rapporteur : Antoine Auchlin, Maître d'enseignement et de recherche (HDR), Université de Genève

Rapporteur : Henri-José Deulofeu, Professeur émérite, Aix-Marseille Université

Examinatrice : Hiyon Yoo, Maîtresse de conférences, Université Paris Diderot

Examinatrice : Emanuela Cresti, Professeure émérite, Université de Florence

Directeur de thèse : Philippe Martin, Professeur émérite, Université Paris Diderot



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Titre : La focalisation prosodique dans la parole interprétée en français.

Résumé : La focalisation prosodique désigne le soulignement d'un constituant dans un énoncé au moyen de différentes ressources prosodiques, en particulier l'accentuation et l'intonation. Plusieurs fonctions sont attribuées à la focalisation : le marquage des différentes catégories de focus, ainsi que des fonctions emphatiques (ici appelées insistance et expressivité). Cette thèse a pour principal but de savoir si la focalisation et ses fonctions présentent des propriétés spécifiques dans le phonogène de la parole interprétée, c'est-à-dire l'oralisation d'un texte écrit mémorisé au préalable par le locuteur (généralement un comédien). Cette question présente un intérêt pour la linguistique et la phonétique à plusieurs titres. Tout d'abord, les différences de réalisation prosodique entre les fonctions de la focalisation sont encore mal connues. Par ailleurs, peu d'études ont été consacrées aux caractéristiques prosodiques de la parole interprétée. Enfin, notre thèse présente un apport sur le plan méthodologique à travers le protocole relativement novateur de ses deux expériences. Dans une expérience de production, des locuteurs ont reproduit des conversations spontanées en parole lue et en parole interprétée. Un groupe d'experts en prosodie a ensuite relevé les occurrences de focalisation dans le corpus et a effectué une classification fonctionnelle de ces occurrences. Nous avons également mené une expérience de perception afin de comparer la réalisation des fonctions de la focalisation indépendamment du phonogène. Malgré un taux d'accord entre experts relativement faible (ce qui soulève plusieurs questions méthodologiques et théoriques), nos analyses révèlent plusieurs résultats originaux. La fréquence d'occurrence de la focalisation est la plus élevée en parole interprétée, suivie de la parole lue. Ce résultat confirme notre prédiction et suggère que la parole interprétée est un phonogène favorable à l'étude de la focalisation. Une forte relation est observée entre la fonction d'insistance et le trait d'accentuation initiale, ce qui confirme de nombreuses études précédentes. Le phonogène se révèle en revanche avoir très peu d'influence sur la réalisation de la focalisation et de ses fonctions. Ce résultat est dû selon nous à un manque de données et au fait que certains traits prosodiques n'ont pas été pris en compte dans l'analyse.

Mots clefs : prosodie, phonogène, parole spontanée et lue, focus, emphase, insistance, expressivité

Title : Prosodic highlighting in interpreted speech in French.

Abstract : Prosodic highlighting refers to the distinction of a constituent through various prosodic means, especially accentuation and intonation. It is taken to fulfill several functions: marking the different types of focus, as well as emphatic functions (named here “insisting” and “expressiveness”). The main goal of this thesis is to determine whether prosodic highlighting and its functions display specific features in interpreted speech, a speaking style that can be defined as the oralization of a written text previously memorized by the speaker (typically an actor). This question is relevant for linguistics and phonetics on several counts. First, little is still known about prosodic differences between functions of prosodic highlighting. Moreover, few studies have analyzed the prosodic characteristics of interpreted speech. Finally, through their innovative protocols, the two experiments described in this thesis present a methodological contribution. A production experiment consisted in having speakers replicate spontaneous conversations in read and interpreted speech. A group of experts then annotated the occurrences of prosodic highlighting in the corpus, and assigned a function to each occurrence. A perception experiment was also led in order to compare the realization of each function independently of speaking style. Despite a relatively low agreement rate between experts (which raises several methodological and theoretical questions), our analyses reveal several important results. The frequency of occurrence of prosodic highlighting is highest in interpreted speech, followed by read speech. This confirms our prediction and suggests that interpreted speech is more suited to the study of prosodic highlighting than other speaking styles. A strong association is observed between insisting and initial secondary accent, which confirms many previous studies. However, there is almost no influence of speaking style on the realization of prosodic highlighting and its functions. We attribute this result to a lack of data and to the fact that some prosodic features were not taken into account in the analysis.

Keywords : prosody, speaking style, read versus spontaneous speech, focus, emphasis, expressiveness

À ma famille

Remerciements

Mes remerciements vont tout d'abord à Philippe Martin, mon directeur de thèse, pour la confiance qu'il m'a accordée en m'acceptant comme doctorant, ainsi que pour ses conseils toujours judicieux. Je remercie ensuite Jean-Marie Marandin, qui m'a également beaucoup aidé dans la préparation de ce travail, aussi bien pour définir les limites du sujet que pour concevoir les expériences. Enfin, je remercie Barbara Hemforth pour son aide dans la réalisation de ces expériences, ainsi que dans l'interprétation des résultats. Je les remercie par ailleurs tous les trois pour leurs relectures attentives des différentes versions du manuscrit.

Je remercie les membres de mon jury de soutenance d'avoir accepté notre invitation et d'avoir bien voulu examiner mon travail, et en particulier Antoine Auchlin et Henri-José Deulofeu d'avoir accepté de rédiger les pré-rapports.

Pendant la réalisation de cette thèse, j'ai pu bénéficier de l'assistance et de la compétence de nombreuses personnes. Je souhaiterais remercier notamment Fabián Santiago Vargas pour l'analyse prosodique de l'expérience de production, Hiyon Yoo pour la préparation de cette même expérience ainsi que celle de perception, et Elisabeth Delais-Roussarie pour l'état de l'art sur la prosodie du français. Pour des discussions très éclairantes, je remercie Wallace Chafe, Ioana Chitoran, Georges Boulakia, Nicolas Ballier, Mary-Annick Morel, Pascal Amsili, Lisa Brunetti, Cristel Portes et Philip Miller. Pour m'avoir appris à manipuler les logiciels qu'ils ont conçus, je remercie Mathieu Avanzi (Analor), Céline De Looze (ADoReVA) et Daniel Hirst (Momel-Intsint). Je remercie également les ingénieur(e)s de recherche du LLF et du Labex EFL : Doriane Gras et Rachel Shen pour l'élaboration des expériences et pour le traitement statistique, et Alexandre Roulois et Loïc Liégeois pour la mise en ligne du Corpus RepTask « Parole interprétée ». Enfin, je remercie bien sûr le groupe d'experts en prosodie qui a annoté les occurrences de focalisation prosodique dans ce corpus, ainsi que les locuteurs de l'expérience de production et les participants du test de perception.

Pour leur joie de vivre et pour tant de bons souvenirs, je remercie les doctorant(e)s de CLILLAC-ARP, de LLF et du comité d'organisation des journées DoSciLa.

Pour m'avoir permis de mettre en application les connaissances acquises au cours de ce doctorat, je remercie le comédien et metteur en scène Arnaud Churin, qui m'a accueilli dans

l'atelier « Exploration de la langue » qu'il a mené ces deux dernières années à l'Ecole d'art dramatique du Théâtre National de Bretagne.

Enfin, je remercie mes amis et ma famille pour leur soutien tout au long de ce travail.

Table des matières

Remerciements	5
Table des illustrations	12
Introduction.....	15
Chapitre 1 : La prosodie du français.....	20
1. Introduction : qu'est-ce que la prosodie ?	20
1.1. Traits acoustiques de la prosodie	20
1.2. Fonctions de la prosodie	21
1.3. Domaines formels de la prosodie	23
1.4. Caractéristiques du français.....	24
2. L'accentuation du français	25
2.1. Catégories d'accents	25
2.2. Unités métriques.....	27
2.3. Contraintes rythmiques	29
2.4. Réalisation des accents.....	31
3. La structure prosodique du français	33
3.1. Définition de la structure prosodique.....	33
3.2. Règles de la structure prosodique	34
3.3. Niveaux de la structure prosodique.....	36
4. L'intonation du français.....	40
4.1. Contours continuatifs	40
4.2. Marquage de la modalité.....	43
4.3. Désaccentuation post-focale ou pré-focale.....	47
4.4. Fonctions expressives	49
5. Résumé.....	51
Chapitre 2 : Influence du phonogène sur la fréquence d'occurrence et la réalisation prosodique de la focalisation	53
1. Introduction.....	53
1.1. Questions de recherche	53
1.2. Arrière-plan et prédictions.....	54

2.	La focalisation prosodique	55
2.1.	Définition	55
2.2.	Traits prosodiques.....	55
3.	Parole spontanée, lue et interprétée.....	57
3.1.	Définitions.....	57
3.2.	Différences prosodiques	59
4.	Méthodologie.....	63
4.1.	Constitution du corpus.....	64
4.1.1.	Elicitation de parole spontanée.....	65
4.1.2.	Reproduction en parole lue et interprétée	67
4.2.	Relevé des occurrences de focalisation	71
4.2.1.	Sélection des experts.....	71
4.2.2.	Annotation du corpus.....	72
4.3.	Analyse prosodique.....	75
4.3.1.	Alignement texte-parole	75
4.3.2.	Mesures de hauteur et de durée.....	77
4.3.3.	Catégories phonologiques	78
5.	Résultats.....	85
5.1.	Taux d'accord entre les experts.....	86
5.2.	Fréquence d'occurrence	88
5.3.	Réalisation prosodique	90
5.3.1.	Hauteur.....	90
5.3.2.	Durée syllabique	92
5.3.3.	Contour intonatif	97
5.3.4.	Etendue du contour.....	99
5.3.5.	Accentuation initiale.....	101
5.4.	Résultats de l'étude sur corpus réduit.....	103
5.4.1.	Taux d'accord entre les experts	104
5.4.2.	Contour intonatif	105
6.	Discussion.....	106
6.1.	Relevé des occurrences de focalisation	106
6.2.	Fréquence d'occurrence de la focalisation	109
6.3.	Réalisation prosodique de la focalisation	110

6.4.	Influence du phonogène sur la fréquence d'occurrence	114
6.5.	Influence du phonogène sur la réalisation prosodique	116
7.	Conclusions.....	117
8.	Résumé.....	120
Chapitre 3 :	Réalisation prosodique des fonctions de la focalisation	123
1.	Introduction.....	123
1.1.	Questions de recherche	123
1.2.	Arrière-plan et prédictions.....	124
2.	Le marquage de focus	124
2.1.	Catégories de focus.....	125
2.1.1.	Focus informationnel.....	126
2.1.2.	Focus contrastif	130
2.1.3.	Focus associatif.....	132
2.1.4.	Verum focus.....	133
2.2.	Réalisation prosodique	134
2.2.1.	Traits du marquage de focus informationnel.....	134
2.2.2.	Traits du marquage de focus contrastif.....	136
2.2.3.	Traits du marquage de focus associatif.....	137
2.2.4.	Traits du marquage de verum focus.....	138
2.2.5.	Poids des méthodes expérimentales.....	138
3.	L'emphase	139
3.1.	Catégories d'emphase	140
3.1.1.	Insistance	140
3.1.2.	Expressivité.....	142
3.2.	Réalisation prosodique	143
3.2.1.	Traits de l'insistance	143
3.2.2.	Traits de l'expressivité	144
3.2.3.	Poids des méthodes expérimentales.....	145
4.	Etude de production.....	147
4.1.	Introduction	147
4.2.	Méthodologie.....	147
4.2.1.	Constitution du corpus	147

4.2.2.	Classification des occurrences de focalisation	148
4.2.3.	Analyse prosodique	151
4.3.	Résultats.....	152
4.3.1.	Taux d'accord entre les experts	152
4.3.2.	Hauteur	155
4.3.3.	Durée syllabique	157
4.3.4.	Contour intonatif	164
4.3.5.	Etendue du contour	171
4.3.6.	Accentuation initiale.....	173
4.3.7.	Résultats de l'étude sur corpus réduit	175
4.4.	Discussion.....	177
4.4.1.	Classification des occurrences de focalisation	177
4.4.2.	Réalisation prosodique des fonctions	180
4.4.3.	Influence du phonogène sur la réalisation	183
4.5.	Conclusions	185
5.	Etude de perception.....	187
5.1.	Introduction	187
5.2.	Méthodologie.....	188
5.2.1.	Participants	188
5.2.2.	Stimuli	189
5.2.3.	Protocole	194
5.3.	Résultats.....	196
5.3.1.	Evaluation des items de contrôle	197
5.3.2.	Evaluation des réalisations	199
5.3.3.	Evaluation des fonctions	201
5.3.4.	Evaluation des réalisations par fonction	203
5.4.	Discussion.....	206
5.4.1.	Evaluation des réalisations et des fonctions	206
5.4.2.	Réalisation phonétique des fonctions	207
5.5.	Conclusions	209
6.	Discussion générale.....	211
7.	Résumé.....	214
	Conclusion générale	220

Bibliographie	228
Annexe A : Matériel de l'expérience de production	242
Annexe B : Autres résultats des études sur corpus réduit.....	251
Annexe C : Matériel de l'expérience de perception	268

Table des illustrations

FIGURE 1. VARIATION DE L'EMPAN SYLLABIQUE DU PIC DE HAUTEUR DES ACCENTS.	33
FIGURE 2. VARIATION DU PROFIL DU PIC DE HAUTEUR DES ACCENTS.	33
FIGURE 3. COURBE DE FREQUENCE FONDAMENTALE D'UN ENONCE ISSU DU CORPUS DE L'EXPERIENCE DE PRODUCTION.	41
FIGURE 4. ILLUSTRATION PAR DELATTRE (1966) DE LA DISTINCTION ENTRE CONTINUATION MINEURE ET MAJEURE.	42
FIGURE 5. ILLUSTRATION PAR MARTIN (2009) DU PRINCIPE DE CONTRASTE DE PENTE.	43
FIGURE 6. COURBES DE FREQUENCE FONDAMENTALE DE QUATRE ENONCES DECLARATIFS ISSUS DU CORPUS DE L'EXPERIENCE DE PRODUCTION.	44
FIGURE 7. COURBES DE FREQUENCE FONDAMENTALE DE TROIS ENONCES INTERROGATIFS ISSUS DU CORPUS DE L'EXPERIENCE DE PRODUCTION.	45
FIGURE 8. COURBES DE FREQUENCE FONDAMENTALE DE QUATRE ENONCES EXCLAMATIFS ISSUS DU CORPUS DE L'EXPERIENCE DE PRODUCTION.	46
FIGURE 9. COURBES DE FREQUENCE FONDAMENTALE DE DEUX ENONCES IMPERATIFS ISSUS DU CORPUS DE L'EXPERIENCE DE PRODUCTION.	46
FIGURE 10. COURBES DE FREQUENCE FONDAMENTALE DE DEUX ENONCES ISSUS DU CORPUS DE L'EXPERIENCE DE PRODUCTION CONTENANT UNE DESACCENTUATION POST-FOCALE APRES UN CONTOUR TERMINAL.	48
FIGURE 11. COURBE DE FREQUENCE FONDAMENTALE D'UN ENONCE ISSU DU CORPUS DE L'EXPERIENCE DE PRODUCTION CONTENANT UNE DESACCENTUATION PRE-FOCALE AVANT LE CONTOUR TERMINAL.	48
FIGURE 12. COURBE DE FREQUENCE FONDAMENTALE D'UN ENONCE ISSU DU CORPUS DE L'EXPERIENCE DE PRODUCTION CONTENANT UNE DESACCENTUATION POST-FOCALE APRES UN CONTOUR INTERROGATIF MONTANT.	49
FIGURE 13. COURBES DE FREQUENCE FONDAMENTALE DE TROIS ENONCES ISSUS DU CORPUS DE L'EXPERIENCE DE PRODUCTION EXPRIMANT DES ATTITUDES.	51
FIGURE 14. CAPTURE D'ECRAN DU LOGICIEL PRAAT CONTENANT L'ALIGNEMENT TEXTE-PAROLE D'UN ENONCE, AINSI QUE LES ANNOTATIONS DES EXPERTS.	75
FIGURE 15. CAPTURE D'ECRAN DU LOGICIEL PRAAT CONTENANT LES INFORMATIONS ACOUSTIQUES D'UN ENONCE, AINSI QUE SA SEGMENTATION EN MOTS, SYLLABES ET PHONES OBTENUE A L'AIDE DE L'EXTENSION EASYALIGN.	76
FIGURE 16. FOCALISATIONS ISSUES DU CORPUS COMPORTANT UN CONTOUR MONTANT H*H% ET H*H-.	80
FIGURE 17. FOCALISATIONS ISSUES DU CORPUS COMPORTANT UN CONTOUR DESCENDANT L*L% ET L*L-.	80
FIGURE 18. FOCALISATIONS ISSUES DU CORPUS COMPORTANT UN CONTOUR MONTANT L*H% ET L*H-.	80
FIGURE 19. FOCALISATIONS ISSUES DU CORPUS COMPORTANT UN CONTOUR MONTANT-DESCENDANT H*L% ET H*L-.	81

FIGURE 20. FOCALISATIONS ISSUES DU CORPUS COMPORTANT UN CONTOUR MONTANT H*, UN CONTOUR DESCENDANT L* ET UN TON HAUT INITIAL HI.	81
FIGURE 21. FOCALISATIONS ISSUES DU CORPUS PRESENTANT DIFFERENTES ETENDUES SYLLABIQUES DU CONTOUR INTONATIF.	83
FIGURE 22. CAPTURE D'ECRAN DU LOGICIEL ANALOR.	85
FIGURE 23. HAUTEUR MOYENNE DES SYLLABES FOCALISEES ET DES AUTRES SYLLABES DU CORPUS.	91
FIGURE 24. HAUTEUR MOYENNE DES SYLLABES FOCALISEES EN PAROLE INTERPRETEE, EN PAROLE LUE ET EN PAROLE SPONTANEE.	92
FIGURE 25. DUREE MOYENNE DES SYLLABES FOCALISEES ET DES AUTRES SYLLABES DU CORPUS.	93
FIGURE 26. DUREE MOYENNE DES SYLLABES FOCALISEES EN PAROLE INTERPRETEE, EN PAROLE LUE ET EN PAROLE SPONTANEE.	94
FIGURE 27. DUREE MOYENNE DES SYLLABES DE PAROLE INTERPRETEE, DE PAROLE LUE ET DE PAROLE SPONTANEE.	96
FIGURE 28. POURCENTAGE D'OCCURRENCE DES DIFFERENTS CONTOURS INTONATIFS PORTES PAR LES CONSTITUANTS FOCALISES.	98
FIGURE 29. POURCENTAGES D'OCCURRENCE DES DIFFERENTS CONTOURS INTONATIFS EN PAROLE INTERPRETEE, EN PAROLE LUE ET EN PAROLE SPONTANEE.	99
FIGURE 30. POURCENTAGE D'OCCURRENCE DES DIFFERENTS TYPES D'ETENDUE SYLLABIQUE DU CONTOUR INTONATIF SUR LES CONSTITUANTS FOCALISES.	100
FIGURE 31. POURCENTAGE D'OCCURRENCE DES DIFFERENTES ETENDUES SYLLABIQUES DU CONTOUR INTONATIF EN PAROLE INTERPRETEE, EN PAROLE LUE ET EN PAROLE SPONTANEE.	101
FIGURE 32. POURCENTAGE D'OCCURRENCE DES CONSTITUANTS FOCALISES PORTANT UN ACCENT INITIAL ET NE PORTANT PAS D'ACCENT INITIAL.	102
FIGURE 33. POURCENTAGE D'OCCURRENCE DE LA PRESENCE D'ACCENT INITIAL SUR LE CONSTITUANT FOCALISE EN PAROLE INTERPRETEE, EN PAROLE LUE ET EN PAROLE SPONTANEE.	103
FIGURE 34. POURCENTAGE D'OCCURRENCE DES DIFFERENTS CONTOURS INTONATIFS PORTES PAR LES CONSTITUANTS FOCALISES DANS LE CORPUS REDUIT.	106
FIGURE 35. COURBES DE FREQUENCE FONDAMENTALE DE DEUX ENONCES ISSUS DU CORPUS DE L'EXPERIENCE SE TERMINANT PAR UN CONTOUR INTONATIF H*H-.	111
FIGURE 36. COURBES DE FREQUENCE FONDAMENTALE DE QUATRE CONTOURS H*L%, H*H% ET L*L% ISSUS DU CORPUS DE L'EXPERIENCE.	112
FIGURE 37. CAPTURE D'ECRAN DU LOGICIEL PRAAT CONTENANT L'ALIGNEMENT TEXTE-PAROLE D'UN ENONCE, AINSI QUE LES ANNOTATIONS DES EXPERTS.	151
FIGURE 38. HAUTEUR MOYENNE DES SYLLABES FOCALISEES POUR CHAQUE FONCTION.	156
FIGURE 39. HAUTEUR MOYENNE DES SYLLABES FOCALISEES SELON LA FONCTION ET LE PHONOGENRE.	157
FIGURE 40. DUREE MOYENNE DES SYLLABES FOCALISEES POUR CHAQUE FONCTION.	159
FIGURE 41. DUREE MOYENNE DES SYLLABES FOCALISEES SELON LA FONCTION ET LE PHONOGENRE.	162

FIGURE 42. POURCENTAGES D’OCCURRENCE DES DIFFERENTS TYPES DE CONTOUR INTONATIF SUR LES CONSTITUANTS FOCALISES POUR CHAQUE FONCTION.	165
FIGURE 43. POURCENTAGES D’OCCURRENCE DES DIFFERENTS TYPES DE FRONTIERE PROSODIQUE SUR LES CONSTITUANTS FOCALISES POUR CHAQUE FONCTION.	168
FIGURE 44. POURCENTAGES D’OCCURRENCE DES DIFFERENTS TYPES D’ETENDUE SYLLABIQUE DU CONTOUR INTONATIF POUR CHAQUE FONCTION.	172
FIGURE 45. POURCENTAGE D’OCCURRENCE DE LA PRESENCE D’ACCENT INITIAL SUR LE CONSTITUANT FOCALISE POUR CHAQUE FONCTION.	174
FIGURE 46. COURBES DE FREQUENCE FONDAMENTALE DE LA VERSION ORIGINALE ET DE LA VERSION REENREGISTREE D’UN DES ENONCES UTILISES DANS L’EXPERIENCE DE PERCEPTION.	190
FIGURE 47. CAPTURE D’ECRAN DU MENU DE MANIPULATION PROSODIQUE DU LOGICIEL PRAAT.	193
FIGURE 48. CAPTURE D’ECRAN DE L’APPLICATION E-STUDIO DU LOGICIEL E-PRIME.	196
FIGURE 49. NOTES STANDARDISEES MOYENNES ATTRIBUEES AUX QUATRE REALISATIONS PROSODIQUES DES ITEMS DE CONTROLE.	198
FIGURE 50. NOTES STANDARDISEES MOYENNES ATTRIBUEES AUX QUATRE REALISATIONS PHONETIQUES (TOUTES FONCTIONS CONFONDUES).	200
FIGURE 51. NOTES STANDARDISEES MOYENNES ATTRIBUEES AUX TROIS FONCTIONS ANALYSEES (TOUTES REALISATIONS CONFONDUES).	202
FIGURE 52. NOTES STANDARDISEES MOYENNES ATTRIBUEES AUX QUATRE REALISATIONS PHONETIQUES POUR CHAQUE FONCTION.	204

Introduction

Lorsque nous nous exprimons au quotidien, il nous arrive souvent de souligner certains mots par notre façon de les prononcer. Ce type de soulignement peut remplir plusieurs fonctions : insister sur une information, la mettre en avant par rapport au reste du contenu, la contraster avec une autre information, ou encore exprimer une émotion. Ces soulignements, qui sont en partie involontaires, contribuent au bon fonctionnement de la communication.

Imaginons maintenant une situation différente : celle d'un acteur interprétant un texte qu'il a appris par cœur. Existe-t-il, entre la parole de cet acteur et la parole ordinaire, des différences concernant les soulignements que nous avons décrits ? Sont-ils plus fréquents dans un cas que dans l'autre ? Présentent-ils des réalisations acoustiques différentes ? De plus, ces réalisations varient-elles selon la fonction remplie par le soulignement ?

Ces questions se trouvent au cœur des recherches que nous présentons dans cette thèse. La *parole interprétée* est le terme que nous proposons pour désigner l'oralisation d'un texte écrit qui a été mémorisé au préalable par le locuteur. Ce type de parole est produit notamment par des acteurs. La parole interprétée se distingue de la parole spontanée par le fait que le locuteur sait à l'avance ce qu'il va dire. Elle se distingue également de la parole lue par le fait que le locuteur a mémorisé le texte. La *focalisation prosodique* est le terme linguistique qui désigne la mise en relief d'un constituant dans un énoncé. La focalisation fait appel à différentes ressources prosodiques : l'accentuation, l'intonation, la formation des groupes prosodiques, ainsi que les variations de tempo et de registre tonal. Les fonctions attribuées à la focalisation se divisent en deux ensembles. Le premier comprend des fonctions de nature sémantico-pragmatique. Il s'agit du marquage des différentes catégories de *focus* (informationnel, contrastif, associatif et verum). Le second, qualifié d'*emphase* dans cette thèse, comprend des fonctions de nature expressive. Il s'agit de la fonction d'insistance, ainsi que l'expression des émotions et des attitudes.

Le choix de notre sujet est motivé par deux raisons. Premièrement, la réalisation prosodique des fonctions attribuées à la focalisation est encore mal connue. Deuxièmement, nous pensons que la focalisation est un phénomène prosodique qui différencie de manière

importante la parole interprétée de la parole spontanée. Comme nous le verrons au cours de cette thèse, les résultats de nos études confirment en partie cette intuition.

Dans cette thèse, la focalisation prosodique est envisagée à la fois en tant que phénomène global de soulignement, ainsi qu'à travers les différentes fonctions que la littérature lui attribue. Deux aspects de la focalisation sont étudiés : sa fréquence d'occurrence et sa réalisation prosodique. Nous souhaitons répondre aux questions suivantes :

- La focalisation présente-t-elle une fréquence d'occurrence supérieure en parole interprétée ?
- La focalisation présente-t-elle des réalisations prosodiques spécifiques en parole interprétée ?
- Les fonctions attribuées à la focalisation présentent-elles des réalisations différentes ?
- Les fonctions de la focalisation présentent-elles des réalisations spécifiques en parole interprétée ?

De manière plus générale, nous souhaitons également connaître les propriétés de la focalisation dans la grammaire du français :

- Quelle est la fréquence d'occurrence de la focalisation en français ?
- Quelle est la réalisation prosodique de la focalisation en français ?

Ces questions présentent un intérêt pour la linguistique et la phonétique à plusieurs titres. La réalisation des fonctions de la focalisation fait l'objet de beaucoup d'études en prosodie. Cependant, ces études s'intéressent généralement à la réalisation d'une seule fonction à la fois. Cette thèse se distingue par le fait qu'elle compare la réalisation de toutes les fonctions de la focalisation. La question de l'existence de réalisations distinctes pour chaque fonction est particulièrement importante. En effet, en l'absence de différences prosodiques, l'attribution de ces fonctions à la focalisation serait remise en cause. Par ailleurs, ce travail s'inscrit dans le cadre des nombreuses recherches en phonostylistique, ou l'étude des différences phonétiques entre les styles de parole. Parmi ces recherches, peu d'études ont été consacrées à la parole interprétée. Enfin, cette thèse présente un apport sur le plan méthodologique. En effet, l'expérience de production et l'expérience de perception que nous avons menées sont relativement novatrices. Dans la première, inspirée du protocole RepTask (cf Laurens, Marandin, Patin et Yoo 2011), des locuteurs ont reproduit des conversations

spontanées en parole lue et en parole interprétée. Le relevé des occurrences de focalisation prosodique, ainsi que la classification fonctionnelle de ces occurrences, ont ensuite été effectués par un groupe d'experts en prosodie. Dans la seconde expérience, des jugements d'acceptabilité ont été obtenus concernant la correspondance entre différentes réalisations phonétiques du même énoncé et le contexte d'occurrence de cet énoncé. Cette expérience nous a permis de tester l'existence de relations entre ces réalisations et les fonctions attribuées à la focalisation.

La thèse est organisée de la manière suivante. Dans le Chapitre 1, nous présentons les concepts prosodiques nécessaires à notre étude de la focalisation et de ses fonctions. Après une introduction globale à la prosodie et aux caractéristiques du français, le chapitre est divisé selon les trois principaux domaines de la prosodie : l'accentuation, la structure prosodique, et l'intonation. Nous décrivons notamment la réalisation des différents types d'accent (rythmique, focal et emphatique) en français, ainsi que celle des différents niveaux de la structure prosodique (groupe accentuel, intonatif et intermédiaire). Nous abordons ensuite les mouvements intonatifs du français, notamment les contours continuatifs et terminaux, ainsi que le phénomène de désaccentuation. La plupart de ces catégories font partie de l'analyse prosodique que nous présentons dans les chapitres suivants.

Dans le Chapitre 2, nous analysons l'influence du phonogène (parole spontanée, lue ou interprétée) sur la fréquence d'occurrence et la réalisation prosodique de la focalisation, envisagée en tant que phénomène global de soulignement. Nous nous intéressons également aux propriétés de la focalisation indépendamment du phonogène. Nous commençons par présenter l'état de l'art concernant la réalisation de la focalisation en français. Un des objectifs de notre étude est de confirmer la présence de certains traits attribués à la focalisation (notamment l'augmentation de hauteur et de durée, ainsi que l'accentuation initiale). Nous présentons également l'état de l'art concernant les différences prosodiques entre parole spontanée, lue et interprétée. Plusieurs différences sont observées, mais aucune étude ne porte explicitement sur la focalisation prosodique, d'où l'intérêt de notre étude.

Nous décrivons ensuite l'expérience de production évoquée plus haut. Les résultats révèlent tout d'abord un taux d'accord relativement bas entre les experts en prosodie concernant le relevé des occurrences de focalisation. Cette faiblesse du taux d'accord réduit la portée des résultats de l'analyse. Cette dernière révèle malgré tout plusieurs résultats originaux. Le

pourcentage de syllabes focalisées dans le corpus est de 11,3%, soit environ un tiers du pourcentage de proéminences en français. La parole interprétée présente le pourcentage le plus élevé, suivie de la parole lue et de la parole spontanée. Ce résultat correspond à notre prédiction, et suggère que la fréquence d'occurrence de la focalisation augmente avec le degré de « préparation » du phonogène. Concernant l'analyse prosodique, on observe tout d'abord que la focalisation est réalisée par une forte augmentation de hauteur et de durée syllabique, des résultats qui confirment la littérature. L'augmentation de durée est plus importante en parole lue qu'en parole interprétée, ce qui suggère que la réalisation de la focalisation au moyen d'une forte augmentation de durée est un trait caractéristique de la lecture à haute voix. Au niveau phonologique, les résultats sont également cohérents avec la littérature. La focalisation est réalisée par une variété de contours intonatifs. De manière majoritaire, ces contours sont de forme montante, commencent par un ton haut, et sont portés par la dernière syllabe du constituant focalisé. Parmi les constituants de plusieurs syllabes, une proportion importante, bien que minoritaire, comporte un accent initial. En revanche, contrairement à notre prédiction, le phonogène se révèle n'avoir aucune influence sur les catégories phonologiques de la focalisation.

Dans le Chapitre 3, une analyse de la réalisation prosodique des fonctions de la focalisation est proposée. Cette analyse prend en compte l'influence du phonogène sur la réalisation. Nous commençons par présenter un état de l'art de la définition théorique et la réalisation prosodique de chaque fonction. Comme nous l'avons expliqué plus haut, bien que plusieurs résultats aient déjà été observés, aucune étude n'a comparé la réalisation de toutes les fonctions à la fois dans le même corpus. De plus, aucune étude ne s'est déjà intéressée, à notre connaissance, à l'influence du phonogène sur la réalisation des fonctions.

Nous décrivons ensuite les deux études que nous avons menées. La première se base sur l'expérience de production décrite dans le chapitre précédent, et fait appel à une classification fonctionnelle des occurrences de focalisation par le groupe d'experts en prosodie. La seconde étude se base sur l'expérience de perception déjà mentionnée plus haut. Les résultats de l'étude de production révèlent tout d'abord un taux d'accord entre experts globalement bas pour toutes les fonctions. Seules trois fonctions présentent un taux d'accord suffisamment haut pour pouvoir être analysées : l'insistance, l'expressivité, et le marquage de focus (en tant que fonction globale). Comme pour l'étude du chapitre précédent, cette faiblesse du taux

d'accord réduit la portée des résultats de l'analyse. Elle soulève également plusieurs questions méthodologiques et théoriques. L'analyse prosodique révèle toutefois plusieurs résultats originaux. Une forte relation est observée entre la fonction d'insistance et l'accentuation initiale. Cette relation est significative à la fois en production et en perception. Ce résultat confirme de nombreuses études précédentes sur l'insistance. Une relation est également observée entre l'insistance et l'augmentation de hauteur, ainsi qu'entre l'expressivité et l'augmentation de durée. Ces derniers résultats sont numériques seulement en production et significatifs en perception. Ceci pourrait être dû aux nombreux facteurs aléatoires de l'expérience de production, au caractère relativement artificiel de l'expérience de perception, ou encore à des variations interindividuelles chez les locuteurs et les auditeurs. Enfin, contrairement à notre prédiction, on n'observe aucune influence du phonogène sur la réalisation des fonctions (un résultat probablement dû à un manque de données).

Plusieurs propositions sont faites pour renforcer les deux expériences. La méthodologie de l'expérience de production peut être améliorée, notamment concernant les étapes de la constitution du corpus et du relevé des occurrences de focalisation. De plus, des traits phonétiques et phonologiques supplémentaires peuvent être pris en compte dans les analyses proposées, notamment celui de la désaccentuation post-focale ou pré-focale. Enfin, les perspectives de notre recherche sont présentées.

Chapitre 1 : La prosodie du français

1. Introduction : qu'est-ce que la prosodie ?

1.1. Traits acoustiques de la prosodie

Le terme de *prosodie* désigne un ensemble de faits acoustiques de la parole humaine. On parle souvent de niveau *suprasegmental* pour décrire la prosodie, c'est-à-dire tout ce qui ne concerne pas les *segments* que sont les consonnes et les voyelles. Concrètement, la prosodie comprend les cinq paramètres suivants (cf Cruttenden 1986) :

- **la hauteur** : Il s'agit de la « mélodie » de la parole que l'auditeur entend. On emploie aussi souvent le terme d'*intonation* (qui peut cependant avoir d'autres sens, comme nous le verrons). La hauteur est en réalité un phénomène perceptif et non acoustique : ce que l'auditeur perçoit ne correspond pas exactement au signal acoustique produit par le locuteur. Le corrélat acoustique de la hauteur est la *fréquence fondamentale* (ou F0) : il s'agit de la mesure de fréquence de vibration des cordes vocales, lors de la production des sons dits *voisés* (les voyelles et une partie des consonnes). Son unité de mesure est le Hertz (Hz), ou le nombre de vibrations par seconde. On l'appelle « fondamentale » car il s'agit de la fréquence la plus basse de la parole (et donc auditivement la plus grave). Les sons voisés peuvent être considérés comme des sons *périodiques complexes*, c'est-à-dire composés de plusieurs sons *simples*. Ils contiennent une multitude d'autres fréquences plus hautes que la F0, que l'on nomme *fréquences harmoniques*.
- **la durée** : La parole contient des variations dans le domaine du temps, en particulier des allongements sur certaines syllabes, ou des variations de *tempo* à l'échelle d'un énoncé entier ou d'un groupe d'énoncés. Le terme de *durée* désigne, parmi ces variations temporelles, celles qui sont perçues par l'auditeur. Son unité est la seconde (s).
- **l'intensité** : Il s'agit des variations dans le volume de la voix. Son unité est le décibel (dB). Comme pour la hauteur, l'intensité est un phénomène perceptif ne correspondant pas à

la réalité acoustique. Le corrélat acoustique de l'intensité est l'*amplitude*, c'est-à-dire la variation dans la pression de l'air qui provient des poumons.

- **la qualité vocalique** : Il s'agit des variations perçues par l'auditeur dans la « couleur » de la voix, telles que le chuchotement, la nasalisation, la « voix craquée » ou encore la « voix de tête » (cf Laver 1980). Les variations de qualité vocalique sont dues à des modifications du dispositif articulatoire glottique (les cordes vocales) et supra-glottique (le conduit vocal et le conduit nasal).
- **les pauses** : Bien qu'on pourrait considérer qu'elles font partie du paramètre de la durée, on les considère souvent comme un paramètre à part, faisant partie intégrante de la prosodie (cf Goldman-Eisler 1968, Clark 1994). Il existe deux types de pauses : les pauses *silencieuses* (sans aucun son de la voix) et les pauses *voisées* (avec vibration des cordes vocales).

1.2. Fonctions de la prosodie

La prosodie sert à remplir un grand nombre de fonctions. On peut compter quatre fonctions principales (cf Chafe 2003) :

- **la délimitation d'unités de parole** : Lorsqu'on écoute de la parole, il apparaît clairement que celle-ci n'est pas un flot sonore ininterrompu, mais se compose de groupes plus ou moins longs. On peut le constater dans l'exemple (1) ci-dessous, un énoncé attesté extrait du corpus de l'expérience de production qui sera présentée au chapitre 2. Les frontières de groupes sont représentées par des barres verticales. C'est la prosodie qui réalise ces frontières, notamment au moyen de pauses, d'allongements et de contours intonatifs montants ou descendants. Cette fonction de la prosodie sera abordée plus en détail ci-dessous dans la partie sur la structure prosodique du français (cf § 3). On voit en (1) que les groupes prosodiques correspondent souvent, mais pas toujours, à des groupes syntaxiques majeurs. La locutrice a par exemple inséré une frontière à l'intérieur du groupe nominal *cours de philosophie indienne*.

(1) donc j'étais à ce cours-là | et c'était assez drôle | parce que tous les étudiants | se sont précipités au cours | de philosophie indienne | parce que c'était plutôt original

- **l'indication des relations entre unités** : Cette fonction découle logiquement de la fonction précédente. En plus de délimiter des unités, la prosodie permet au locuteur d'indiquer la hiérarchie qui existe entre elles : certaines unités sont dépendantes d'autres, et de petites unités s'assemblent pour former des unités plus grandes. Cette hiérarchie prosodique est réalisée en particulier par les contours intonatifs. Dans l'exemple (2) ci-dessous, l'énoncé de l'exemple précédent est repris et sa hiérarchie est représentée. Les unités mineures sont à nouveau délimitées par des barres verticales, et les unités majeures sont délimitées par des parenthèses. On voit que la locutrice a d'abord produit un groupe autonome, *donc j'étais à ce cours-là*. Puis tous les groupes suivants s'assemblent pour former un groupe supérieur, se trouvant au même niveau que le premier dans la hiérarchie prosodique.

(2) (donc j'étais à ce cours-là) (et c'était assez drôle | parce que tous les étudiants | se sont précipités au cours | de philosophie indienne | parce que c'était plutôt original)

- **l'assignation de proéminences** : Cette fonction fera particulièrement l'objet de cette thèse, car elle est justement réalisée par le procédé de focalisation prosodique. Le locuteur éprouve souvent le besoin de mettre en valeur certains éléments de l'énoncé : des mots, des groupes de mots, voire même des syllabes. Il le fait en leur assignant des proéminences, notamment au moyen de l'accentuation. Cette mise en valeur peut avoir plusieurs raisons, comme on peut le voir dans les exemples (3) et (4) ci-dessous (également extraits du corpus de l'expérience de production), où les mots proéminents sont représentés par des majuscules. En (3), le locuteur accentue *voisin* car il s'agit de la seule information nouvelle de l'énoncé, puisque *le maire de Bormes* a déjà été formulé dans l'énoncé précédent. En (4), la locutrice souhaite simplement insister sur *exactement* pour souligner auprès de son interlocuteur l'importance de cette information. Dans le premier cas, la proéminence remplit la fonction de marquage de focus informationnel étroit (cf chap. 3, § 2.1.1), et dans le second cas elle remplit la fonction d'insistance (cf chap. 3, § 3.1.1).

(3) Le maire de Bormes ! C'est le VOISIN, tu vois, le maire de Bormes.

(4) Et les profs, c'est EXACTEMENT l'inverse, quoi.

- **l'expression d'émotions et d'attitudes** : Cette fonction, qui est généralement celle qui vient en premier à l'esprit des non-spécialistes lorsqu'on leur demande à quoi sert la prosodie, est aussi celle à laquelle ont été consacrées le moins d'études. Il semble néanmoins évident que la prosodie remplit des fonctions expressives. Dans ce domaine, on distingue souvent les *émotions* (telles que la joie ou la colère) des *attitudes* (telles que l'ironie ou le doute). Les premières constituent des *indices* (involontaires) de l'état du locuteur tandis que les secondes constituent des *signaux* (volontaires) envoyés à l'auditeur (cf Léon 1993 : 21-22).

A ces fonctions principales s'ajoutent beaucoup d'autres fonctions. La prosodie sert par exemple à indiquer la modalité de l'énoncé : assertion, interrogation, exclamation, injonction... Ce sujet sera abordé en détail plus bas (cf § 4.2). La prosodie peut également informer l'auditeur sur les origines géographiques ou sociales du locuteur. On parle alors d'accent « du Midi », « du Nord » ou « parisien », et d'accent « snob » ou « populaire ». Certaines prosodies peuvent également être associées à des professions ou des activités particulières : journaliste, avocat, allocution politique, lecture à haute voix... Nous aborderons ce domaine dans le chapitre suivant (cf § 3.1). Enfin, selon Fónagy et Bérard (2006), la prosodie remplit des fonctions d'ordre esthétique. Par exemple, lorsqu'un locuteur raconte une anecdote, « *la répétition et la variation de structures mélodiques [...] est comparable au rôle de la répétition et de la variation dans les compositions musicales* ». La parole contient également des « clichés mélodiques » associés à certains énoncés particuliers (ex : « A table ! », « Oh qu'il est mignon ! »).

1.3. Domaines formels de la prosodie

En plus de la décrire selon ses différents paramètres acoustiques et ses différentes fonctions, on divise souvent la prosodie en plusieurs domaines qui constituent chacun un ensemble

cohérent de faits phonologiques et phonétiques. Les trois domaines principaux de la prosodie sont les suivants (cf Frota et Prieto 2015) :

- **l'accentuation** : Il s'agit de l'assignation de différents types d'accents (lexicaux, rythmiques, focaux, emphatiques) à des mots ou à des groupes de mots.
- **le phrasé** (terme que nous adaptons de l'anglais *phrasing*) : Il s'agit de la formation de groupes prosodiques de différents niveaux (groupe accentuel, groupe intermédiaire et groupe intonatif), et par conséquent du marquage de la structure prosodique des énoncés.
- **l'intonation** : Il s'agit de l'ensemble des contours mélodiques de la langue (continuatifs, terminaux, expressifs). Notons que le terme d'intonation est aussi parfois employé pour désigner la prosodie dans sa totalité.

Parmi les autres domaines formels de la prosodie qui sont souvent étudiés, on peut citer les variations de *registre tonal* ou le *rythme*. Le registre tonal, c'est-à-dire l'étendue des variations de hauteur du locuteur, peut s'élever vers l'aigu ou s'abaisser vers le grave, et s'élargir ou se rétrécir. Le rythme, domaine qui recoupe en partie celui de l'accentuation, désigne la récurrence de certains motifs dans la prosodie, par exemple l'alternance entre syllabes proéminentes et syllabes faibles (cf Delais-Roussarie et Yoo 2014).

1.4. Caractéristiques du français

La suite de ce chapitre aborde successivement les domaines de l'accentuation, du phrasé et de l'intonation en français. Dans cette langue, il existe un syncrétisme entre l'accentuation et l'intonation, car les contours intonatifs sont toujours portés par les syllabes accentuées (e.g. Martin 1975, Delais-Roussarie et al 2015). Pour le comprendre, comparons les exemples (5a), (5b) et (5c) ci-dessous. Ces trois exemples contiennent le même énoncé en français, en espagnol et en anglais. Du point de vue intonatif, il s'agit d'une assertion qui se termine, dans chacune des trois langues, par un contour terminal descendant sur la dernière syllabe de l'énoncé. Cependant, du point de vue accentuel, les trois énoncés varient selon la syllabe (soulignée dans les exemples) qui est accentuée dans le dernier mot. Il s'agit de la dernière syllabe en français, de l'avant-dernière syllabe en espagnol, et de l'antépénultième syllabe en

anglais. Par conséquent, tandis que la syllabe accentuée et la syllabe portant le contour intonatif sont séparées en espagnol et en anglais (ce n'est cependant pas toujours le cas), elles sont identiques en français.

- (5) a. Nous allons étudier la prosodie.
 b. Vamos a estudiar prosodia.
 c. We are going to study prosody.

Une autre caractéristique du français est l'absence d'accent lexical dans cette langue. Les accents lexicaux sont des accents qui tombent toujours sur la même syllabe d'un mot donné, et qui sont à ce titre stockés dans le lexique au même titre que les propriétés phonologiques, morphologiques ou sémantiques de ce mot. En (5a), l'accent sur la dernière syllabe de *prosodie* tombe ici parce que ce mot se trouve à la fin du groupe prosodique qui le contient. Par opposition, en (5b) et en (5c), les accents sur la deuxième syllabe de *prosodia* et sur la première syllabe de *prosody* sont des accents lexicaux qui tombent toujours sur ces syllabes, quelle que soit la position de ces mots dans l'énoncé. En espagnol et en anglais, comme dans beaucoup de langues, la même forme peut d'ailleurs avoir plusieurs significations selon la syllabe qui porte l'accent. En espagnol, on peut citer le contraste entre *terminó* (« il/elle a terminé »), *termino* (« je termine ») et *término* (« terme »). En anglais, on trouve plusieurs paires nom-verbe telles que *record* et *record* (« disque » et « enregistrer ») ou *convict* et *convict* (« condamné(e) » et « condamner »).

En plus du syncrétisme entre accentuation et intonation et de l'absence d'accent lexical, le français connaît un faible écart acoustique entre syllabes accentuées et inaccentuées par rapport à d'autres langues (Delais-Roussarie et Di Cristo, à paraître). Cela a pour conséquence que le français a parfois été considéré comme une langue sans accent (e.g. Rossi 1980), voire sans intonation. Cependant, nous allons voir que c'est loin d'être le cas !

2. L'accentuation du français

2.1. Catégories d'accents

Il existe plusieurs catégories d'accents (cf Di Cristo 1999a) :

- **l'accent lexical** : Comme nous venons de le voir, il s'agit de la catégorie, absente du français, qui est distinctive au niveau lexical.
- **l'accent rythmique (ou métrique)** : Cet accent sert à délimiter les groupes prosodiques (et participe à ce titre au rythme de la parole). Il tombe généralement sur la syllabe finale des groupes, et parfois sur une des syllabes initiales. L'exemple (6) ci-dessous reprend l'énoncé que nous avons vu plus haut et indique les accents rythmiques réalisés par la locutrice.

(6) donc j'étais à ce cours-là | et c'était assez drôle | parce que tous les étudiants | se sont précipités | au cours | de philosophie indienne | parce que c'était plutôt original

- **l'accent focal** : Cet accent sert à indiquer un *focus* dans l'énoncé. Les concepts de focus et de ses différents sous-types (*informationnel, contrastif, associatif* et *verum*) ont leur importance dans cette thèse, puisque le marquage de focus est une des fonctions attribuées à la focalisation prosodique. Ces concepts seront définis dans le chapitre 3 (cf § 2.1). Pour l'instant, examinons les exemples (7) à (10) ci-dessous, dans lesquels l'énoncé de l'exemple (3) est repris puis modifié pour illustrer les différents contextes dans lesquels on peut rencontrer un accent focal. En (7), l'accent sur *voisin* marque l'information nouvelle de l'énoncé (focus informationnel). En (8), l'accent sur *La Londe* marque un contraste avec *Bormes* dans l'énoncé précédent (focus contrastif). En (9), l'accent sur *Bormes* marque le fait qu'il s'agit de l'argument de l'adverbe *aussi* (focus associatif). En (10), l'accent sur *c'est* sert à souligner la vérité de la proposition (verum focus).

(7) Le maire de Bormes ! C'est le voisin, le maire de Bormes...

(8) – C'est le maire de Bormes.

– Non, c'est le maire de La Londe.

(9) J'ai aussi parlé au maire de Bormes.

(10) – Ce n'est pas le maire de Bormes.

– Si, c'est le maire de Bormes.

- **l'accent emphatique** : Comme son nom l'indique, cet accent sert à marquer une *emphase* sur un élément, une fonction qui est elle aussi attribuée au phénomène de focalisation prosodique. Dans le chapitre 3 (cf § 3.1), nous distinguerons deux types d'emphase : *l'insistance* et *l'expressivité*. L'insistance a déjà été illustrée dans l'exemple (4), repris ci-dessous en (11). L'expressivité est illustrée par l'exemple (12) ci-dessous. En (11), la locutrice insiste sur l'importance de *exactement*. En (12), l'accent sur *hurlé* lui sert à exprimer une émotion (probablement la colère). On remarque que, contrairement aux exemples précédents, les accents sont ici initiaux et non finaux.

(11) Et les profs, c'est exactement l'inverse, quoi.

(12) Et il lui a hurlé dessus !

Nous allons voir que ces différents accents connaissent des réalisations acoustiques différentes. Mais auparavant, il est nécessaire de passer en revue certaines unités et certains principes de l'accentuation du français.

2.2. Unités métriques

Le rythme concerne le niveau superficiel de la langue et renvoie donc à la réalisation prosodique effective des énoncés. Le fonctionnement de l'accentuation repose également sur un domaine sous-jacent qui renvoie à la réalisation potentielle des énoncés : il s'agit du *mètre*. En français comme dans d'autres langues, ce domaine comprend trois unités de base (cf Di Cristo 1999a) :

- **le pied** : Il s'agit de l'unité métrique minimale. Le pied ne peut contenir qu'un seul accent. En français, l'accent tombe toujours à la fin du pied, contrairement à l'anglais où il tombe toujours au début du pied (Di Cristo 2013 : 135). Le pied est donc constitué de la syllabe accentuée et des syllabes qui la précèdent. L'exemple (13) ci-dessous reprend l'énoncé attesté que nous avons déjà vu plusieurs fois plus haut. Un découpage potentiel en pieds y est indiqué à l'aide de parenthèses.

(13) (donc j'étais) (à ce cours-là) (et c'était assez drôle) (parce que tous) (les étudiants) (se sont pré-) (cipités au cours) (de phi-) (losophie indienne) (parce que c'était plutôt) (original)

- **le mot prosodique** : Cette unité domine le pied et constitue le domaine de *projection* de l'accent final. Tandis que le pied peut se terminer par un accent tombant au début d'un mot, le mot prosodique se termine forcément par un accent tombant à la fin d'un mot. Le mot prosodique contient obligatoirement un mot appartenant à la classe des mots dits *accentuables* (cf Mertens 1993). Il s'agit de tous les mots lexicaux, de certains pronoms (ex : *tout, chacun*, et les pronoms forts tels que *moi, toi* ou *eux*), de certaines prépositions polysyllabiques (ex : *pendant*) et de certains auxiliaires polysyllabiques (ex : *avait, étions*). La représentation sous-jacente des mots accentuables prévoit qu'ils peuvent recevoir un accent final et, s'ils comportent plusieurs syllabes, un accent initial. On parle de principe de *bipolarisation* (Di Cristo 1999a). L'accent final porte sur la dernière syllabe du mot (ou éventuellement sur l'avant-dernière syllabe si la dernière syllabe contient un schwa, comme dans *étable*). Par ailleurs, d'après un principe dit de *subordination accentuelle* (ibid.), l'accent final est toujours plus fort que l'accent initial. On dit que l'accent final est *primaire* et l'accent initial *secondaire*. Ces deux principes sont illustrés par les exemples (14a) à (14c) ci-dessous. Les accents primaires sont soulignés et en gras, tandis que les accents secondaires sont uniquement soulignés. Ces principes peuvent aussi s'appliquer à des unités signifiantes plus larges qu'un mot, telles qu'un mot composé, un groupe syntaxique, voire même une phrase, comme les exemples (15a) à (15c) le montrent¹. Dans l'exemple (16), l'énoncé attesté est repris et un découpage potentiel en mots prosodiques est proposé à l'aide de crochets. Il faut noter que ce découpage n'est pas celui qui a

¹ Les mots non-accentuables peuvent en fait eux aussi recevoir un accent, dans deux cas de figure. Le premier cas est lorsqu'un mot prosodique commence par un mot grammatical. Celui-ci peut alors recevoir un accent initial, comme en (i) et (ii) ci-dessous. Le second cas est lorsqu'un mot prosodique se termine par un clitique. Celui-ci porte alors l'accent final primaire, comme en (iii).

- (i) se sont précipités
- (ii) de philosophie indienne
- (iii) à ce cours-là

vraiment été réalisé par la locutrice. Nous verrons ce dernier dans la partie sur la structure prosodique (cf § 3), où l'équivalent du mot prosodique au niveau superficiel de la langue (le *groupe accentuel*) sera présenté.

(14) a. drôle

b. étudiants

c. précipités

(15) a. philosophie indienne

b. plutôt original

c. c'était plutôt original

(16) [donc j'étais] [à ce cours-là] [et c'était assez drôle] [parce que tous] [les étudiants] [se sont précipités au cours] [de philosophie indienne] [parce que c'était plutôt] [original]

- **l'unité intonative** : Cette unité, la plus haute dans la hiérarchie métrique, est assemblée à partir de mots prosodiques (elle peut aussi ne contenir qu'un seul mot prosodique). L'unité intonative se termine par un accent dit *nucléaire*, qui est plus fort que l'accent final de mot prosodique. Dans l'exemple (17) ci-dessous, un découpage potentiel en unités intonatives est proposé pour l'énoncé précédent. Les unités intonatives sont délimitées par des crochets et les accents nucléaires sont indiqués en gras et en petites majuscules. Ce découpage est, comme celui en mots prosodiques, différent de celui qui a été effectivement réalisé par la locutrice. Nous verrons ce dernier plus bas (cf § 3.3), lorsque l'équivalent rythmique de l'unité intonative (le *groupe intonatif*) sera présenté.

(17) [donc j'étais à ce cours-LA] [et c'était assez drôle parce que tous les étudiants se sont précipités au cours de philosophie indienne] [parce que c'était plutôt original]

2.3. Contraintes rythmiques

Les unités métriques sont uniquement des *gabarits* accentuels qui déterminent de manière sous-jacente la façon dont les énoncés peuvent être accentués. Au niveau rythmique,

l'accentuation effective des énoncés doit respecter plusieurs contraintes, qui peuvent parfois modifier les unités métriques :

- **la contrainte d'évitement des collisions accentuelles** : Cette contrainte désigne le fait que l'accentuation de deux syllabes adjacentes tend à être évitée (Martin 1987, Di Cristo 1999a, Rossi 1999 : 178, Post 2000). Deux cas de figure sont possibles. Si les deux syllabes sont séparées par une frontière de groupe prosodique, alors la contrainte ne s'applique pas. C'est le cas pour *là* et *et* dans l'exemple (18) ci-dessous. En revanche, si elles se trouvent à l'intérieur du même mot prosodique, une solution doit être trouvée. Les exemples (19a) à (19d) illustrent le fait que, alors que le mot accentuable *assez* reçoit un accent sur sa seconde syllabe lorsqu'il est prononcé en isolation, cela ne peut pas être le cas s'il est suivi d'une syllabe accentuée. Soit l'accent sur *assez* est supprimé, comme en (19c), soit il est déplacé sur la syllabe initiale, comme en (19d).

(18) [donc j'étais à ce cours-là] [et c'était assez drôle]

(19) a. assez

b. * assez drôle

c. assez drôle

d. assez drôle

- **la contrainte d'évitement des vides accentuels** : Inversement, un trop grand nombre de syllabes inaccentuées tend également à être évité (Rossi 1999 : 194, Martin 2015 : 98, Delais-Roussarie et Di Cristo, à paraître). Des séquences syllabiques comme (20a) sont en effet difficilement prononçables, ne serait-ce que pour des raisons physiologiques. La contrainte est respectée en assignant un ou plusieurs accents supplémentaires sur la séquence concernée, comme en (20b) et (20c).

(20) a. * parce que tous les étudiants se sont précipités

b. parce que tous les étudiants se sont précipités

c. parce que tous les étudiants se sont précipités

- **la contrainte d'équilibre rythmique (ou eurhythmie)** : D'après cette contrainte, les groupes prosodiques successifs tendent à être équilibrés en durée (Martin 1987, Di Cristo 1999a, Rossi 1999 : 178). Deux stratégies sont possibles pour respecter la contrainte. La première stratégie consiste à former des groupes ayant des nombres de syllabes assez proches, comme en (21) ci-dessous. La deuxième stratégie consiste à ralentir le débit de parole sur les groupes ayant un faible nombre de syllabes, comme *donc* et *parce que* dans les exemples (22a) et (22b).

(21) [donc j'étais à ce cours-là] [et c'était assez drôle] [parce que tous les étudiants] [se sont précipités] [au cours de philosophie indienne] [parce que c'était plutôt original]

(22) a. [donc] [j'étais à ce cours-là]

b. [parce que] [c'était plutôt original]

2.4. Réalisation des accents

Les différents types d'accents du français que nous avons passés en revue plus haut (rythmiques, focaux et emphatiques) connaissent des réalisations phonétiques différentes. Il existe également des différences de réalisation selon la position, initiale ou finale, de l'accent sur le constituant accentué. Nous regroupons toutes ces différences ci-dessous (cf Tableau 1) en présentant les résultats de certaines études (Rossi 1999, Astésano 2001, De Looze 2010, Delais-Roussarie et Di Cristo, à paraître) pour chaque accent selon sa catégorie fonctionnelle et sa position. Les résultats sont notamment tirés de l'étude instrumentale détaillée menée par Astésano (2001) sur un large corpus comprenant trois styles de parole (lecture à haute voix, interview et journal radiophonique). Cette étude porte sur les principaux paramètres prosodiques (hauteur, durée et intensité) et aussi sur la « constituance syllabique et infrasyllabique » et les « configurations tonales » des accents. La constituance syllabique désigne la durée de la syllabe et de ses éléments (attaque, noyau et coda). La configuration tonale désigne la forme du pic de hauteur de l'accent, c'est-à-dire son empan (étendue en nombre de syllabes, cf Figure 1) et le caractère symétrique ou asymétrique de son profil (cf Figure 2).

	accent initial	accent final
accent rythmique	<ul style="list-style-type: none"> - augmentation de hauteur et de durée - attaque plus allongée que la rime - pic de hauteur asymétrique (pente ascendante plus abrupte) - mouvement mélodique montant ancré sur la syllabe accentuée 	<ul style="list-style-type: none"> - augmentation de hauteur et de durée - attaque moins allongée que la rime - pic de hauteur symétrique
accent focal	<ul style="list-style-type: none"> - augmentation de hauteur et de durée - allongement de la syllabe supérieur à celui de l'accent rythmique - attaque plus allongée que la rime - étendue syllabique du pic de hauteur moins importante que pour l'accent rythmique, pic plus saillant - pic de hauteur asymétrique (pente ascendante plus abrupte) 	<ul style="list-style-type: none"> - augmentation de hauteur et de durée - allongement de la syllabe supérieur à celui de l'accent rythmique - attaque moins allongée que la rime - étendue syllabique du pic de hauteur moins importante que pour l'accent rythmique, pic plus saillant - pic de hauteur symétrique
accent emphatique	<ul style="list-style-type: none"> - augmentation de hauteur, de durée et d'intensité - allongement de la syllabe supérieur à celui de l'accent rythmique - attaque plus allongée que la rime - augmentation significative de la durée de la consonne (si présente) - coup de glotte si la syllabe commence par une voyelle - étendue syllabique du pic de hauteur moins importante que pour l'accent rythmique, pic plus saillant - pic de hauteur asymétrique (pente ascendante plus abrupte) 	<ul style="list-style-type: none"> - augmentation de hauteur, de durée et d'intensité - allongement de la syllabe supérieur à celui de l'accent rythmique - attaque moins allongée que la rime - étendue syllabique du pic de hauteur moins importante que pour l'accent rythmique, pic plus saillant - pic de hauteur symétrique

Tableau 1. Caractéristiques phonétiques des accents du français selon leur catégorie fonctionnelle et leur position sur le constituant accentué.



Figure 1. Variation de l’empan syllabique du pic de hauteur des accents (cf Astésano 2001). Le pic de l’accent de droite s’étend sur moins de syllabes que le pic de l’accent de gauche et est par conséquent plus saillant.

Figure 2. Variation du profil du pic de hauteur des accents (cf Astésano 2001). Le pic de l’accent de gauche est symétrique tandis que le pic de l’accent de droite est asymétrique (sa pente ascendante est plus abrupte que sa pente descendante).

On voit dans le Tableau 1 que l’accent emphatique se distingue des deux autres catégories d’accent par le fait qu’il met en jeu une augmentation d’intensité en plus de l’augmentation de hauteur et de durée. L’accent rythmique se distingue quant à lui par un allongement de la syllabe moins important, et par le fait que son pic de hauteur s’étend sur un nombre de syllabes plus important, ce qui a pour effet de le rendre moins « saillant ». Entre les accents initiaux et finaux, on note que l’attaque est plus allongée que la rime chez les accents initiaux, alors que c’est l’inverse chez les accents finaux. Les accents initiaux présentent par ailleurs un pic de hauteur asymétrique (la pente ascendante du pic est plus abrupte que la pente descendante) alors que le pic est symétrique chez les accents finaux.

3. La structure prosodique du français

Nous avons vu plus haut (cf § 2.1) qu’une des catégories accentuelles, celle de l’accent rythmique, a pour fonction de délimiter les groupes prosodiques. Nous avons également vu (cf § 2.2) qu’au niveau métrique sous-jacent, le mot prosodique et l’unité intonative ne sont pas marqués par le même accent final : celui de l’unité intonative, dit accent nucléaire, est plus fort que celui du mot prosodique. Le domaine de l’accentuation est donc fortement relié au domaine que nous allons aborder maintenant : le phrasé, ou le marquage de la structure prosodique de surface des énoncés. Nous allons voir que le domaine de l’intonation joue lui aussi un rôle important dans ce marquage. Néanmoins, le phrasé présente beaucoup de caractéristiques qui justifient qu’on l’étudie comme un domaine à part.

3.1. Définition de la structure prosodique

Le concept de structure prosodique est introduit par Martin (1975). Il est ensuite développé notamment par Selkirk (1978, 1984) et Nespors et Vogel (1986), dont l’approche se situe dans

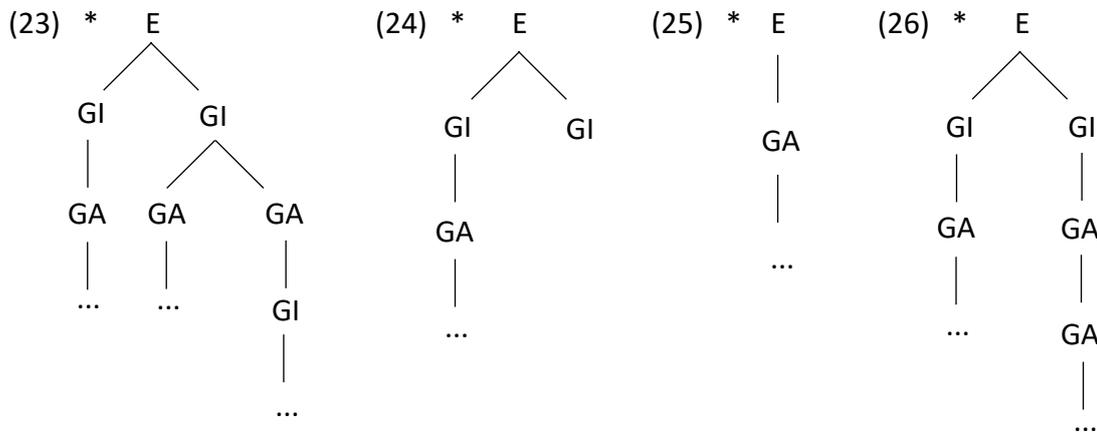
le cadre de la grammaire générative (notamment Chomsky et Halle 1968). Dans cette seconde approche, la représentation phonologique d'un énoncé est décrite comme une hiérarchie de constituants se trouvant à l'interface entre les modules syntaxique et phonologique du langage (cf Delais-Roussarie et al 2011). L'énoncé est le constituant le plus haut de cette hiérarchie, et la syllabe le constituant le plus bas. Entre l'énoncé et la syllabe, un rôle important est joué par le *groupe accentuel* et par le constituant qui le domine directement, le *groupe intonatif*. Comme nous l'avons vu plus haut (cf § 2.2), ces deux constituants sont les équivalents respectifs, à la surface de l'énoncé, des unités sous-jacentes que sont le mot prosodique et l'unité intonative. Le groupe accentuel et le groupe intonatif sont définis à partir d'informations syntaxiques. Dans une version ultérieure de la théorie (e.g. Beckman et Pierrehumbert 1986, Jun 1998), ils sont également définis à partir d'informations prosodiques, notamment les tons (hauts ou bas) de l'énoncé.

3.2. Règles de la structure prosodique

Dans la section suivante, nous allons passer en revue les propriétés syntaxiques et prosodiques des différents niveaux de la structure prosodique. Auparavant, il est nécessaire de mentionner quelques règles que doit respecter la structure prosodique pour être bien formée :

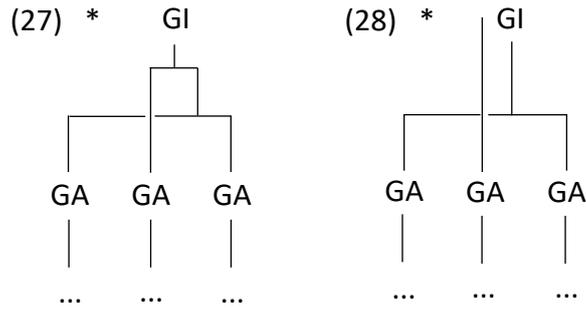
- Selkirk (1984) formule les quatre principes suivants (*strict layer hypothesis*) :
 - **principe d'étagement** : Un constituant de la hiérarchie prosodique ne peut dominer un constituant de niveau supérieur. Dans l'exemple (23) ci-dessous, la structure prosodique de l'énoncé (noté « E ») est mal formée car un groupe accentuel (noté « GA ») domine un groupe intonatif (noté « GI »).
 - **principe de dominance** : Tout constituant doit dominer un constituant de niveau inférieur. Dans l'exemple (24), la structure est mal formée car un groupe intonatif ne domine pas de constituant inférieur.

- **principe d'exhaustivité** : Un constituant ne peut dominer immédiatement un constituant se trouvant à plus d'un niveau en-dessous de lui. Dans l'exemple (25), la structure est mal formée car l'énoncé domine immédiatement un groupe accentuel.
- **principe de non-récurtivité** : Un constituant ne peut dominer un constituant de même niveau que lui. Dans l'exemple (26), la structure est mal formée car un groupe accentuel domine un autre groupe accentuel.



- Martin (1987) formule quant à lui les deux contraintes suivantes :

- **contrainte de planarité** : Un groupe accentuel ne peut entrer dans une relation de dépendance avec un groupe accentuel qui ne le suit pas directement dans l'énoncé. Une telle relation ne pourrait en effet pas être réalisée prosodiquement. Dans l'exemple (27) ci-dessous, la structure est mal formée car le deuxième groupe accentuel passe « par-dessus » la relation entre les deux autres groupes accentuels.
- **contrainte de connexité** : Tout groupe accentuel doit entrer dans une relation de dépendance avec un groupe accentuel se trouvant à sa suite. Dans l'exemple (28), la structure est mal formée car le deuxième groupe accentuel est indépendant du reste de la structure.



3.3. Niveaux de la structure prosodique

Nous présentons ci-dessous (cf Tableau 2) les propriétés syntaxiques, accentuelles et intonatives du groupe accentuel et du groupe intonatif en français (cf Di Cristo et Hirst 1993, Di Cristo 1999a, Post 2000, Post 2002, Jun et Fougeron 2000, Delais-Roussarie et al 2011, Delais-Roussarie et al 2015). Concernant l’intonation, nous adoptons le système de transcription ToBI (*Tones and Break Indices*, cf Silverman et al 1992) qui est fréquemment utilisé dans les recherches sur la prosodie. Nous nous basons plus précisément sur la dernière version de ToBI adapté pour le français (appelée F-ToBI) qui a été proposée par Delais-Roussarie et al (2015). Pour rappel, dans ToBI, les mouvements intonatifs sont représentés par des tons hauts (H) ou bas (L).

	groupe accentuel	groupe intonatif
propriétés syntaxiques	<ul style="list-style-type: none"> - correspond généralement à un groupe nominal, verbal, adjectival ou adverbial - contient généralement un mot lexical et tous les mots grammaticaux qu'il gouverne - taille moyenne : entre 2,3 et 2,6 mots et entre 3,5 et 3,9 syllabes (varie selon le style de parole et la vitesse d'élocution) 	<ul style="list-style-type: none"> - correspond généralement à une phrase racine ou à une proposition subordonnée, coordonnée ou juxtaposée - exceptions : groupes nominaux ou prépositionnels n'entrant pas dans la bonne formation de la phrase noyau (ex : ajouts, éléments disloqués, termes d'adresse) - taille moyenne : entre 2,0 et 2,7 groupes accentuels et entre 7,3 et 10,1 syllabes
propriétés accentuelles	<ul style="list-style-type: none"> - délimité par un accent final obligatoire et un accent initial optionnel - accent final : allongement syllabique - accent initial : affecte la première ou la deuxième syllabe du premier mot ou du premier mot lexical 	<ul style="list-style-type: none"> - délimité par un accent nucléaire final plus fort que l'accent final du groupe accentuel, avec un allongement syllabique plus important, et éventuellement suivi d'une pause
propriétés intonatives	<ul style="list-style-type: none"> - forme tonale par défaut : / (L) (Hi) (L) H* /, où les trois premiers tons sont optionnels et où H* est un accent mélodique démarcatif² - six réalisations phonétiques³ : <ul style="list-style-type: none"> - [L Hi L H*]  - [L H*]  - [L L H*]  - [L Hi H*]  - [Hi L H*]  - [L Hi L*]  	<ul style="list-style-type: none"> - se termine par un ton de frontière L% ou H%, précédé d'un accent mélodique L*, H* ou H+H* (montée sur la syllabe pénultième puis redescende) - cinq contours de fin de groupe : <ul style="list-style-type: none"> - [L* L%]  - [L* H%]  - [H* H%]  - [L H* L%]  - [H+H* H%] 

Tableau 2. Propriétés syntaxiques et prosodiques du groupe accentuel et du groupe intonatif.

² Le premier ton L affecte la ou les syllabes précédant la syllabe portant l'accent initial. Il est absent si l'accent initial tombe sur la première syllabe du groupe. La présence du ton haut Hi dépend de plusieurs facteurs dont le rythme, le style de parole et le locuteur. Le ton Hi affecte la syllabe recevant l'accent initial. Le second ton L affecte la syllabe pénultième (surtout sur les mots courts) ou bien la syllabe portant le ton H* (surtout sur les mots longs). Sa présence est plus fréquente lorsque le groupe accentuel contient un accent initial.

³ La dernière réalisation est due au fait que le ton final H* peut en fait être réalisé comme un ton bas L* lorsque le groupe accentuel se trouve en fin de groupe intonatif, ou bien en milieu de groupe intonatif dans le cadre du principe de contraste de pente (cf § 4.1).

Dans l'exemple (29) ci-dessous, les frontières de groupes accentuels et intonatifs sont indiquées (en rouge et noir respectivement) dans l'énoncé attesté que nous avons déjà analysé plusieurs fois plus haut, ainsi que dans les énoncés qui le suivent dans l'enregistrement de l'expérience de production. Tous les accents rythmiques initiaux et finaux sont également indiqués, ainsi que les contours intonatifs des groupes (en se basant sur les cases du bas du Tableau 2). On voit que la plupart des groupes intonatifs contiennent plusieurs groupes accentuels, sauf trois qui n'en contiennent qu'un seul (*donc j'étais à ce cours-là, en plus, et j'étais assez étonnée*). Tous les groupes intonatifs portent un contour montant, sauf un qui se termine par un contour montant-descendant (*et le prof était tellement horrible*).

(29) $[Hi\ L\ H^*\ H\%]$ $[L\ L\ H^*]$
 [[donc j'étais à ce cours-là]_{GA}]_{GI} [[et c'était assez drôle]_{GA}

$[L\ H^*]$ $[L\ H^*]$ $[L\ H^*]$
 [parce que tous les étudiants]_{GA} [se sont précipités]_{GA} [au cours]_{GA}

$[L\ Hi\ L\ H^*]$ $[L\ L\ H^*\ H\%]$ $[L\ H^*]$
 [de philosophie indienne]_{GA} [parce que c'était plutôt original]_{GA}]_{GI} [[et le prof]_{GA}

$[L\ L\ H^*]$ $[L\ L\ H^*\ L\%]$ $[L\ H^*\ H\%]$ $[L\ H^*]$
 [était tellement]_{GA} [horrible]_{GA}]_{GI} [[en plus]_{GA}]_{GI} [[on lui avait prévu]_{GA}

$[Hi\ L\ H^*\ H\%]$ $[L\ H^*]$ $[L\ L\ H^*\ H\%]$
 [pour son séminaire]_{GA}]_{GI} [[on lui avait prévu une]_{GA} [séminaire de M1]_{GA}]_{GI}

$[L\ Hi\ L\ H^*\ H\%]$ $[L\ L\ H^*]$
 [[une toute petite salle]_{GA}]_{GI} [[ce qui était étonnant]_{GA}

$[L\ H^*\ H\%]$ $[L\ L\ H^*]$ $[L\ L\ H^*\ H\%]$
 [parce que y'avait énormément d'étudiants]_{GA}]_{GI} [[à Paris IV]_{GA} [en philo]_{GA}]_{GI}

$[L\ H^*]$ $[L\ Hi\ L\ H^*\ H\%]$ $[L\ H^*]$ $[L\ H^*]$
 [[et la salle]_{GA} [était toute petite]_{GA}]_{GI} [[et elle était blindée]_{GA} [à craquer]_{GA}

$[L\ Hi\ H^*\ H\%]$ $[L\ H^*\ H\%]$
 [le premier jour]_{GA}]_{GI} [[et j'étais assez étonnée]_{GA}]_{GI}

Plusieurs auteurs (e.g. Jun et Fougeron 2000, Michéas 2011, Delais-Roussarie et al 2015) incluent un troisième niveau dans la structure prosodique : le *groupe intermédiaire*. Comme son nom l'indique, ce groupe se situe à un niveau intermédiaire entre le groupe accentuel et le groupe intonatif. Il contient un ou plusieurs groupes accentuels, et un groupe intonatif peut contenir un ou plusieurs groupes intermédiaires. Comme les deux autres niveaux de la structure prosodique, le groupe intermédiaire possède des propriétés syntaxiques et prosodiques spécifiques (cf Delais-Roussarie et al 2015) :

- **propriétés syntaxiques** : Certaines constructions syntaxiques favorisent la formation d'un groupe intermédiaire, comme l'illustrent les exemples (30) à (35) ci-dessous (les groupes intermédiaires sont indiqués en rouge et sont notés « gi » en minuscules, pour les différencier des groupes intonatifs). Il peut s'agir d'un sujet ou d'un objet relativement long comme en (30) et (31), d'un élément disloqué comme en (32), ou bien d'un élément non-final d'une énumération ou d'une disjonction comme en (33) et (34).

(30) [[[les étudiants]_{GA} [du cours]_{GA} [de philosophie indienne]_{GA}]_{gi} [se sont précipités]_{GA}]_{GI}

(31) [[ils se sont précipités]_{GA} [[au cours]_{GA} [de philosophie indienne]_{GA}]_{gi}]_{GI}

(32) [[[les étudiants]_{GA}]_{gi} [ils se sont précipités]_{GA}]_{GI}

(33) [[[le premier étudiant]_{GA}]_{gi} [[le deuxième étudiant]_{GA}]_{gi} [et le troisième étudiant]_{GA}]_{GI}

(34) [[est-ce un cours]_{GA} [[de philosophie indienne]_{GA}]_{gi} [ou de philosophie orientale]_{GA}]_{GI}

- **propriétés accentuelles** : L'accent final du groupe intermédiaire est marqué par un allongement syllabique plus important que celui du groupe accentuel, mais moins important que celui du groupe intonatif.
- **propriétés intonatives** : Le groupe intermédiaire se termine généralement par un accent phrastique H-, c'est-à-dire un contour montant moins ample que celui du groupe intonatif. Il peut se terminer par un contour descendant L- dans certains cas, par exemple lorsqu'il délimite un élément non-final d'une énumération comme en (33). Lorsqu'un énoncé

contient un marquage de focus informationnel étroit, comme dans l'exemple (7) vu plus haut, le constituant recevant la focalisation prosodique se termine également par un contour descendant de groupe intermédiaire. Enfin, le contour peut être descendant lorsque s'applique le principe de contraste de pente, que nous exposons dans la partie suivante (cf § 4.1).

4. L'intonation du français

La partie précédente sur la structure prosodique du français nous a déjà donné l'occasion d'aborder le domaine de l'intonation puisque, comme nous l'avons vu, les trois niveaux de la structure prosodique ont chacun leurs propriétés intonatives distinctives. Dans la section qui suit (cf § 4.1), nous allons voir plus précisément comment, au moyen des contours continuatifs, l'intonation contribue à indiquer la structure prosodique de l'énoncé. Ce n'est cependant pas la seule fonction de l'intonation : celle-ci sert aussi à indiquer la modalité de l'énoncé, joue un rôle dans l'assignation de proéminences à travers le phénomène de désaccentuation, et permet d'exprimer les émotions et les attitudes.

4.1. Contours continuatifs

Comme nous l'avons vu dans la partie précédente, les groupes accentuels et les groupes intermédiaires se terminent généralement par des contours montants. C'est également souvent le cas pour les groupes intonatifs lorsqu'ils se trouvent en milieu d'énoncé (Post 2000, Jun et Fougeron 2000). En plus de contribuer (avec l'accentuation) à délimiter les groupes prosodiques, ces contours montants que l'on dit *continuatifs* remplissent deux fonctions. Premièrement, ils permettent au locuteur de faire savoir à l'auditeur que l'énoncé n'est pas terminé et qu'il faut attendre une suite. Deuxièmement, ils indiquent les relations de dépendance entre les groupes prosodiques. En effet, la présence d'un contour continuatif à la fin d'un groupe signifie que ce groupe est relié au groupe qui suit. Ce processus est illustré ci-dessous (cf

Figure 3) par la courbe de fréquence fondamentale de notre énoncé attesté, obtenue à l'aide du logiciel d'analyse acoustique Praat (cf Boersma 2002). Les contours montants à la fin de chaque groupe accentuel (sur les mots *cours-là, drôle, étudiants, précipités, cours, indienne* et

original) indiquent à l'auditeur que le groupe auquel ils appartiennent dépend de ce qui suit pour que l'énoncé ait un sens.

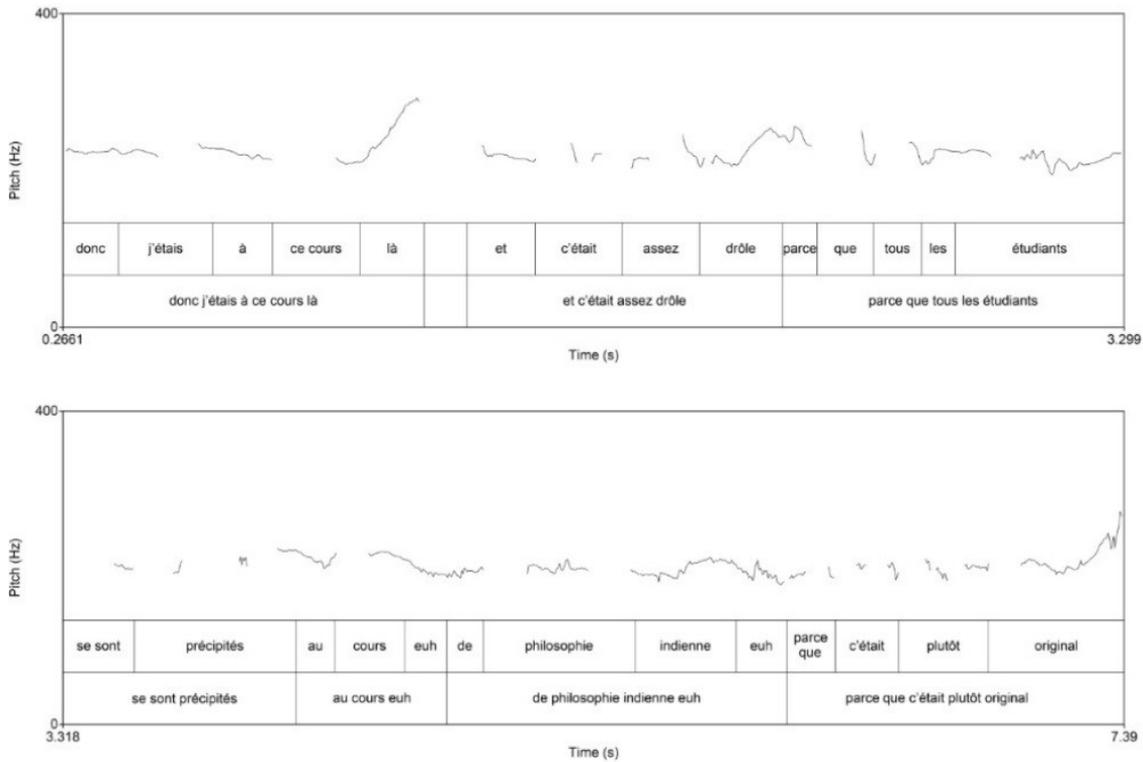


Figure 3. Courbe de fréquence fondamentale (en Hertz) d'un énoncé issu du corpus de l'expérience de production. En bas de la courbe, deux « tires » contiennent le découpage de l'énoncé en mots et en groupes accentuels.

Dans ces relations de dépendance, on voit que certains contours montants sont plus « amples » que les autres. C'est le cas des contours sur *cours-là* et sur *original* dans la Figure 3. Pour expliquer cela, Delattre (1966), distingue deux types de continuation : la continuation *mineure* et la continuation *majeure*. Lorsqu'un ou plusieurs groupes portant un contour de continuation mineure sont suivis par un groupe portant un contour de continuation majeure, tous ces groupes sont reliés ensemble en un grand groupe qui dépend de la suite de l'énoncé. Ce processus est illustré ci-dessous (cf Figure 4). Le contour de continuation mineure relie le groupe *si ces œufs* au groupe suivant *étaient frais*, et le contour de continuation majeure relie ces deux groupes au groupe suivant *j'en prendrais* (qui porte un contour déclaratif descendant). Cette hiérarchie entre deux types de continuation correspond en grande partie à la hiérarchie entre groupe accentuel, groupe intermédiaire et groupe intonatif. Un contour de continuation majeure marque généralement la fin d'un groupe intonatif ou intermédiaire,

tandis qu'un contour de continuation mineure marque la fin d'un groupe accentuel. C'est le cas dans la Figure 3 : les contours de continuation majeure sur *cours-là* et sur *original* marquent la fin d'un groupe intonatif, comme nous l'avons vu plus haut dans l'exemple (29).

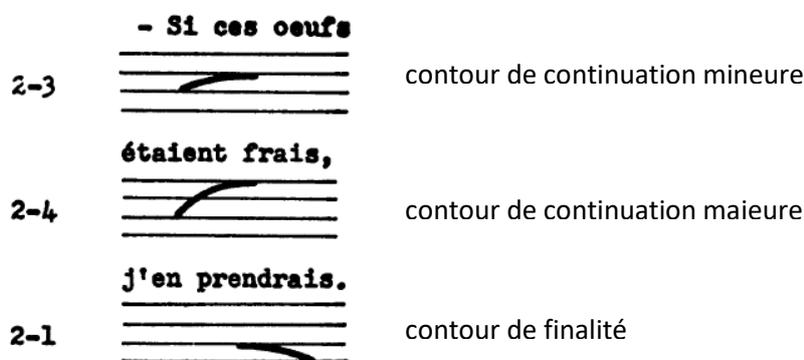


Figure 4. Illustration par Delattre (1966) de la distinction entre continuation mineure et majeure. Les contours sont représentés sur une portée de type musical à quatre niveaux. Les deux contours continuatifs partent du niveau 2 mais le contour de continuation majeure va jusqu'au niveau 4 tandis que celui de continuation mineure ne va que jusqu'au niveau 3.

D'après Martin (1975, 1987, 2009 : 108), certains contours continuatifs peuvent être descendants, en vertu d'un mécanisme qu'il nomme « contraste de pente ». Le contraste de pente indique une relation de dépendance entre deux unités. Par exemple, une unité se terminant par un contour descendant est dépendante d'une unité suivante se terminant par un contour montant. Ce mécanisme a lieu à la fois au niveau du groupe accentuel et au niveau du groupe intonatif. Une illustration est proposée ci-dessous (cf Figure 5). On voit, dans la structure prosodique au-dessus de l'énoncé, que ce dernier comprend deux groupes intonatifs, chacun composé de deux groupes accentuels. Comme l'énoncé est une assertion, le deuxième groupe intonatif se termine par un contour descendant sur *associations*. Afin de respecter le principe de contraste de pente, le premier groupe intonatif se termine par un contour montant sur *Marie-Ségolène*. Le contraste de pente est également respecté à l'intérieur de chaque groupe intonatif : le premier groupe accentuel *les éléphants* porte un contour descendant (puisque le contour qui suit est montant), et le troisième groupe accentuel *ont été regroupés* porte un contour montant (puisque le contour qui suit est descendant). Si l'énoncé était une question et se terminait par un contour montant, tout serait inversé : afin de respecter le contraste de pente, il y aurait un contour descendant sur

regroupés, un contour descendant sur *Marie-Ségolène* et un contour montant sur *éléphanteaux*.

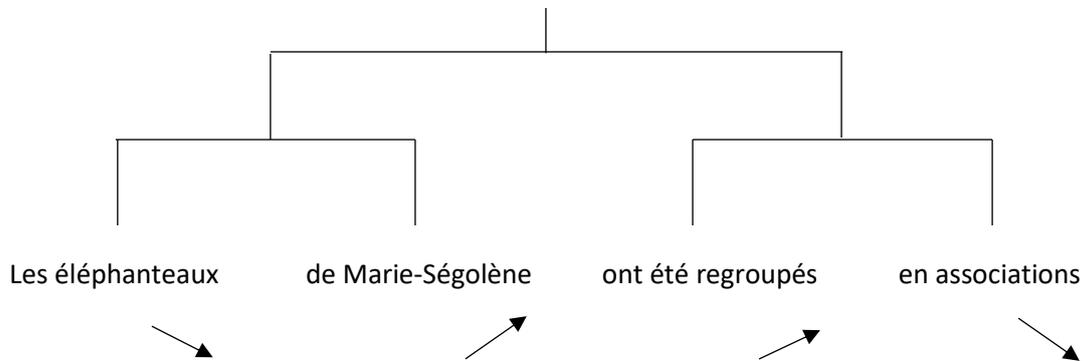


Figure 5. Illustration par Martin (2009 : 108) du principe de contraste de pente. Les flèches représentent les contours intonatifs par lesquels se terminent les groupes prosodiques. Les contours à la fin du deuxième et du quatrième groupe accentuel sont plus amples que les deux autres contours, car ils se trouvent également en fin de groupe intonatif.

4.2. Marquage de la modalité

Nous venons de voir que les contours intonatifs des groupes prosodiques se trouvant en milieu d'énoncé ont pour fonction d'exprimer une continuation. Logiquement, les contours des groupes se trouvant en fin d'énoncé, que l'on appelle contours *terminaux*, expriment quant à eux une finalité. En employant un contour terminal, le locuteur fait savoir à l'auditeur que l'énoncé est terminé et qu'il ne faut pas attendre de suite. La présence d'un contour terminal est donc un critère de complétude de l'énoncé, de même que la bonne formation syntaxique de ce dernier (Martin 2009 : 124-130). Cependant, comme nous allons le voir dans la partie suivante sur la désaccentuation (cf § 4.3), il existe des cas où un contour terminal peut tomber au milieu de l'énoncé (même si celui-ci n'est pas encore complet syntaxiquement).

En signalant la fin de l'énoncé, les contours terminaux marquent également la modalité de ce dernier. On observe des contours de formes différentes selon la modalité qui est exprimée :

- **modalité déclarative** : Il s'agit généralement d'un contour descendant L*L% qui atteint le bas du registre tonal du locuteur, comme illustré plus haut (cf Figure 4) par Delattre (1966) et comme on peut le voir ci-dessous (cf Figure 6) dans l'énoncé en haut à gauche. Ce n'est cependant pas toujours le cas : dans l'énoncé en haut à droite, on voit que le contour terminal est au contraire légèrement montant. Il arrive aussi, lorsque le locuteur souhaite

« atténuer » l’acte illocutoire, que le contour ne descende que jusqu’au milieu du registre (Post 2002), comme on peut le voir dans l’énoncé en bas à gauche. Enfin, lorsque le locuteur exprime également une attitude telle que le doute ou l’évidence, l’énoncé se termine fréquemment par un contour bitonal montant-descendant LH*L% (Delais-Roussarie et al 2015), comme c’est le cas en bas à droite.

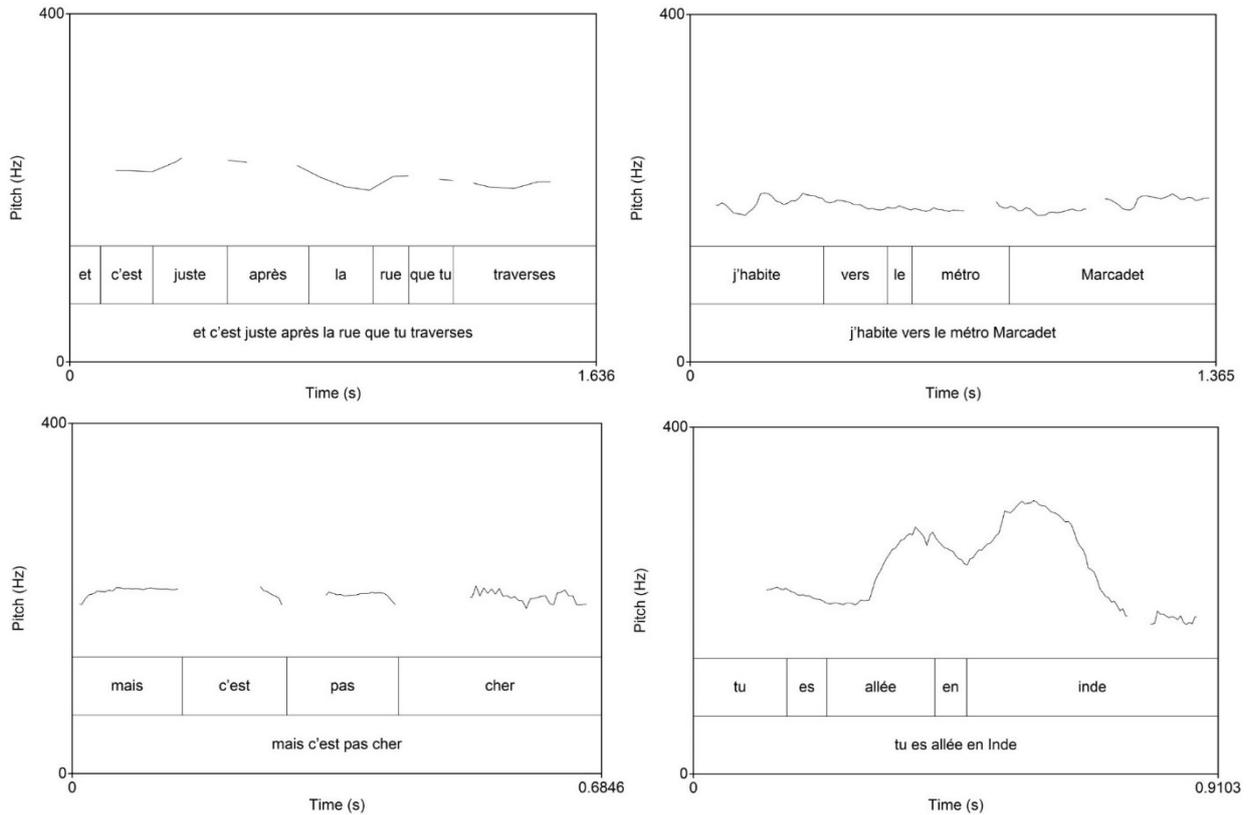


Figure 6. Courbes de fréquence fondamentale de quatre énoncés déclaratifs issus du corpus de l’expérience de production.

- **modalité interrogative** : Si l’énoncé ne contient pas de marques morphosyntaxiques de l’interrogation (mots interrogatifs ou inversion du sujet), on observe généralement un contour montant H*H% qui atteint le haut du registre du locuteur (Delattre 1966, Martin 2009 : 87-89), comme dans l’énoncé en haut à gauche ci-dessous (cf Figure 7). De même que pour les énoncés déclaratifs, le contour peut n’atteindre que le milieu du registre si l’acte illocutoire est atténué (Post 2002), comme dans l’énoncé en haut à droite. En revanche, si la modalité est marquée morphosyntaxiquement, l’énoncé peut se terminer par un contour descendant L*L% afin d’éviter une redondance de la prosodie, comme c’est le cas dans l’énoncé en bas.

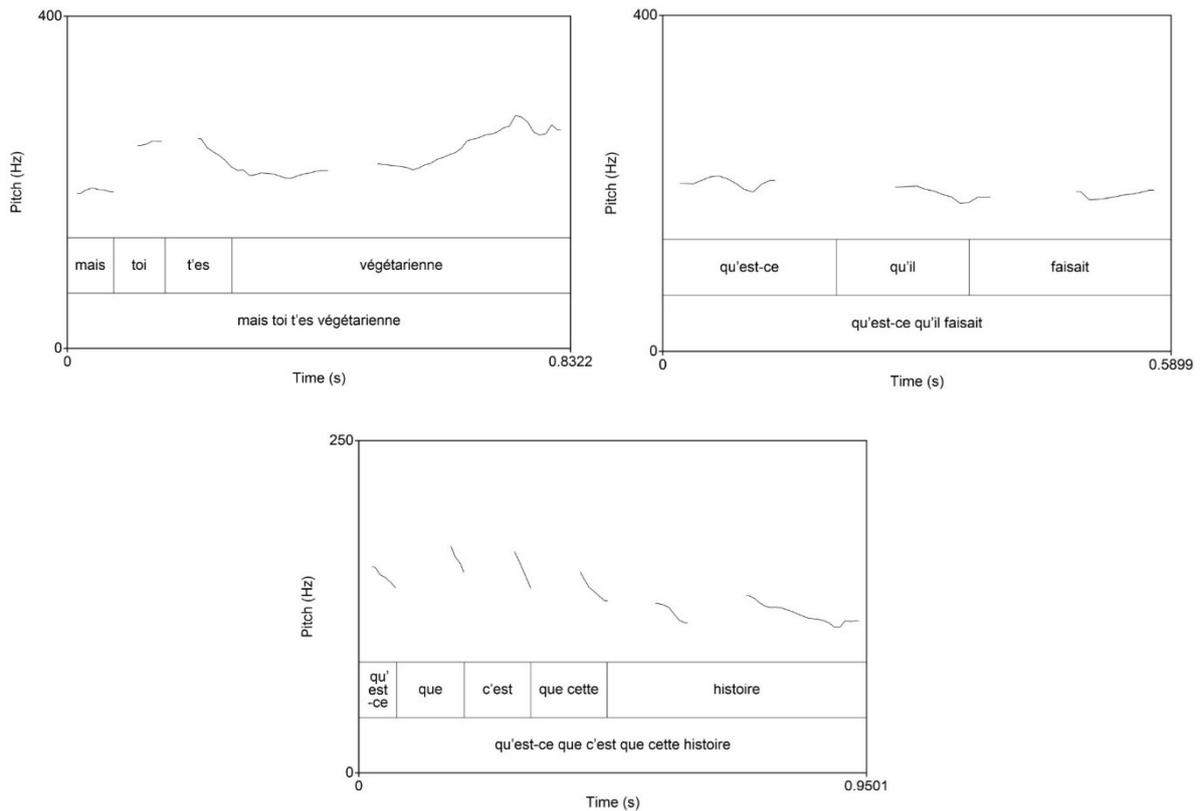


Figure 7. Courbes de fréquence fondamentale de trois énoncés interrogatifs issus du corpus de l'expérience de production.

- **modalité exclamative** : Plusieurs formes sont observées. La première est une courbe partant du haut du registre au début de l'énoncé et atteignant le bas du registre à la fin (Delattre 1966), comme dans l'énoncé en haut à gauche ci-dessous (cf Figure 8). La deuxième est un contour montant-descendant LH*L% à la fin de l'énoncé (Delais-Roussarie et al 2015), comme dans l'énoncé en haut à droite. La troisième est un contour montant et ample H*H% (Martin 1987), comme dans l'énoncé en bas à gauche. Enfin, on observe parfois un contour initial Hi au début de l'énoncé (Delais-Roussarie et al 2015), comme dans l'énoncé en bas à droite.

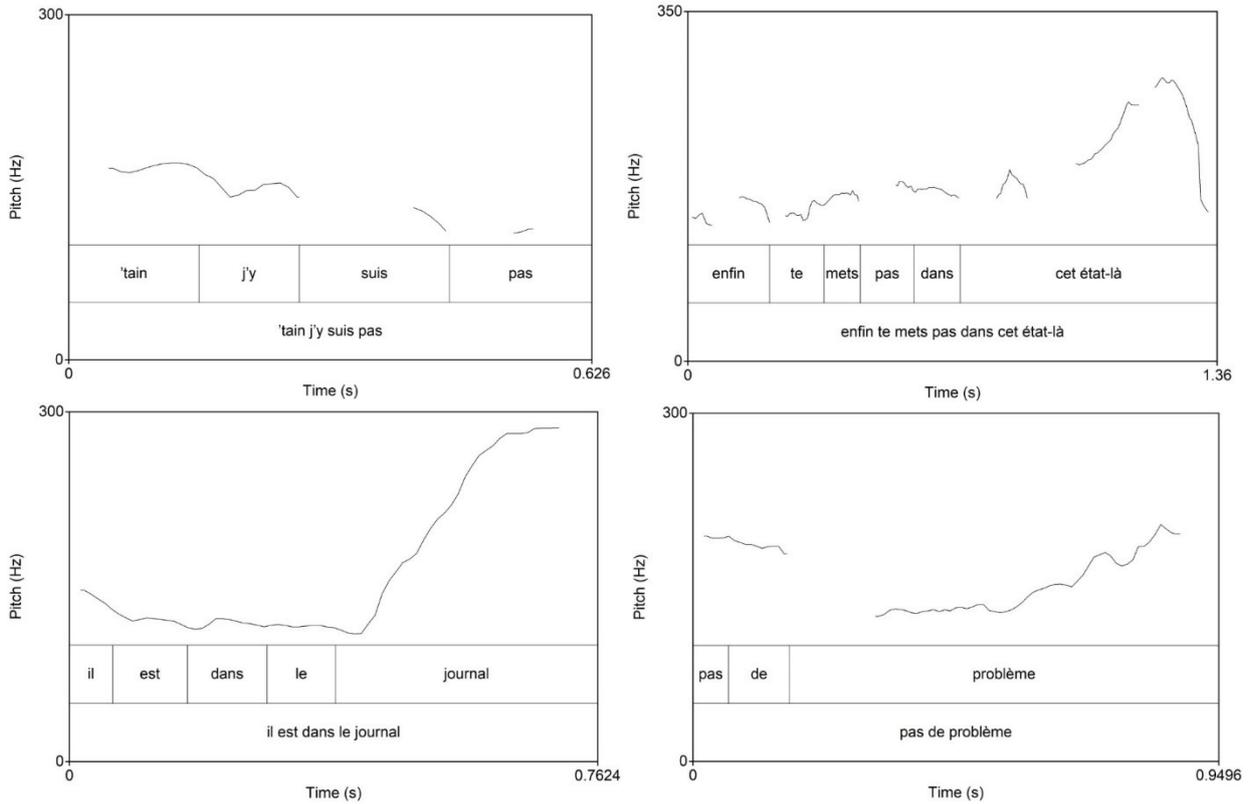


Figure 8. Courbes de fréquence fondamentale de quatre énoncés exclamatifs issus du corpus de l'expérience de production.

- **modalité impérative** : Comme pour la modalité exclamative, on observe souvent une courbe partant du haut du registre et atteignant le bas (Delattre 1966, Martin 1987), comme dans l'énoncé de gauche ci-dessous (cf Figure 9). On observe également un contour montant-descendant LH*L% (Delais-Roussarie et al 2015), comme dans l'énoncé de droite.

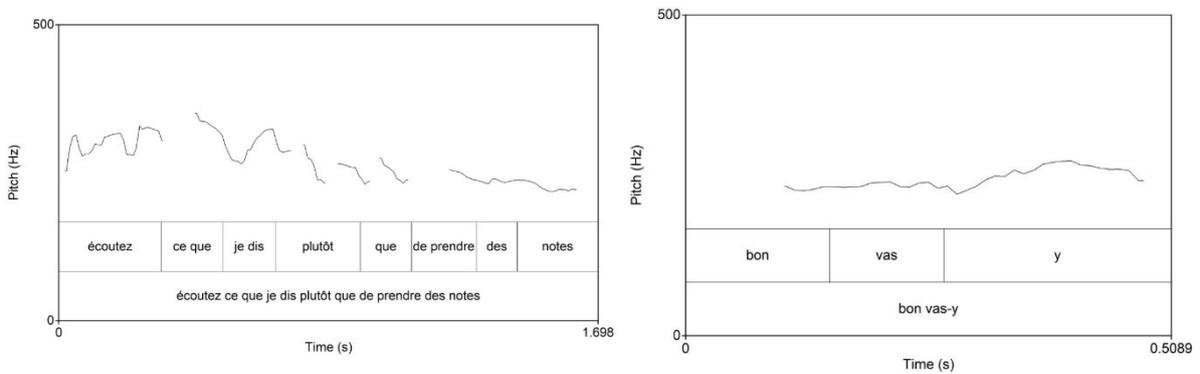


Figure 9. Courbes de fréquence fondamentale de deux énoncés impératifs issus du corpus de l'expérience de production.

4.3. Désaccentuation post-focale ou pré-focale

La désaccentuation est un phénomène intonatif qui a toute son importance dans cette thèse car il s'agit d'un des traits souvent rapportés de la focalisation prosodique, en particulier pour la fonction de marquage de focus informationnel étroit (Delattre 1966, Lambrecht 1994 : 322-333, Di Cristo 1999b, Rossi 1999 : 84-85, Jun et Fougeron 2000, Martin 2009 : 124-130, Delais-Roussarie et al 2015). Ce trait se décrit globalement comme un contour de forme plate se situant dans le bas du registre tonal du locuteur. Cependant, comme nous allons le voir, on observe en réalité une variété de formes selon différents facteurs, notamment la modalité de l'énoncé et la position de la séquence désaccentuée. La désaccentuation peut porter sur la séquence qui précède le constituant qui porte le contour terminal de l'énoncé, mais aussi sur la séquence qui suit ce constituant. En effet, comme nous l'avons mentionné au début de la section précédente (cf § 4.2), il arrive que le contour terminal tombe sur un constituant se trouvant au milieu de l'énoncé (il s'agit là aussi d'un des traits de la focalisation prosodique, cf *ibid.*).

Dans l'énoncé de gauche ci-dessous (cf Figure 10), un contour terminal descendant assertif tombe sur *voisin*, et toute la séquence qui suit est désaccentuée. On parle dans ce cas de désaccentuation *post-focale*. Par rapport au début de l'énoncé, on voit que le registre tonal s'abaisse et se rétrécit à la fois. Dans l'énoncé de droite, où *les journaux* est désaccentué, on voit que la désaccentuation peut aussi se produire à la suite d'un contour montant exclamatif (sur *arrangeait*). Comme nous l'avons vu dans les exemples (3) et (7) plus haut, *le maire de Bormes* dans l'énoncé de gauche est une information « ancienne », car le locuteur vient de la formuler dans l'énoncé précédent. Le marquage du statut informationnel ancien d'une information est une des fonctions fréquemment attribuées à la désaccentuation (e.g. Halliday 1967, Chafe 1974, Lambrecht 1994 : 322-333). Cependant, d'après Bolinger (1986 : 110-127), la désaccentuation remplit une fonction plus large consistant à exprimer l'absence d'intérêt du locuteur pour une information, quelle que soit la raison. Certains auteurs tels que Rossi (1999 : 82-96) considèrent pour leur part que la désaccentuation est un phénomène purement phonologique : la séquence désaccentuée est une « copie à forme réduite » du contour terminal.

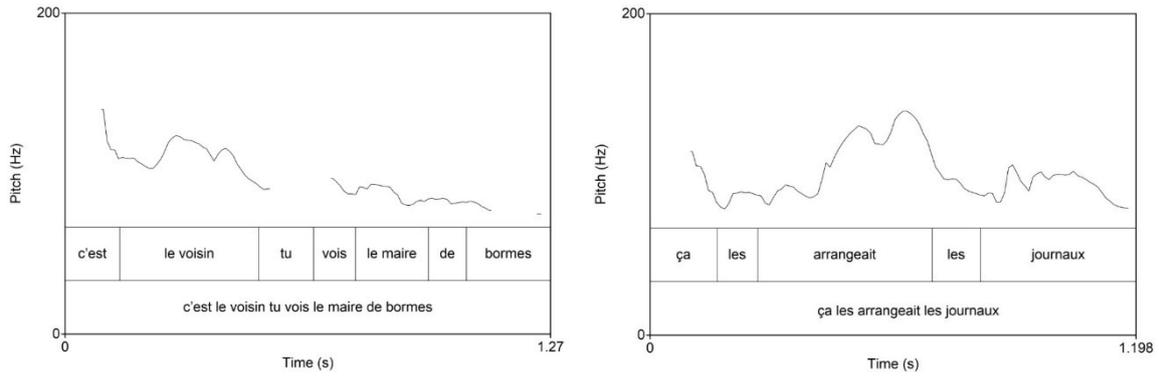


Figure 10. Courbes de fréquence fondamentale de deux énoncés issus du corpus de l'expérience de production contenant une désaccentuation post-focale après un contour terminal assertif (énoncé de gauche) et exclamatif (énoncé de droite).

Dans l'énoncé ci-dessous (cf Figure 11), la séquence désaccentuée (*je m'étais pas*) précède cette fois le contour terminal (*sur battu*). On parle dans ce cas de désaccentuation *pré-focale*. La réalisation prosodique est semblable aux deux énoncés précédents : contour plat dans le bas du registre tonal. Cependant, d'après Ladd (1980 : 55), le rétrécissement et l'abaissement du registre tonal est souvent beaucoup moins marqué lorsque la zone désaccentuée précède le contour terminal que lorsqu'elle le suit.

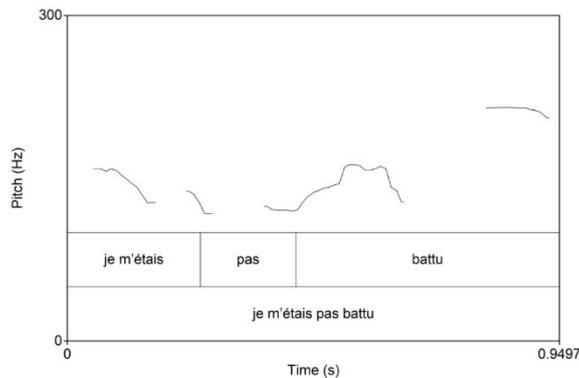


Figure 11. Courbe de fréquence fondamentale d'un énoncé issu du corpus de l'expérience de production contenant une désaccentuation pré-focale avant le contour terminal.

Dans l'énoncé ci-dessous (cf Figure 12), la désaccentuation est de nouveau post-focale, mais l'énoncé est cette fois de modalité interrogative. Notons que le contour terminal (*sur bien*) est montant. Avec un contour interrogatif descendant, la séquence désaccentuée aurait la même réalisation que dans un énoncé assertif ou exclamatif (cf Figure 10). Ici, cependant, on voit

que la séquence désaccentuée (*quand même*) a une réalisation très différente : elle se termine par un contour montant se situant dans le haut du registre tonal et atteignant le même niveau que le contour terminal montant. Il arrive parfois que le contour soit plat et qu'il se situe un peu plus bas que le contour terminal, notamment si la séquence désaccentuée est longue (Delattre 1966, Wunderli 1987, Jun et Fougeron 2000).

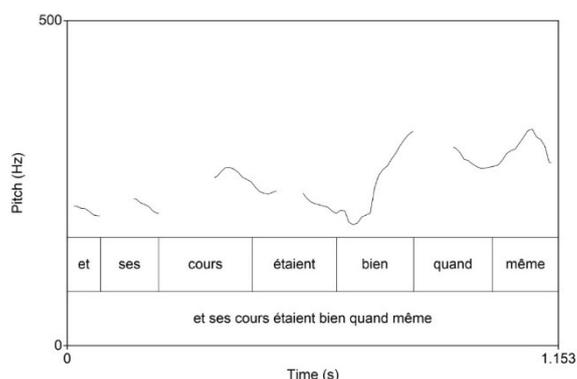


Figure 12. Courbe de fréquence fondamentale d'un énoncé issu du corpus de l'expérience de production contenant une désaccentuation post-focale après un contour interrogatif montant.

4.4. Fonctions expressives

Bien que le rôle des contours intonatifs dans l'expression des émotions et des attitudes semble très important, il existe relativement peu d'études sur le sujet. On trouve plus d'études concernant d'autres domaines de la prosodie tels que l'accentuation, le phrasé, la qualité vocalique ou les variations de registre tonal. En anglais, Bolinger (1986 : 74-88) s'est notamment intéressé aux « accents de puissance » (*accents of power*), qui ont pour fonction d'impressionner ou d'émouvoir l'auditeur et qui portent sur l'énoncé entier. L'étude de Selting (1994) montre quant à elle que les attitudes (les signaux conscients envoyés par le locuteur, cf § 1.2) sont globalement marquées par une forte densité de syllabes accentuées et par des groupes prosodiques courts et isochroniques. Souvent, les accents des énoncés concernés ont également une hauteur et une intensité supérieures à la moyenne de tous les accents. Enfin, plusieurs études ont observé des corrélations entre certains aspects prosodiques et certaines « dimensions » expressives pouvant servir à décrire les émotions et attitudes. Une de ces

dimensions, l'*activation*, désigne le degré d'importance et de nouveauté d'un stimulus et la possibilité par le locuteur de contrôler ce dernier. D'après Scherer (1986), plus l'activation est grande, plus la tension de la voix augmente, ce qui a plusieurs conséquences au niveau acoustique. On a également observé que le registre tonal s'élève de manière proportionnelle à l'*intensité* du stimulus, c'est-à-dire le niveau d'excitation physiologique que celui-ci procure au locuteur (Scherer et al 1984, Bänziger et Scherer 2005, Audibert et al 2010).

Les études sur les fonctions expressives des contours intonatifs du français concernent principalement les attitudes d'évidence, de doute et de surprise. Les résultats suivants sont observés :

- **évidence** : Les résultats ne convergent pas concernant cette attitude. Le contour est descendant d'après Post (2000), descendant-montant d'après Martin (1987), et montant ou montant-descendant d'après Delais-Roussarie et al (2015). Dans l'énoncé en haut à gauche ci-dessous (cf Figure 13), une évidence est vraisemblablement exprimée par le contour montant-descendant sur *sais*.
- **doute** : Le contour est montant-descendant (Martin 1987), comme illustré par le contour sur *convoqué* dans l'énoncé en haut à droite ci-dessous.
- **surprise** : Le contour est montant et ample (Martin 1987, Delais-Roussarie et al 2015), comme illustré par le contour sur *six heures dix* dans l'énoncé en bas ci-dessous.

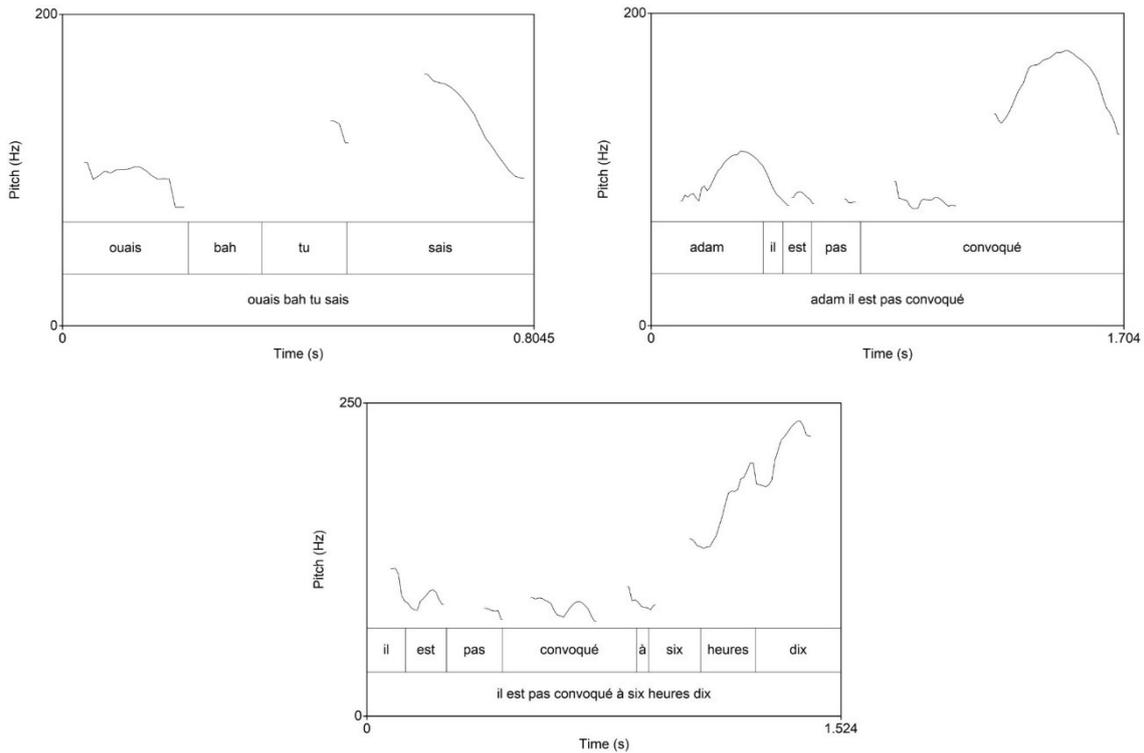


Figure 13. Courbes de fréquence fondamentale de trois énoncés issus du corpus de l'expérience de production exprimant des attitudes d'évidence (énoncé en haut à gauche), de doute (énoncé en haut à droite) et de surprise (énoncé du bas).

5. Résumé

Dans ce chapitre sont définies les notions prosodiques nécessaires à l'étude du phénomène de focalisation prosodique. La première de ces notions est la prosodie elle-même, un ensemble de paramètres acoustiques (la hauteur, la durée, l'intensité, la qualité vocalique et les pauses) qui remplit un certain nombre de fonctions (notamment la délimitation d'unités de parole, l'indication des relations entre unités, l'assignation de proéminences et l'expression d'émotions et d'attitudes) et qui se divise en plusieurs domaines formels (notamment l'accentuation, le phrasé et l'intonation). Le français est caractérisé comme une langue sans accent lexical où il existe un syncrétisme entre accentuation et intonation.

Les principaux domaines de la prosodie sont ensuite successivement abordés pour le français, à commencer par celui de l'accentuation. Plusieurs catégories d'accents sont définies. L'accent rythmique délimite les groupes prosodiques. L'accent focal indique un focus dans l'énoncé.

L'accent emphatique marque une emphase sur un élément. Une distinction est effectuée entre le niveau superficiel du rythme et le niveau sous-jacent du mètre. Au niveau du mètre, trois unités entrant dans une hiérarchie sont définies : le pied, le mot prosodique et l'unité intonative. La distinction entre mots accentuables et non-accentuables est présentée, de même que les principes de bipolarisation et de subordination accentuelle. Au niveau du rythme, trois contraintes sont présentées : l'évitement des collisions accentuelles, l'évitement des vides accentuels et l'équilibre rythmique. Enfin, la réalisation acoustique des accents est décrite selon leur catégorie ainsi que selon leur position initiale ou finale.

Le second domaine abordé est celui de la structure prosodique du français, c'est-à-dire la hiérarchie entre groupes prosodiques à la surface de l'énoncé. La structure prosodique obéit à plusieurs règles : les principes d'étagement, de dominance, d'exhaustivité et de non-récurtivité (cf Selkirk 1984), ainsi que les contraintes de planarité et de connexité (cf Martin 1987). La structure prosodique comprend deux niveaux principaux : le groupe accentuel et le groupe intonatif, le second dominant le premier. Un niveau supplémentaire a été proposé par certains auteurs, celui du groupe intermédiaire. Les propriétés syntaxiques, accentuelles et intonatives de ces trois niveaux sont décrites.

Enfin, le domaine de l'intonation du français et de ses différentes fonctions est abordé. Les contours continuatifs indiquent les relations de dépendance entre groupes prosodiques. On fait souvent à ce sujet une distinction entre continuation mineure et majeure (cf Delattre 1966), et les contours continuatifs peuvent fonctionner selon un mécanisme de contraste de pente (cf Martin 1975, 1987). L'intonation joue également un rôle dans le marquage de la modalité de l'énoncé. Les contours terminaux associés aux différentes modalités sont décrits. Un phénomène important de l'intonation, la désaccentuation, est décrit. La désaccentuation précède ou suit une focalisation avec présence de contour terminal, et connaît des réalisations différentes selon la modalité de l'énoncé. Enfin, les fonctions expressives de l'intonation sont abordées. Malgré le manque de recherches dans ce domaine, certaines associations entre contours intonatifs et certaines attitudes (évidence, doute et surprise) sont décrites.

Chapitre 2 : Influence du phonogène sur la fréquence d'occurrence et la réalisation prosodique de la focalisation

1. Introduction

1.1. Questions de recherche

Dans ce chapitre, la question suivante est posée : la focalisation prosodique présente-t-elle des propriétés spécifiques en parole interprétée ? Nous nous intéressons plus précisément à deux aspects de la focalisation : sa fréquence d'occurrence et sa réalisation prosodique. Nous souhaitons savoir si la fréquence d'occurrence de la focalisation est supérieure en parole interprétée par rapport à d'autres phonogènes, et si la parole interprétée favorise certaines réalisations prosodiques. De manière plus générale, nous souhaitons également savoir quelles sont les propriétés de la focalisation prosodique en français indépendamment du phonogène étudié. Les questions de recherche de cette étude sont donc les suivantes :

Question 1 : Quelle est la fréquence d'occurrence de la focalisation en français ?

Question 2 : Quelle est la réalisation prosodique de la focalisation en français ?

Question 3 : La focalisation présente-t-elle une fréquence d'occurrence supérieure en parole interprétée ?

Question 4 : La focalisation présente-t-elle des réalisations prosodiques spécifiques en parole interprétée ?

Afin d'évaluer l'effet de la parole interprétée sur la focalisation, il est nécessaire de comparer la focalisation dans ce phonogène ainsi que dans d'autres phonogènes. Nous avons choisi d'analyser un corpus comprenant de la parole interprétée, de la parole spontanée et de la parole lue. Le choix de la parole spontanée est assez naturel, étant donné qu'il s'agit du phonogène le plus fréquent, et en quelque sorte du phonogène « par défaut » de la communication. Le choix de la parole lue a une raison indirecte. Étant donné que nous définissons la parole interprétée comme l'oralisation d'un texte écrit qui a été mémorisé au préalable par le locuteur, les différences prosodiques entre la parole interprétée et la parole

spontanée pourraient être dues soit uniquement au processus d'oralisation, soit au processus supplémentaire de mémorisation. L'inclusion de la parole lue, qui constitue également une oralisation de texte écrit, permet de cerner le facteur responsable des différences observées entre la parole interprétée et la parole spontanée. Nous présentons plus bas (cf § 3.2) l'état de l'art concernant les différences prosodiques entre parole spontanée, lue et interprétée.

Notons que la focalisation est envisagée dans ce chapitre comme un phénomène prosodique global et non à travers les différentes fonctions que lui attribue la littérature. La définition que nous adoptons de la focalisation est présentée plus bas (cf § 2.1). Les fonctions attribuées à la focalisation font l'objet du chapitre suivant, dans lequel nous cherchons à savoir s'il existe une réalisation prosodique différente pour chaque fonction en français.

1.2. Arrière-plan et prédictions

Plusieurs sources ont inspiré les questions de recherche de cette étude. Les questions 1 et 2 ont simplement pour but de dégager des généralisations descriptives concernant la focalisation en français. Le choix des paramètres prosodiques sur lesquels l'étude se concentre s'appuie sur les observations rapportées dans la littérature (e.g. Di Cristo 1999b, Rossi 1999, Astésano 2001), que nous exposons de manière détaillée plus bas (cf § 2.2). La nouveauté apportée par cette étude réside dans le fait que les données sont systématisées sur de la parole de laboratoire, et surtout sont établies dans un cadre contrastif qui est crucial pour l'établissement de généralités caractéristiques de la grammaire du français.

La question 3 se base tout d'abord sur la conception courante selon laquelle la parole interprétée est plus expressive que la parole spontanée. Comme nous le verrons plus bas (cf § 3.2), les études sur les différences prosodiques entre ces deux phonogenres n'apportent pas vraiment de fondements à cette conception. Néanmoins, on pourrait s'attendre à ce que la focalisation présente une fréquence d'occurrence supérieure en parole interprétée. Cette prédiction trouve par ailleurs un début de confirmation dans deux études. Goldman et al (2011) montrent que la fréquence d'accents initiaux augmente avec le degré de « préparation » du discours (cf § 3.1). Or l'accentuation initiale est un des traits fréquents de la focalisation, et la parole interprétée implique justement un grand degré de préparation (car le locuteur mémorise le texte). Godement-Berline (2014), une étude-pilote sur la parole interprétée, montre que celle-ci contient un pourcentage de syllabes proéminentes

relativement élevé par rapport aux pourcentages relevés par Simon et al (2010) sur les phonogenres du corpus C-PROM (Avanzi et al 2010). On a là aussi un argument pour prédire une plus haute fréquence d'occurrence de la focalisation en parole interprétée.

La question 4 se base sur le fait que, comme nous le verrons plus bas (cf § 3.2), la parole lue et la parole interprétée se caractérisent par une prosodie globalement plus marquée que la parole spontanée (concernant des paramètres tels que la déclinaison et la variation de la hauteur, les allongements de fin de groupe ou les variations de registre tonal). On peut s'attendre à ce que la focalisation soit elle aussi plus marquée en parole lue et en parole interprétée, par exemple du point de vue de la hauteur et de la durée, ou bien de la présence d'accent initial.

2. La focalisation prosodique

2.1. Définition

Comme nous l'avons expliqué plus haut (cf § 1.1), la focalisation est envisagée dans ce chapitre comme un phénomène global. La définition que nous adoptons de la focalisation est la suivante : il s'agit simplement du soulignement d'un constituant dans un énoncé, pour toute raison sémantico-pragmatique ou emphatique. On peut également parler de mise en valeur, de mise en relief, ou bien de distinction d'une figure sur un fond (cf notamment la théorie psychologique du gestaltisme, e.g. Koffka 1922).

2.2. Traits prosodiques

Les domaines prosodiques que met en jeu la focalisation sont l'accentuation, l'intonation, le phrasé, les variations de tempo et les variations de registre tonal. Les principaux traits prosodiques de la focalisation ont déjà été évoqués au cours du chapitre sur la prosodie du français. Nous rappelons ces traits ci-dessous :

- **augmentation de hauteur, de durée et d'intensité** : Les accents de focalisation se caractérisent par une augmentation de hauteur et de durée plus forte que chez les accents rythmiques, ainsi que par une augmentation d'intensité qui est absente chez les accents rythmiques (Rossi 1999 : 128, Astésano 2001 : 53-57). Les accents de focalisation se

caractérisent également par un allongement syllabique plus important que les accents rythmiques (Astésano 2001 : 192) et par des configurations tonales plus saillantes (ibid., 253-254).

- **accentuation initiale** : Les constituants focalisés qui comportent plusieurs syllabes sont fréquemment marqués par un accent initial sur une des premières syllabes (Di Cristo 1999b, Astésano 2001 : 202). Cet accent initial forme généralement un arc accentuel avec l'accent final (Di Cristo 1999a, Rossi 1999 : 128). Il est réalisé par une augmentation significative de la durée de la consonne, ainsi que par la présence d'un coup de glotte initial si la syllabe commence par une voyelle (Delais-Roussarie et Di Cristo, à paraître). L'accent initial se distingue de l'accent final par le fait que l'attaque syllabique est plus allongée que la rime, alors que c'est l'inverse pour l'accent final (Astésano 2001 : 202). Il se distingue également par le fait que sa configuration tonale est asymétrique (la pente montante de la variation de hauteur est plus abrupte que la pente descendante), alors que celle de l'accent final est symétrique (ibid., 254).
- **contour terminal** : Le constituant focalisé peut porter un contour terminal indiquant la modalité (déclarative, interrogative, exclamative ou impérative) de l'énoncé (Rossi 1999 : 84-87, Post 2002, Martin 2009 : 124-130). Ce contour peut être présent même lorsque le constituant focalisé se trouve au milieu de l'énoncé et non à la fin. La forme du contour terminal dépend de la modalité mais connaît beaucoup de variation. Le contour terminal n'atteint pas toujours le haut ou le bas du registre tonal (Post 2002), et certaines formes peuvent servir à exprimer plusieurs modalités.
- **désaccentuation post-focale ou pré-focale** : Ce phénomène peut affecter la zone qui précède ou qui suit le constituant focalisé (Delattre 1966, Wunderli 1987, Lambrecht 1994 : 322-333, Di Cristo 1999b, Rossi 1999 : 84-85, Jun et Fougeron 2000, Martin 2009 : 124-130, Delais-Roussarie et al 2015). La désaccentuation peut prendre plusieurs formes, selon notamment la modalité de l'énoncé. Lorsque la modalité est déclarative ou exclamative, on observe généralement un abaissement et un rétrécissement du registre tonal du locuteur. Cette réalisation est plus marquée en cas de désaccentuation post-focale que de désaccentuation pré-focale. Lorsque la modalité est interrogative, que la

désaccentuation est post-focale et que le contour du constituant focalisé est montant, on observe au contraire un contour montant ou plat se situant dans le haut du registre tonal.

3. Parole spontanée, lue et interprétée

3.1. Définitions

Cette partie s'inscrit dans le domaine de la *phonostylistique*, ou l'étude des caractéristiques phonétiques des différents styles de parole. Ce domaine a été développé notamment par Troubetzkoy (1939), puis pour le français par Ivan Fónagy (e.g. Fónagy 1964, 1983) et Pierre Léon (e.g. Léon 1969, 1993). Beaucoup d'études ont également été menées de manière indépendante pour d'autres langues, notamment l'anglais. Une distinction est souvent faite en phonostylistique entre deux catégories de styles de parole, les *phonostyles* et les *phonogenres*. La littérature contient plusieurs définitions concernant ces deux catégories. Nous adoptons les définitions proposées par Goldman et al (2011) :

- **phonostyles** : Il s'agit de styles liés à une classe sociale, une profession ou une activité culturelle particulière. On peut citer les exemples de l'accent « snob », de l'accent « populaire », du sermon religieux, de la plaidoirie d'avocat, de la récitation scolaire et des annonces publicitaires. Ces exemples ont été étudiés notamment par Léon (1993). Certaines études portent également sur des phonostyles propres à des individus, notamment des acteurs (e.g. Fónagy 1983 : 164-171, Léon 1993 : 274-276, Boulakia et Galazzi 2003) ou des hommes politiques (e.g. Martin 2009 : 228-232).
- **phonogenres** : Il s'agit de styles liés à une *activité de parole* particulière. Une activité de parole se caractérise par ses conditions de production. Les conditions de production concernent notamment le caractère « médiatique » ou non du discours (diffusion sur des médias de masse), le « type d'audience » (micro, face à face ou public) et le degré de « préparation » du discours (improvisé, semi-préparé, lu ou mémorisé). Parmi les phonogenres, beaucoup d'études ont été consacrées au journal parlé (e.g. Léon 1993 : 164-166, Simon et al 2010, Boula de Mareüil et al 2011, Gendrot et al 2012) ainsi qu'à l'allocution politique (e.g. Duez 1991, Léon 1993 : 166-168, Touati 2003, Simon et al 2010, Feldhausen et Delais-Roussarie 2012). Bien que ces deux styles soient également liés à des

professions (ce qui pourrait les ranger dans la catégorie des phonostyles), on peut en effet les définir selon leurs conditions de production. Dans les deux cas, il s'agit d'un discours médiatique et lu à haute voix. On peut cependant les distinguer selon le type d'audience : le journal parlé est produit devant un micro, tandis que l'allocution politique est produite en face d'un public.

D'après ces définitions, la parole spontanée et la parole lue constituent des phonogenres et non des phonostyles. En effet, on peut caractériser ces deux styles selon leurs conditions de production (cf Tableau 3). La parole interprétée, étant donné qu'elle est liée à une catégorie professionnelle (voire culturelle) particulière, peut être considérée comme un phonostyle. Cependant, l'analyse proposée dans ce chapitre et dans le suivant porte plutôt sur les conditions de production de la parole interprétée. En effet, seule la moitié des locuteurs de parole interprétée que nous avons enregistrés est constituée d'acteurs professionnels (cf § 4.1.2). Ces acteurs ont été de surcroît enregistrés dans un lieu neutre (une chambre sourde), et non dans un lieu lié à leur profession (une scène de théâtre ou un plateau de tournage). En tant que phonogène, la parole interprétée se définit comme l'oralisation d'un texte écrit qui a été mémorisé au préalable par le locuteur. Le paramètre du degré de préparation sépare donc la parole interprétée de la parole spontanée et de la parole lue (cf Tableau 3). Cependant, le caractère médiatique et le type d'audience de la parole interprétée peuvent varier. Dans le cas du théâtre, le discours est de type non-médiatique, et produit devant un public. En revanche, dans le cas du cinéma et de la télévision, le discours est de type médiatique, et produit devant une caméra.

	parole spontanée	parole lue	parole interprétée
caractère médiatique	discours non-médiatique	discours non-médiatique	discours médiatique ou non-médiatique
type d'audience	face à face	micro	public ou caméra
degré de préparation	improvisé	lu à haute voix	mémorisé

Tableau 3. Conditions de production de la parole spontanée, de la parole lue et de la parole interprétée.

Notons que plusieurs études récentes sur les phonogenres en français adoptent une approche dite « multidimensionnelle » (e.g. Degand et Simon 2009, Simon et al 2010, Goldman et al 2011, Goldman et al 2014, Pršir et al 2014). Cette approche repose sur une classification des styles de parole selon différentes « dimensions situationnelles ». Ces dimensions comprennent les conditions de production (caractère médiatique, type d'audience, degré de préparation), ainsi que d'autres dimensions telles que le degré de formalité du discours, le degré d'intimité avec l'interlocuteur, l'ancrage référentiel dans la situation, ou encore le degré de liberté thématique (cf Labov 1972, Joos 1972, Eskenazi 1993, Koch et Oesterreicher 2001). L'approche multidimensionnelle consiste à étudier les relations entre ces dimensions et les différents traits prosodiques, plutôt que d'étudier les caractéristiques prosodiques de chaque phonogène séparément. Des relations sont notamment observées entre le degré de formalité et le phrasé (Degand et Simon 2009), le caractère médiatique et l'accentuation (Goldman et al 2011), le degré de préparation et l'accentuation (ibid.), le caractère public ou privé du discours et le registre tonal (Simon et al 2010), ainsi que la contrainte de temps et les pauses (Pršir et al 2014).

3.2. Différences prosodiques

Les différences prosodiques entre la parole spontanée et la parole lue ont fait l'objet de nombreuses études^{4 5}. Notons que la plupart de ces études portent sur l'anglais. Nous présentons ci-dessous (cf Tableau 4) les résultats observés concernant différents aspects prosodiques tels que l'intonation, l'accentuation, la durée ou la qualité vocalique. Comme on peut le voir, les résultats sont souvent contradictoires, en particulier concernant l'intonation, les variations de registre tonal et la durée. Cependant, on observe plusieurs différences. La

⁴ Notons que ces études ont généralement une visée technologique : l'amélioration de systèmes de reconnaissance et de synthèse de la parole. En effet, ces systèmes sont généralement entraînés sur de la parole lue, mais ont pour but de recevoir de la parole spontanée (dans le cas de la reconnaissance), ou de produire de la parole ressemblant à de la parole spontanée (dans le cas de la synthèse). Pour obtenir une bonne qualité, il est nécessaire de savoir en quoi les deux phonogenres divergent.

⁵ Notons également que le protocole expérimental employé par la plupart de ces études (e.g. Batliner et al 1995, Laan 1997, Wagner et Windmann 2016) est très proche du protocole RepTask (cf Laurens et al 2011) que nous avons employé dans notre expérience de production (cf § 4.1). Ce protocole consiste à demander à des locuteurs de reproduire une conversation spontanée qui a vraiment eu lieu. Notre expérience se distingue par le fait que nous n'avons pas seulement demandé aux locuteurs de lire le texte à haute voix, mais aussi de l'apprendre par cœur.

parole lue se caractérise par une plus forte déclinaison de la hauteur que la parole spontanée, un plus grand nombre de contours descendants, un plus grand nombre de syllabes accentuées, des groupes prosodiques plus courts, des allongements de fin de groupes plus marqués, un nombre beaucoup plus faible de marques d'hésitation, et moins de perturbation de la qualité vocalique.

intonation	registre tonal	durée et débit d'articulation
<ul style="list-style-type: none"> - plus de variations en parole lue (Laan 1997) - moins de variations en parole lue (Lieberman et al 1985) - plus de déclinaison en parole lue (Swerts et al 1996, Laan 1997) - plus de contours descendants en parole lue (Lieberman et al 1985, Blaauw 1994, de Ruiter 2015) 	<ul style="list-style-type: none"> - plus élevé en parole lue (Swerts et al 1996, Laan 1997) - moins élevé en parole lue (Batliner et al 1995) - plus étendu en parole lue (Batliner et al 1995, Laan 1997, Goldman et al 2011) - plus réduit en parole lue (Simon et al 2010, Wagner et Windmann 2016) 	<ul style="list-style-type: none"> - plus d'allongement sur les accents finaux en parole lue (Astésano 2001 : 192) - débit plus élevé en parole lue (Mixdorff et Pfitzinger 2005, Simon et al 2010, Wagner et Windmann 2016) - débit moins élevé en parole lue (Batliner et al 1995, Laan 1997, Dellwo et al 2015) - beaucoup moins de marques d'hésitation en parole lue (Wagner et Windmann 2016)
accentuation	phrasé	qualité vocalique
<ul style="list-style-type: none"> - variation dans la localisation des accents (Howell et Kadi-Hanifi 1991) - plus de syllabes accentuées en parole lue (Mixdorff et Pfitzinger 2005, Wagner et Windmann 2016) 	<ul style="list-style-type: none"> - peu de différences (Blaauw 1994, Swerts et al 1996) - groupes prosodiques plus courts en parole lue (Howell et Kadi-Hanifi 1991) 	<ul style="list-style-type: none"> - moins de <i>shimmer</i> et moins de réduction de la qualité spectrale des voyelles en parole lue⁶ (Laan 1997)

Tableau 4. Différences observées entre la parole spontanée et la parole lue selon différents aspects prosodiques.

⁶ Le *shimmer* est la variation d'amplitude du signal sonore entre deux cycles. Il reflète la qualité de la vibration des cordes vocales. La réduction de la qualité spectrale des voyelles désigne la déviation du premier et du deuxième formant vers la position des formants du schwa (environ 500 et 1500 Hz). Cette déviation est la conséquence d'une prononciation relativement relâchée.

Par comparaison avec la parole lue, la parole interprétée a fait l'objet de moins d'études⁷. De plus, ces études portent parfois sur un autre aspect que les différences prosodiques entre la parole interprétée et la parole spontanée :

- Audibert et al (2008) et Jürgens et al (2011) testent la capacité de sujets à différencier auditivement la parole spontanée et la parole interprétée. Les sujets parviennent à effectuer cette différenciation dans la première étude, mais pas dans la seconde.
- Audibert et al (2008) comparent également des jugements d'« intensité émotionnelle » obtenus auprès de sujets concernant la parole spontanée et la parole interprétée. La parole interprétée est jugée plus intense que la parole spontanée.
- Certaines études s'intéressent enfin aux différences entre la parole produite par des acteurs professionnels et celle produite par des acteurs amateurs :
 - Jürgens et al (2015) testent la capacité de sujets à identifier, parmi ces deux types de locuteurs, le fait qu'il s'agit de parole interprétée. Les acteurs professionnels sont plus souvent correctement identifiés que les acteurs amateurs (bien que le taux global d'identification soit assez faible).
 - Krahmer et Swerts (2008) comparent des jugements d'intensité émotionnelle concernant les deux types de locuteurs. Les acteurs professionnels sont jugés plus « extrêmes » et moins naturels que les acteurs amateurs⁸.

⁷ Notons que ces études ont généralement une visée méthodologique. Elles se situent en effet dans le cadre des recherches sur l'expression vocale des émotions. Ces recherches se basent souvent sur de la parole interprétée produite par des acteurs. Le choix de ce phonogène est motivé par les problèmes éthiques que pourrait poser l'élicitation d'émotions chez des locuteurs de parole spontanée (cf Wilting et al 2006). Afin de pouvoir appliquer les résultats observés à la parole spontanée, il est nécessaire de connaître les différences entre les deux phonogènes.

⁸ Ce résultat amène les auteurs à contester l'utilisation d'acteurs professionnels dans les études sur l'expression vocale des émotions.

Seules deux études, Audibert et al (2010) et Jürgens et al (2011), s'intéressent aux différences prosodiques entre la parole spontanée et la parole interprétée. Les résultats suivants sont observés :

- La parole interprétée présente un registre tonal plus élevé que la parole spontanée (Audibert et al 2010).
- La parole interprétée présente de plus fortes variations de hauteur que la parole spontanée (Jürgens et al 2011).
- Les autres différences concernent la qualité vocalique :
 - En parole interprétée, les voyelles présentent un premier formant plus large qu'en parole spontanée (Jürgens et al 2011), ainsi qu'un deuxième formant plus bas (Audibert et al 2010).
 - On observe également, en parole interprétée, une plus forte amplitude des fréquences harmoniques basses, des pics fréquentiels moins hauts, et plus de shimmer⁹ (Jürgens et al 2011).

4. Méthodologie

Le protocole de l'étude présentée dans ce chapitre a tout d'abord consisté à constituer un corpus de parole interprétée, de parole lue et de parole spontanée. Un relevé des occurrences de focalisation prosodique a ensuite été effectué dans le corpus par un groupe d'experts en prosodie. Puis une analyse phonétique et phonologique des occurrences de focalisation a été réalisée. Nous présentons ces étapes successives ci-dessous.

⁹ Selon les auteurs, ces résultats ne sont pas dus à un plus grand degré d'excitation physiologique en parole interprétée, ni à une exagération de la prosodie. Ce qui caractérise la parole interprétée ne serait donc pas une plus forte expressivité, mais plutôt un effort pour rendre la prosodie moins monotone. Cet effort pourrait s'expliquer, toujours selon les auteurs, par une volonté de ne pas ennuyer le public, ou simplement par l'entraînement vocal qui est souvent pratiqué par les acteurs.

4.1. Constitution du corpus

Pour comparer le phénomène de la focalisation prosodique en parole interprétée, en parole lue et en parole spontanée, nous avons choisi d'analyser les mêmes textes dans ces trois phonogenres. Deux conversations spontanées ont ainsi été enregistrées puis retranscrites. Des locuteurs ont ensuite lu à haute voix, puis interprété de mémoire, la retranscription de ces conversations. Ce protocole permet de mettre en évidence l'influence du phonogenre sur la fréquence d'occurrence et la réalisation prosodique de la focalisation. Toute influence du contenu syntaxique, sémantique ou discursif du texte est en effet neutralisée, puisque ce dernier est toujours le même.

La tâche de reproduction des conversations spontanées en parole lue et en parole interprétée est inspirée du protocole RepTask (pour *replication task*) qui a été proposé et employé par Laurens, Marandin, Patin et Yoo (2011). Comme nous l'avons mentionné plus haut (cf note 5), des protocoles similaires ont en réalité été employés par un certain nombre d'études portant sur les différences phonétiques entre parole spontanée et parole lue (e.g. Lieberman et al 1985, Laan 1997, Mixdorff et Pfitzinger 2005, Wagner et Windmann 2016). Dans l'expérience de Laurens et al (2011), les locuteurs ont une expérience de l'art dramatique et lisent à haute voix de manière naturelle et expressive la transcription de la conversation spontanée. L'hypothèse de l'étude est que les choix prosodiques faits par les locuteurs de parole de laboratoire convergent avec les choix faits par les locuteurs de parole spontanée. Deux cas de figure sont étudiés : un contour continuatif montant typique de la formulation d'une « liste » par le locuteur, ainsi que le rétrécissement du registre tonal sur des énoncés déclaratifs dits de « reprise » (lors des changements de thème de discours). Les résultats confirment l'hypothèse, ce qui suggère que la parole de laboratoire reflète de manière relativement fidèle la prosodie de la parole spontanée.

Dans la suite de cette section, nous exposons en détail l'étape de l'élicitation des conversations spontanées, puis l'étape de reproduction de ces dernières en parole lue et en parole interprétée. Le corpus de l'expérience est disponible en intégralité sur le site du Laboratoire de Linguistique Formelle ainsi que sur la base de données Ortolang, aux adresses suivantes :

- <http://www.lif.cnrs.fr/reptask/parole-interpretee>

- <https://hdl.handle.net/11403/corpus-reptaskparoleinterpretee>

4.1.1. Elicitation de parole spontanée

Nous avons enregistré deux conversations spontanées. Ce choix avait pour but d'inclure des locuteurs des deux sexes (les locuteurs de parole lue et interprétée ayant reproduit le texte d'un seul des locuteurs dans chaque conversation, cf § 4.1.2). Nous présentons ci-dessous (cf Tableau 5) des informations sur les locuteurs et sur les conditions d'enregistrement de chaque conversation. Nous présentons ensuite le début des extraits qui ont été retenus pour la tâche de reproduction en parole lue et en parole interprétée (pour plus de lisibilité, le texte est légèrement simplifié, et des marques de ponctuation sont ajoutées). Comme on peut le voir, il s'agit de dialogues « à bâtons rompus » entre des locuteurs se connaissant déjà. Le fait que les locuteurs échangent librement a pour avantage que, à tout instant de la conversation, la quantité d'informations partagées (*common ground*, cf Stalnaker 1978) est minimale. Comme le font remarquer Laurens et al (2011), cela facilite la tâche de reproduction du dialogue, car le texte est compréhensible même si l'on n'a pas accès à la situation de communication originale. Dans les extraits choisis, un des deux locuteurs raconte une anecdote insolite qui lui est arrivée, tandis que l'autre locuteur est relativement silencieux. Ce choix présente deux avantages. Premièrement, les locuteurs qui ont reproduit les dialogues ont interprété le rôle du locuteur qui raconte l'anecdote, et ont donc été peu interrompus par les répliques de l'autre locuteur. Deuxièmement, le fait que les extraits présentent un contenu intéressant ainsi qu'une cohérence discursive (un début et une fin bien limités) les a rendus plus faciles à mémoriser pour les locuteurs de parole interprétée.

		production A ¹⁰	production B
locuteurs	sexe	hommes	femmes
	âge	30 et 40 ans	32 et 40 ans
	origine	région Provence-Côte d’Azur	région parisienne
	occupation	enseignants-chercheurs	doctorantes
conditions	lieu	chambre sourde	pièce calme
	matériel	micros-casques	micro portatif Zoom H2
consigne donnée aux locuteurs		raconter des situations insolites dans lesquelles les locuteurs se sont un jour trouvés	échanger librement (thème de départ : restaurants asiatiques du 13 ^{ème} arrondissement de Paris)

Tableau 5. Informations sur les deux productions spontanées de l’expérience.

Extrait de la production A :

- A propos de mon nom, du coup j’ai une anecdote : quand j’étais... je devais être... ouais, j’étais au lycée... Je jouais au foot, tu vois, à Bormes... Et pour savoir qui jouait... Tu vois, donc t’avais l’entraînement la semaine... Et le samedi, dans le journal, dans la Prov... enfin à l’époque c’était Bormes-Matin... à Bormes-les-Mimosas, tu avais les convocations des joueurs. Donc, « sont convoqués à telle heure pour jouer contre telle équipe... »
- Ah ça paraissait dans le journal ? Tu savais quand tu... ? Non, toi tu le savais d’avance.
- Non, tu savais pas.
- Ah ouais. C’est dans le journal que tu apprenais que tu...

¹⁰ Nous n’avons pas enregistré nous-même cette production. Elle provient du corpus CID (Bertrand et al 2008). Il s’agit de la production qui est analysée par Laurens et al (2011) dans leur étude.

– Si tu veux, tu t’entraînais la semaine, tu t’entraînais par exemple le mardi et le vendredi, tu vois, ou le mercredi et le vendredi... Enfin ça dépendait. Au départ c’était le mercredi, après quand t’étais junior, tout ça, c’était le soir.

Extrait de la production B :

– Donc j’étais à ce cours-là, et c’était assez drôle parce que tous les étudiants se sont précipités au cours de philosophie indienne parce que c’était plutôt original... Et le prof était tellement horrible...

– Ah bon !

– Bah, en plus on lui avait prévu pour son séminaire, on lui avait prévu une – séminaire de M1 – une toute petite salle, ce qui était étonnant parce qu’y avait...

– Hmm, ça attire en fait... Ouais, ils se doutent que...

– Enormément d’étudiants, à Paris 4 en philo, et la salle était toute petite ! Et elle était blindée à craquer le premier jour et j’étais assez étonnée...

– Ils ont pas mesuré, en fait pour eux ils se sont dits on fait un petit truc comme ça...

– Hmm et en fait...

– Mais c’est ce qui attire les étudiants !

– Mais en fait ils avaient raison, parce que au bout de deux semaines y avait plus personne, tellement il était insupportable.

4.1.2. Reproduction en parole lue et interprétée

Afin de faciliter la tâche de reproduction, nous avons transcrit orthographiquement les extraits qui ont été retenus dans les productions spontanées. Afin d’améliorer la lisibilité du texte, des marques de ponctuation ont été ajoutées. Afin que les textes ne soient pas trop difficiles à reproduire, les chevauchements de parole n’ont pas été retranscrits, et les passages incompréhensibles (ou bien jugés trop complexes) ont été supprimés ou simplifiés. La plupart des disfluences (répétitions, amorces, corrections, hésitations) ont été retranscrites, sauf lorsque nous avons jugé qu’elles rendaient le texte trop complexe. Nous présentons ci-dessous la version non-simplifiée et non-ponctuée du premier extrait qui a été présenté dans

la section précédente. Nous avons souligné les séquences qui se chevauchent entre les deux locuteurs. Les amorces sont délimitées par un trait d'union.

Extrait non-simplifié et non-ponctué de la production A :

- à propos de mon nom moi 'tain du coup j'ai une anecdote quand euh quand j'étais euh je devais être ouais j'étais au lycée je jouais je jouais au foot tu vois à Bormes et euh donc t'avais si tu veux le sam- le pour savoir qui jouait tu vois donc t'avais l'entraînement la semaine
- hm-hm
- et le samedi dans le journal tu vois dans la Prov- enfin à l'époque c'était euh Bormes Matin tu vois à Bormes-les-Mimosas tu avais les convocations des joueurs tu vois donc unt- euh donc sont convoqués à telle heure pour jouer contre unt- telle équipe
- ah ça paraissait dans le jour-
- ça paraissait dans le journal
- si tu savais quand tu c'ét- non toi tu le savais d'avance ouais
- non tu savais pas
- ah ouais c'est dans le journal que t'apprenais que tu
- parce que si tu veux tu t'entraînais tu t'entraînais la semaine alors tu t'entraînais par exemple le mardi et le vendredi tu vois
- hm-hm
- ou le mercredi et le vendredi 'fin ça dépendait au départ c'était le mercredi après quand t'étais junior tout ça c'était en le soir tout ça

La reproduction des dialogues spontanés en parole lue et en parole interprétée a été effectuée par quatre locuteurs (deux femmes et deux hommes), tous originaires de Paris. Nous présentons ci-dessous (cf Tableau 6) des informations supplémentaires sur ces locuteurs. On voit que tous les locuteurs possèdent, à un degré différent, une expérience de l'art dramatique. Ceci a pu introduire un biais dans leurs productions, dont certaines caractéristiques prosodiques pourraient être dues à un éventuel phonostyle du comédien. Nous souhaitons néanmoins que les locuteurs possèdent une expérience de l'art dramatique,

afin qu'ils soient en mesure de mémoriser les textes (qui sont relativement longs). Nous avons choisi des locuteurs n'ayant pas le même degré d'expérience, afin d'être en mesure d'évaluer l'effet de ce paramètre (une question qui fera peut-être l'objet d'une étude ultérieure). Chaque locuteur a reçu une rémunération de 75 euros. Cette somme relativement élevée s'explique par la consigne de mémorisation qui a été donnée aux locuteurs, et aussi car la tâche de reproduction était constituée de deux parties se déroulant à des jours différents.

	locuteur A1	locutrice B1	locuteur A2	locutrice B2
âge	26 ans	24 ans	36 ans	26 ans
occupation	assistant de production d'émission de radio	étudiante de Licence	architecte	comédienne professionnelle
expérience de l'art dramatique	a participé à un cours de théâtre amateur	élève dans une école d'art dramatique	acteur semi-professionnel	
dialogue reproduit	dialogue A	dialogue B	dialogue A	dialogue B

Tableau 6. Informations sur les locuteurs des productions de parole lue et de parole interprétée de l'expérience.

Les enregistrements ont été réalisés dans une chambre sourde à l'Université Paris-Diderot, au moyen d'un micro de studio Rode NT1-A, d'une interface audio Roland Quad-Capture et du logiciel Audacity. Les locuteurs ont été enregistrés un par un. Ils ont uniquement reproduit, dans les dialogues spontanés, le texte du locuteur qui raconte l'anecdote l'insolite. Nous avons prévu au départ de leur faire entendre les tours de parole de l'autre locuteur à travers des haut-parleurs¹¹. Cependant, un problème technique nous a empêché d'employer cette

¹¹ Ces tours de parole avaient été au préalable réenregistrés par nous-même avec une prosodie relativement peu expressive. Ceci avait pour but d'éviter toute influence de la prosodie des enregistrements originaux sur la prosodie des locuteurs reproduisant les dialogues. Une autre option aurait consisté à prononcer nous-même les tours de

méthode lors du premier enregistrement. Afin de ne pas introduire de biais dans l'expérience, nous avons décidé de ne pas employer cette méthode pour les enregistrements suivants. Nous avons simplement demandé aux locuteurs de faire une pause à chaque tour de parole de l'autre locuteur, et d'imaginer qu'ils entendaient ce tour de parole.

L'expérience s'est déroulée en deux parties :

- **tâche de lecture à voix haute** : Les locuteurs ont découvert le texte lors de cette tâche. Ils ont été autorisés à le lire silencieusement avant que la tâche ne commence. Ils ont reçu pour consigne de dire le texte comme s'ils étaient le personnage et participaient réellement à la conversation. Nous leur avons demandé de respecter le mieux possible les mots exacts du texte (y compris les disfluences). Pendant l'enregistrement, nous avons communiqué avec les locuteurs depuis la régie. En cas d'erreur dans le texte, nous leur avons demandé de reprendre la phrase depuis le début.

- **tâche d'interprétation** : Les locuteurs ont effectué cette tâche environ une semaine après la tâche de lecture. Ce délai était nécessaire afin qu'ils puissent mémoriser le texte. Nous avons demandé aux locuteurs de mémoriser le texte au mot près dans la mesure du possible. Nous leur avons également demandé d'être prêts à en donner une véritable interprétation, c'est-à-dire pas une restitution neutre mais une version personnelle contenant de la spontanéité et de l'expressivité. Lors de l'enregistrement, les locuteurs n'ont pas été autorisés à conserver le texte sous les yeux pour s'aider. Nous leur avons demandé d'enregistrer deux versions du texte : d'abord une restitution neutre (une « italienne » dans le jargon théâtral), puis l'interprétation proprement dite¹². Comme dans la tâche de lecture, nous avons demandé aux locuteurs de respecter le mieux possible les mots exacts du texte, et de reprendre la phrase depuis le début en cas d'erreur.

parole lors de la tâche de reproduction. Cependant, un biais aurait alors été introduit, puisque de la variation prosodique aurait forcément eu lieu d'un locuteur à l'autre.

¹² L'enregistrement de la restitution neutre avait pour but d'être en mesure d'identifier de manière plus précise le facteur de variation prosodique : le processus de mémorisation, ou la simulation de spontanéité et d'expressivité. En effet, tandis que le premier facteur concerne à la fois la restitution neutre et l'interprétation, le second facteur concerne uniquement l'interprétation. Cette question n'est pas explorée dans cette thèse mais fera peut-être l'objet d'une étude ultérieure.

4.2. Relevé des occurrences de focalisation

Le relevé des occurrences de focalisation a été effectué par un groupe d'experts en prosodie. Les experts ont annoté les productions spontanées, lues et interprétées en se référant tous aux mêmes consignes. Les annotations des experts ont ensuite été recoupées afin de ne conserver que les occurrences de focalisation ayant été relevées par une majorité d'experts.

4.2.1. Sélection des experts

Nous avons sollicité dix experts pour annoter le corpus de l'expérience de production. Les experts ont relevé les occurrences de focalisation prosodique et ont également établi une classification fonctionnelle de ces occurrences (cf chap. 3, § 4.2.2). Le recours à ces experts était nécessaire afin que l'annotation du corpus ait un caractère objectif, ce qui n'aurait pas été le cas si nous l'avions effectuée nous-même. Tous les experts possèdent une expérience de l'étude de la prosodie (bien qu'ils ne possèdent pas tous le même niveau et ne soient pas tous phonéticiens ou phonologues). Nous présentons ci-dessous (cf Tableau 7) le niveau de chaque expert, ainsi que son domaine de spécialisation.

expert	niveau	domaine
experte 1	enseignante-chercheuse	phonétique et phonologie
expert 2	doctorant	pragmatique
expert 3	enseignant-chercheur	phonétique et phonologie
experte 4	étudiante en Master	phonétique et phonologie
expert 5	doctorant	phonétique et phonologie
expert 6	enseignant-chercheur	phonétique et phonologie
experte 7	étudiante en Master	phonétique et phonologie
experte 8	doctorante	phonétique et phonologie
expert 9	étudiant en Master	phonétique et phonologie
experte 10	enseignante-chercheuse	pragmatique

Tableau 7. Informations sur les experts en prosodie qui ont relevé les occurrences de focalisation prosodique dans le corpus de l'expérience.

4.2.2. Annotation du corpus

Le corpus comprend en tout dix enregistrements : deux productions spontanées, quatre versions lues et quatre versions interprétées. Chaque enregistrement a été annoté par quatre experts en prosodie. Idéalement, les enregistrements auraient tous dû être annotés par les mêmes experts. Cependant, cela aurait nécessité que chaque expert annote tous les enregistrements, ce qui était impossible pour des raisons pratiques. Nous indiquons ci-dessous (cf Tableau 8) la répartition des enregistrements parmi les experts.

phonogène	locuteur	groupe d'experts
parole spontanée	locuteur A	experts 1, 2, 5 et 6
	locutrice B	
parole lue	locuteur A1	experts 1, 2, 7 et 8
	locutrice B1	
	locuteur A2	experts 3, 4, 7 et 8
	locutrice B2	
parole interprétée	locuteur A1	experts 3, 4, 9 et 10
	locutrice B1	
	locuteur A2	experts 5, 6, 9 et 10
	locutrice B2	

Tableau 8. Répartition des enregistrements du corpus parmi les experts en prosodie pour l'annotation.

Les experts ont reçu les consignes suivantes :

- relever les occurrences de focalisation prosodique réalisées au moyen des traits suivants (cf § 2.2), isolément ou en combinaison :
 - augmentation de hauteur, de durée ou d'intensité sur une partie ou la totalité du constituant focalisé
 - accentuation sur une des syllabes initiales du constituant focalisé
 - contour intonatif terminal à la fin du constituant focalisé
 - abaissement et rétrécissement du registre tonal sur la séquence précédant ou la séquence suivant le constituant focalisé
- annoter les enregistrements dans n'importe quel ordre et sans utiliser de logiciel d'analyse acoustique (ou bien en masquant le spectrogramme, la courbe de fréquence fondamentale et la courbe d'intensité¹³)

¹³ Huit des dix experts ont choisi cette option, ce qui leur a permis d'effectuer l'annotation en utilisant les alignements texte-parole que nous avons réalisés pour chaque enregistrement (cf § 4.3.1). Notons que cela a pu

- ne pas relever les accents purement rythmiques, c'est-à-dire ayant pour seule fonction de marquer la frontière initiale ou finale d'un groupe prosodique¹⁴

Nous avons ensuite recoupé entre elles les annotations des experts, afin de conserver les occurrences de focalisation ayant été relevées par le plus d'experts¹⁵. Nous avons choisi de conserver les occurrences ayant été relevées par au moins trois experts sur quatre. On peut voir ci-dessous (cf Figure 14) le début d'un fichier contenant les annotations des experts pour un des enregistrements. Il s'agit de l'énoncé que nous avons cité plusieurs fois en exemple au cours du premier chapitre. On voit que trois experts ont relevé une focalisation sur le mot *cours*. Cette occurrence de focalisation est donc conservée pour l'analyse prosodique.

influencer en partie leur jugement, car ces alignements montrent (bien que de manière approximative et non-quantifiée) la durée relative des syllabes les unes par rapport aux autres (cf Figure 14 et Figure 15 plus bas). Nous avons néanmoins autorisé les experts à utiliser cette option, en raison de sa plus grande commodité. Les deux autres experts ont effectué leur annotation à l'intérieur de la transcription orthographique des enregistrements.

¹⁴ Ces accents ne rentrent en effet pas dans la définition que nous avons adoptée de la focalisation prosodique (le soulignement d'un constituant pour toute raison sémantico-pragmatique ou emphatique). Par ailleurs, sans cette précision, les experts risquaient de relever un nombre trop important de focalisations.

¹⁵ Un problème d'unité s'est présenté concernant le recoupement des annotations. En effet, pour relever les occurrences de focalisation, quatre experts ont annoté des mots, tandis que les autres experts ont annoté soit des mots, soit des constituants plus larges. Idéalement, une consigne précise aurait dû être donnée aux experts à ce sujet. Afin de résoudre le problème, nous avons décidé de choisir comme unité la syllabe. Il s'agit en effet de l'unité que nous avons utilisée dans l'analyse prosodique. Nous avons converti les annotations des experts à cet effet.

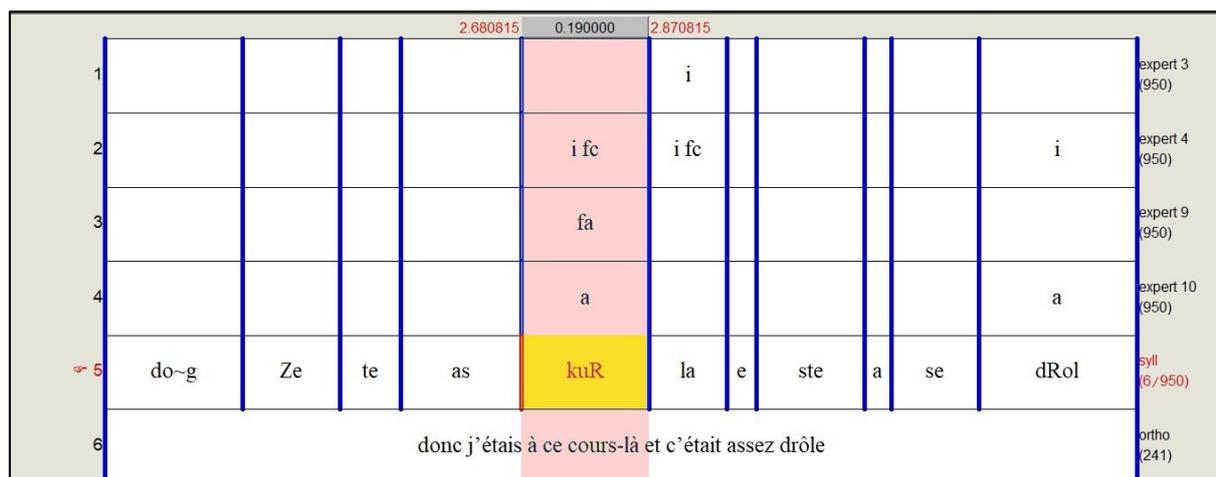


Figure 14. Capture d'écran du logiciel Praat contenant l'alignement texte-parole d'un énoncé, ainsi que les annotations des experts. En bas, la tire orthographique contient la transcription orthographique de l'énoncé. La tire juste au-dessus contient la segmentation syllabique de l'énoncé (les syllabes sont transcrites phonétiquement à l'aide de l'alphabet SAMPA). Les tires du dessus contiennent les annotations des quatre experts. Les lettres indiquent la fonction assignée par les experts aux occurrences de focalisation (« i » : insistance, « fc » : focus contrastif, « fa » : focus associatif, « a » : autre fonction).

4.3. Analyse prosodique

4.3.1. Alignement texte-parole

Cette étape est nécessaire afin de pouvoir effectuer des mesures phonétiques sur le corpus de manière automatisée. Elle permet également à l'expérimentateur de se repérer facilement à l'intérieur des enregistrements lorsque ceux-ci sont de longue durée. L'alignement consiste à segmenter les enregistrements en unités (par exemple des mots ou des syllabes) sur un logiciel d'analyse acoustique. On fait ensuite correspondre chaque unité avec sa transcription orthographique ou phonétique. Le résultat de l'alignement est illustré ci-dessous (cf Figure 15) en reprenant l'énoncé de la figure précédente.

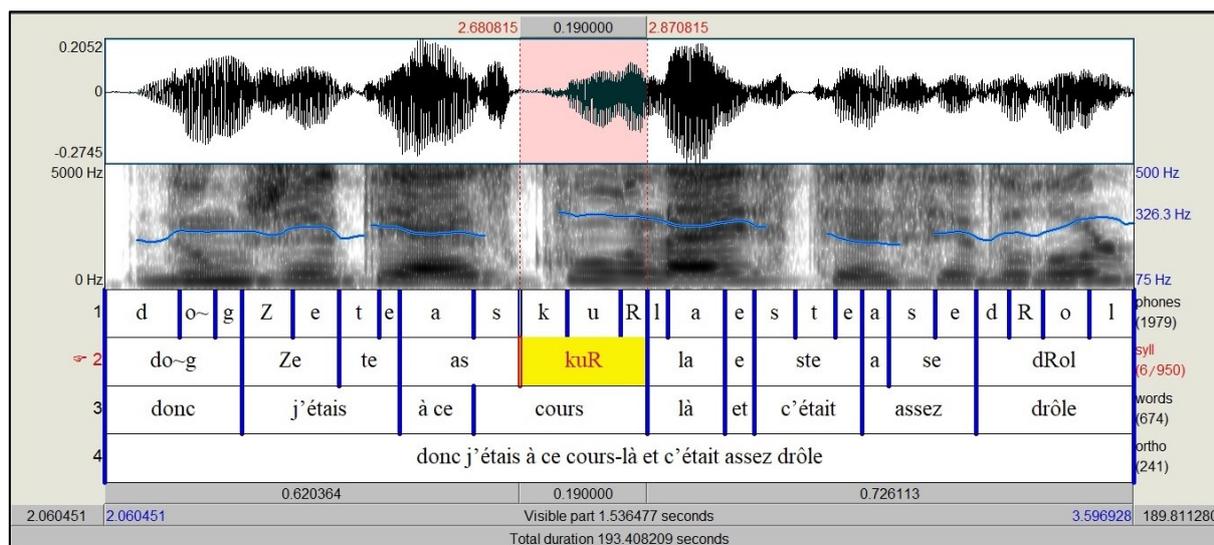


Figure 15. Capture d'écran du logiciel Praat contenant les informations acoustiques d'un énoncé, ainsi que sa segmentation en mots, syllabes et phones obtenue à l'aide de l'extension EasyAlign. La fenêtre du haut contient l'oscillogramme de l'énoncé. La fenêtre en-dessous contient le spectrogramme, ainsi que la courbe de fréquence fondamentale de l'énoncé (en bleu). En-dessous se trouvent les quatre tires de segmentation de l'énoncé. La syllabe [kuR] a été sélectionnée, ce qui permet d'obtenir facilement sa durée (indiquée tout en haut) ainsi que d'effectuer d'autres mesures.

Nous avons segmenté les enregistrements du corpus en énoncés, en mots, en syllabes et en phones. Nous avons effectué la segmentation sur le logiciel Praat (cf Boersma 2002) à l'aide de l'extension EasyAlign (cf Goldman 2011). Actuellement développé pour le français, l'espagnol et le portugais (et partiellement pour l'anglais), EasyAlign permet de réaliser la segmentation de manière automatique à partir de la transcription orthographique de l'enregistrement. Bien que l'utilisateur doive ensuite effectuer une correction manuelle, le processus est beaucoup moins long que si on segmentait l'enregistrement manuellement. EasyAlign fonctionne en trois étapes :

- **macro-segmentation** : L'enregistrement est segmenté en énoncés ou en groupes intonatifs. L'utilisateur indique au préalable les frontières entre unités à l'intérieur de la transcription orthographique. Un algorithme estime la position de ces frontières dans l'enregistrement en se basant sur la durée de celui-ci, ainsi que sur la longueur de la transcription orthographique. L'algorithme privilégie par ailleurs les pauses, qu'il détecte automatiquement.

- **conversion phonétique** : La transcription orthographique est convertie en transcription phonétique (en alphabet SAMPA). Cette conversion est effectuée à l'aide d'un dictionnaire intégré, ainsi que d'un ensemble de règles de prononciation (tous deux issus d'un système de synthèse de la parole). Afin de faciliter l'étape suivante, les liaisons et les schwas optionnels sont automatiquement indiqués. L'utilisateur a la possibilité de corriger manuellement la transcription phonétique.
- **segmentation** : La transcription phonétique est alignée, phone par phone, avec la portion de son correspondante dans l'enregistrement. L'enregistrement est également segmenté en mots. Ces deux opérations sont effectuées à l'aide d'un moteur de reconnaissance de la parole qui repère les frontières temporelles de chaque phone et de chaque mot. Enfin, une segmentation en syllabes est obtenue à partir de la segmentation en phones. Pour effectuer la syllabification, l'algorithme applique le principe selon lequel chaque syllabe doit forcément contenir une voyelle et une seule. Pour le découpage syllabique des groupes consonantiques, l'algorithme applique le principe de sonorité (cf Clements 1990).

4.3.2. Mesures de hauteur et de durée

Nous avons mesuré la hauteur moyenne et la durée des occurrences de focalisation prosodique¹⁶. Ces mesures ont été effectuées de manière automatisée sur Praat à l'aide de l'extension Prosogram (cf Mertens 2004). Prosogram permet d'obtenir une série de mesures sur toutes les syllabes du corpus, à partir d'une segmentation préalable du corpus en syllabes et en phones. Nous avons fourni à Prosogram la segmentation du corpus que nous avons obtenue à l'aide d'EasyAlign (comme décrit dans la section précédente). Les mesures effectuées par Prosogram incluent notamment la hauteur moyenne de la syllabe, sa hauteur minimale, sa hauteur maximale, son pic d'intensité, sa durée, la durée de la voyelle et la durée de la rime. Les mesures de hauteur sont uniquement effectuées sur les noyaux vocaliques des

¹⁶ Nous n'avons pas mesuré l'intensité moyenne des occurrences de focalisation prosodique. En effet, ces mesures n'auraient pas été entièrement fiables. Ceci est dû au fait que la distance entre le micro et le locuteur n'a pas été contrôlée lors de nos enregistrements, ce qui a pu créer des variations d'intensité indépendantes de la production des locuteurs. Pour contrôler la distance micro-locuteur, il aurait fallu employer un matériel très contraignant pour les locuteurs, ce qui les aurait gênés dans leurs tâches de production (et aurait donc nui à notre étude).

voyelles. Pour déterminer les noyaux vocaliques, Prosogram sélectionne la portion voisée du pic d'intensité de la voyelle. La portion voisée est déterminée de la façon suivante : il s'agit de la portion qui commence à l'instant où l'intensité est à 3 dB en-dessous du pic, et qui se termine à l'instant où l'intensité est à 9 dB en-dessous du pic. Cette méthode a pour avantage d'éliminer les variations micro-prosodiques (négligeables) en début de voyelle, et de conserver les variations de hauteur à la fin des voyelles accentuées (qui sont, elles, pertinentes).

La hauteur moyenne et la durée syllabique ont été normalisées par rapport à plusieurs paramètres :

- La hauteur moyenne a été normalisée par rapport aux locuteurs. Pour effectuer cette normalisation, nous avons converti la hauteur moyenne en demi-tons. Contrairement à l'échelle hertzienne, l'échelle tonale est indépendante des différences de registre tonal entre les locuteurs.
- La durée syllabique a été normalisée par rapport à la structure syllabique. Cette normalisation était nécessaire car, toutes choses égales par ailleurs, une syllabe de structure complexe (contenant par exemple un groupe consonantique en attaque ou en coda) a tendance à être plus longue qu'une syllabe ayant une structure plus simple. Pour effectuer cette normalisation, nous avons divisé la durée de chaque syllabe par le nombre de phones de la syllabe.
- La durée syllabique a également été normalisée par rapport au débit de parole des locuteurs. Pour effectuer cette normalisation, nous avons, pour chaque locuteur, standardisé en scores-z les valeurs de durée préalablement normalisées par rapport à la structure syllabique. La standardisation en scores-z a été effectuée sur le logiciel R (cf R Core Team 2018), à l'aide de la fonction générique *scale*.

4.3.3. Catégories phonologiques

Nous avons effectué une classification des occurrences de focalisation prosodique selon trois catégories :

- **le type de contour intonatif** : La classification a été effectuée de manière auditive, ainsi qu'en visualisant la courbe de hauteur des occurrences de focalisation prosodique sur le logiciel Praat. Nous nous sommes basé sur le système de transcription prosodique ToBI, dans sa version adaptée au français (F-ToBI, cf Delais-Roussarie et al 2015). Cette version a été présentée dans le premier chapitre (cf § 3.3). Pour rappel, les mouvements intonatifs sont représentés dans ToBI par des tons hauts ou bas (notés respectivement H ou L). Trois niveaux de la structure prosodique sont postulés : le groupe accentuel, le groupe intermédiaire et le groupe intonatif. Le groupe accentuel se termine par un accent mélodique (noté H* ou L*) et peut contenir un ton haut initial¹⁷ (noté Hi). Le groupe intonatif se termine par un ton de frontière (noté H% ou L%) qui est plus ample que l'accent mélodique et qui s'accompagne d'un allongement plus important. Le groupe intermédiaire se termine par un ton phrastique (noté H- ou L-), dont l'amplitude et l'allongement se situent entre ceux de l'accent mélodique et du ton de frontière. Nous avons retenu onze contours intonatifs possibles¹⁸ pour les constituants focalisés (ces contours sont illustrés ci-dessous à l'aide d'occurrences de focalisation tirées du corpus de l'expérience) :
- ton montant H*H% ou H*H- (cf Figure 16)
 - ton descendant L*L% ou L*L- (cf Figure 17)
 - ton montant L*H% ou L*H- (cf Figure 18)
 - ton montant-descendant H*L% ou H*L- (cf Figure 19)
 - ton montant H* (cf Figure 20)
 - ton descendant L* (cf Figure 20)
 - ton haut initial Hi (cf Figure 20)

¹⁷ Ce contour concerne les cas où le constituant focalisé comporte une accentuation initiale marquant le début d'un groupe accentuel. Le ton de fin de groupe accentuel se trouve en-dehors du constituant délimité par l'annotation des experts (dans l'exemple présenté dans la Figure 20, le ton haut Hi se trouve sur *un* et le ton de fin de groupe accentuel tombe à la fin de *samedi*).

¹⁸ Idéalement, un douzième contour aurait également dû être retenu parmi les contours possibles. Il s'agit du contour montant-descendant H+H*H%, avec montée sur la syllabe pénultième. Toutefois, ce contour ne concerne que certains types d'énoncé (énoncés vocatifs ou expression d'une contradiction, cf Delais-Roussarie et al 2015).

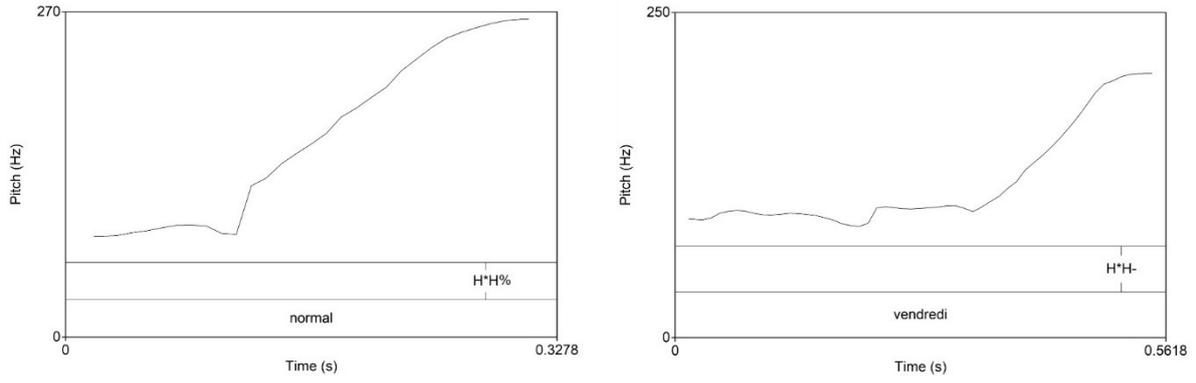


Figure 16. Focalisations issues du corpus comportant un contour montant H*H% (à gauche) et H*H- (à droite).

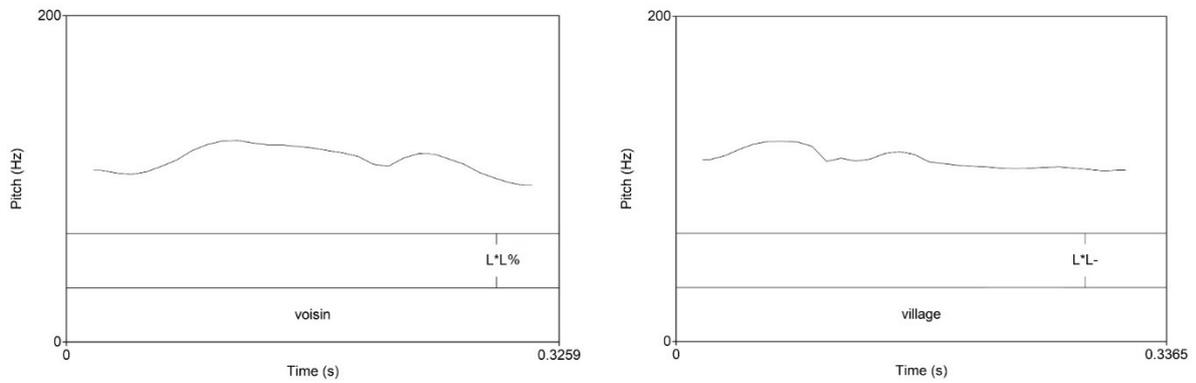


Figure 17. Focalisations issues du corpus comportant un contour descendant L*L% (à gauche) et L*L- (à droite).

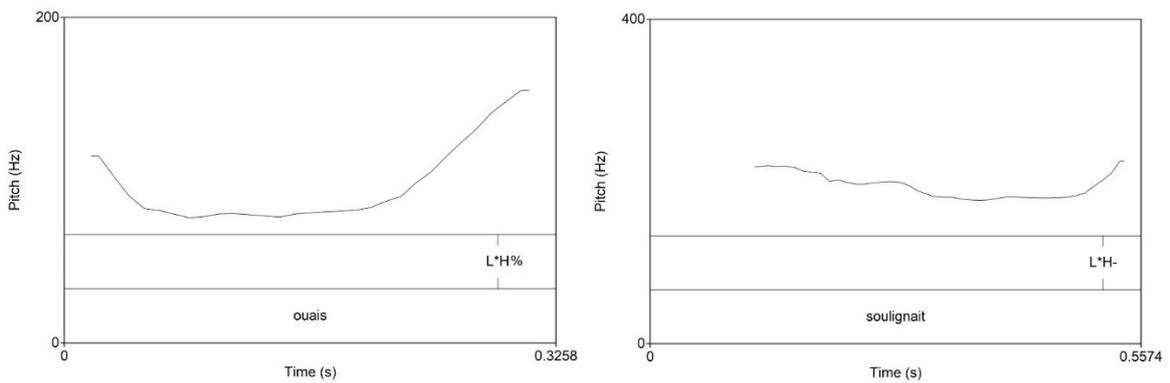


Figure 18. Focalisations issues du corpus comportant un contour montant L*H% (à gauche) et L*H- (à droite).

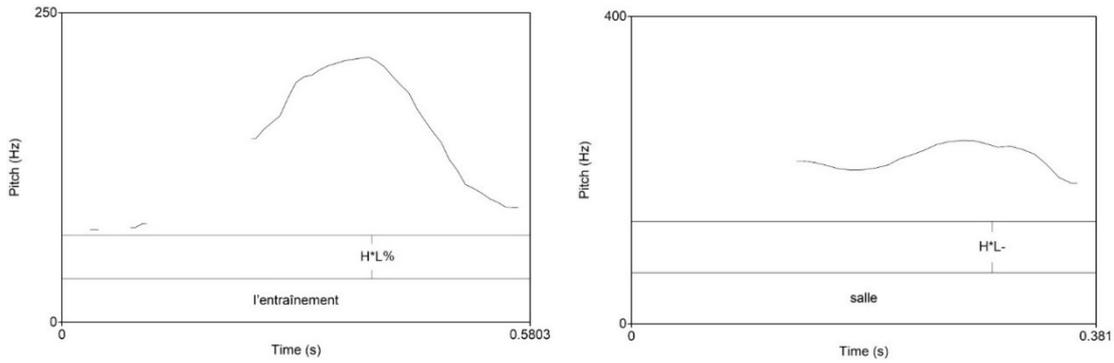


Figure 19. Focalisations issues du corpus comportant un contour montant-descendant $H^*L\%$ (à gauche) et H^*L- (à droite).

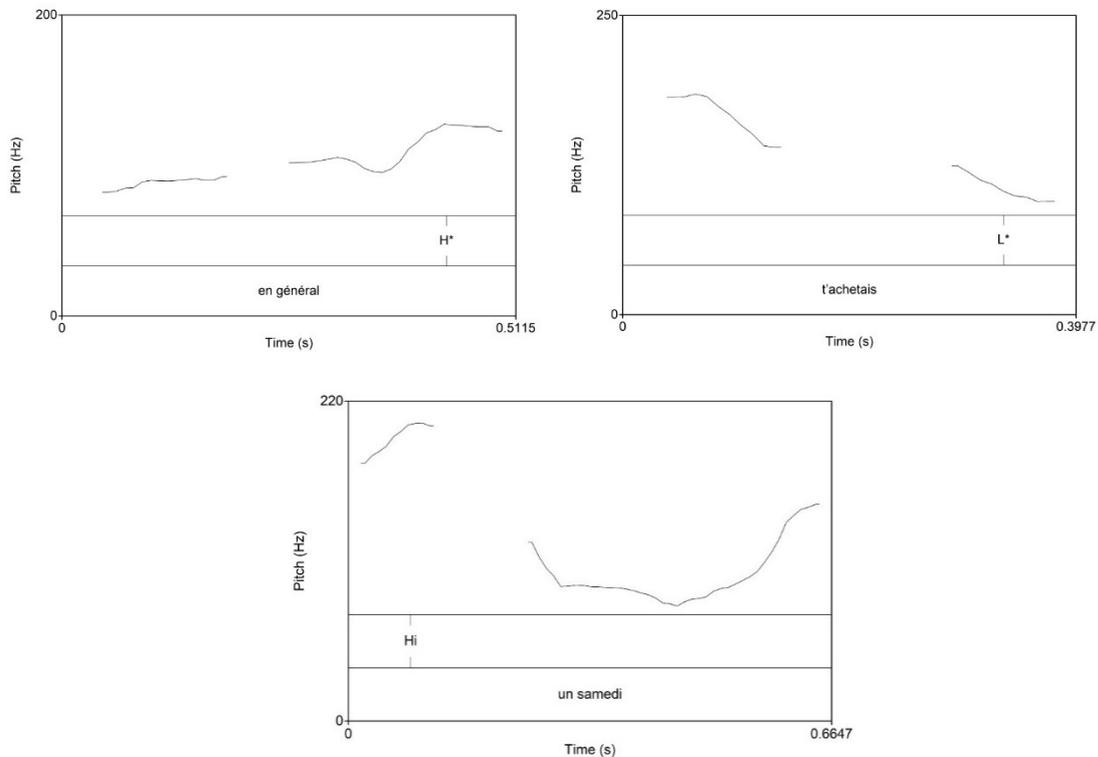


Figure 20. Focalisations issues du corpus comportant un contour montant H^* (en haut à gauche), un contour descendant L^* (en haut à droite) et un ton haut initial Hi (en bas).

- **l'étendue syllabique du contour intonatif** : Cette catégorie concerne les cas où le constituant focalisé comporte plusieurs syllabes. Elle désigne la ou les syllabes du constituant qui portent le contour. Comme pour le type de contour intonatif, la classification a été effectuée de manière auditive ainsi qu'en visualisant la courbe de

hauteur des occurrences de focalisation prosodique sur Praat. Nous avons identifié dans le corpus six types possibles d'étendue du contour intonatif (cf Figure 21) :

- dernière syllabe
- deux dernières syllabes
- première syllabe
- deux premières syllabes
- deuxième syllabe
- constituant entier

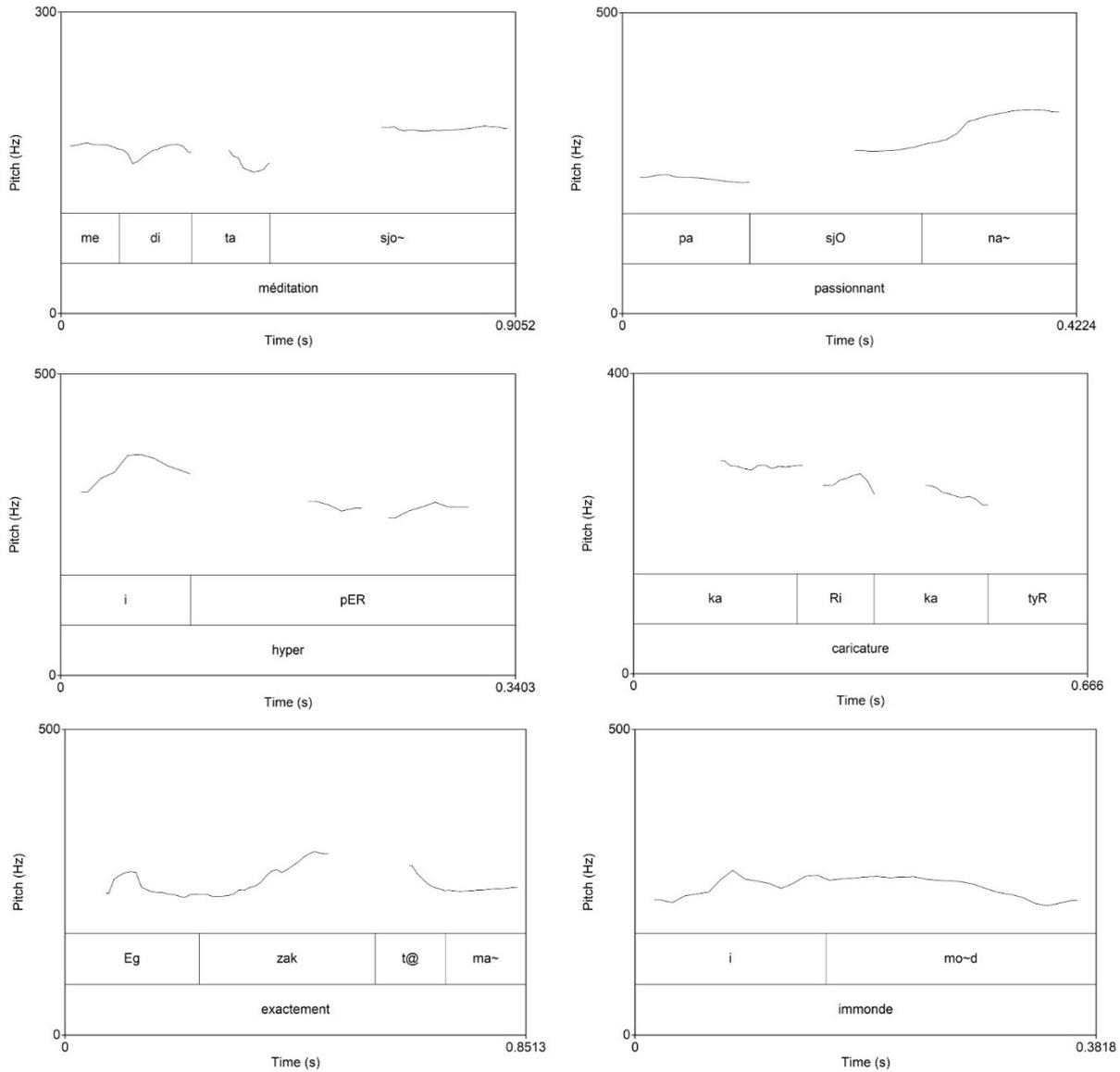


Figure 21. Focalisations issues du corpus présentant différentes étendues syllabiques du contour intonatif. De gauche à droite et de haut en bas, le contour est porté par la dernière syllabe (méditation), les deux dernières syllabes (passionnant), la première syllabe (hyper), les deux premières syllabes (caricature), la deuxième syllabe (exactement) et le constituant entier (immonde).

- **la présence d'accent initial** : Pour déterminer la présence d'un accent initial¹⁹ sur le constituant focalisé, nous nous sommes basé sur une détection automatique des

¹⁹ Notons que ce trait est différent du ton haut initial Hi, qui fait partie des contours intonatifs possibles sur le constituant focalisé. Comme nous l'avons expliqué plus haut (cf note 17), le ton Hi marque la frontière initiale d'un groupe de mots qui est plus large que le constituant focalisé relevé par les experts. Un constituant qui porte un ton Hi porte aussi, par définition, un accent initial. Cependant, un accent initial peut également être porté par un constituant qui porte un autre contour que le ton Hi.

proéminences dans le corpus obtenue à l'aide du logiciel Anamor (cf Avanzi et al 2008). La détection est effectuée par Anamor à partir d'une segmentation préalable de l'enregistrement en syllabes et en phones (nous lui avons fourni la segmentation obtenue à l'aide d'EasyAlign, cf § 4.3.1). La notion de proéminence d'Anamor est plus large que la notion de focalisation prosodique que nous avons adoptée (cf § 2.1), car elle comprend également les accents rythmiques (marquant la frontière initiale ou finale d'un groupe prosodique). Anamor se base sur les paramètres acoustiques suivants, tirés des recherches récentes sur la proéminence en français (cf Avanzi 2012 : 95-96) : la hauteur et la durée relatives de la syllabe par rapport aux syllabes environnantes²⁰, la valeur du glissando montant²¹ (s'il est présent), et la présence d'une pause silencieuse après la syllabe. Le caractère proéminent ou non d'une syllabe est déterminé en attribuant une note à chaque paramètre (en fonction d'un seuil fixé). Un degré de force de 0 à 10 est également attribué à la syllabe en calculant la moyenne des quatre notes. On peut voir ci-dessous (cf Figure 22) les proéminences détectées par Anamor dans l'énoncé que nous avons déjà vu plus haut (cf Figure 14 et Figure 15).

²⁰ Plus précisément, la hauteur et la durée de la syllabe sont comparées à la moyenne de la hauteur et de la durée des trois syllabes qui précèdent et des trois syllabes qui suivent. Ce choix se base sur le fait qu'il est souvent considéré que le groupe accentuel contient un nombre maximal de sept syllabes en français (e.g. Wioland 1985, Martin 2015 : 58).

²¹ Le glissando désigne le degré d'inclinaison de la pente de la courbe de fréquence fondamentale.

(cf § 4.1.2). Deuxièmement, les locuteurs des productions lues et interprétées n'ont pas toujours parfaitement respecté le texte, malgré nos consignes en ce sens.

- Les fréquences des types syllabiques diffèrent en partie des moyennes observées pour le français (cf Léon 2007 : 133). Si le type CV est bien le plus fréquent en français (59,9%), il est suivi du type CVC (17,1%) et non du type V, puis du type CCV (14,2%). Notons que la structure syllabique peut avoir une influence sur les mesures de durée. Toutefois, comme nous l'avons mentionné plus haut (cf § 4.3.2), nous avons normalisé nos mesures de durée par rapport à ce paramètre.

		parole spontanée	parole lue	parole interprétée	corpus entier
nombre de mots		1139	2246	2258	5643
nombre de syllabes		1533	3127	3140	7800
fréquence du type de syllabe	CV	57,5%	56,0%	56,6%	56,6%
	V	16,3%	14,2%	14,6%	14,8%
	CVC	12,3%	14,2%	13,7%	13,6%
	CCV	9,7%	10,0%	9,6%	9,8%
	autres	4,2%	5,6%	5,5%	5,2%

Tableau 9. Informations globales sur le corpus.

5.1. Taux d'accord entre les experts

Nous avons mesuré le taux d'accord entre les experts en prosodie concernant le relevé des occurrences de focalisation. Nous avons employé le test statistique du kappa de Fleiss (cf Fleiss 1971). Ce test permet d'évaluer le taux d'accord entre plus de deux experts, contrairement au test du kappa de Cohen (cf Cohen 1960) qui ne permet de l'évaluer que pour deux experts. Nous avons obtenu les kappas sous le logiciel R, à l'aide de la fonction *kappam.fleiss* de la bibliothèque « irr » (Gamer et al 2012). Nous indiquons ci-dessous (cf Tableau 10) le taux

d'accord pour chaque phonogène ainsi que pour le corpus entier²². On voit que tous les kappas sont significatifs d'après leur score-z²³.

	parole spontanée	parole lue	parole interprétée	corpus entier
taux d'accord (kappa de Fleiss)	0,374 (z = 36,1)	0,230 (z = 31,5)	0,285 (z = 39,2)	0,284 (z = 61,6)

Tableau 10. Taux d'accord entre les experts en prosodie concernant le relevé des occurrences de focalisation. Les scores-z sont indiqués sous les kappas. Tous les scores-z sont significatifs à $p < 0,001$.

On voit que le taux d'accord varie selon le phonogène. Il est le plus haut en parole spontanée (0,374), suivie de la parole interprétée (0,285) et de la parole lue (0,230). Il est en revanche plus difficile d'évaluer le taux d'accord global pour le corpus entier (0,284). En effet, il n'existe pas de référence universellement admise pour l'interprétation du kappa. Landis et Koch (1977) proposent une échelle²⁴ à laquelle se réfèrent beaucoup d'études. Cependant, selon Sim et Wright (2005), les seuils choisis dans cette échelle sont arbitraires, et la valeur du kappa dépend par ailleurs de plusieurs facteurs tels que le nombre de catégories analysées (plus il y a de catégories, plus le kappa est faible). Afin d'évaluer le taux d'accord pour le corpus entier, nous avons choisi de le comparer au taux d'accord obtenu dans d'autres études sur la prosodie dans lesquelles le test du kappa de Fleiss est également utilisé. Deux études récentes, Cole et al (2010) et Avanzi et al (2011), obtiennent des kappas respectifs de 0,54 et 0,67. Par comparaison, dans notre étude, le taux d'accord pour le corpus entier (0,284) est donc

²² Comme nous l'avons mentionné plus haut (cf § 4.2.2), chaque enregistrement n'a pas été annoté par le même groupe d'experts. Le test du kappa de Fleiss permet cependant de calculer le taux d'accord pour plusieurs enregistrements à la fois, même si les experts varient d'un enregistrement à l'autre.

²³ Pour que le kappa soit significatif à un degré de confiance de 95%, le score-z doit être supérieur à 1,96 en valeur absolue (ce qui correspond à une valeur-p d'un maximum de 0,05). On voit que c'est largement le cas ici.

²⁴ L'échelle de Landis et Koch (1977) distingue six degrés d'accord : mauvais (inférieur ou égal à 0,00), léger (de 0,00 à 0,20), acceptable (de 0,21 à 0,40), modéré (de 0,41 à 0,60), important (de 0,61 à 0,80) et presque parfait (de 0,81 à 1,00).

relativement faible. Comme nous le verrons plus bas (cf § 5.4), nous avons décidé pour cette raison de répéter l'étude sur un corpus réduit constitué des enregistrements ayant obtenu les meilleurs taux d'accord.

5.2. Fréquence d'occurrence

Nous indiquons ci-dessous (cf Tableau 11) la fréquence d'occurrence de focalisation prosodique pour chaque phonogène ainsi que pour le corpus entier. La fréquence d'occurrence a été déterminée en calculant le pourcentage de syllabes focalisées. Nous avons décidé de calculer le pourcentage de syllabes focalisées (et non de mots) car la syllabe est l'unité qui a été choisie pour recouper les annotations du corpus par les experts en prosodie (cf § 4.2.2, note 15). Le pourcentage de syllabes focalisées dans tout le corpus est de 11,27%. On voit que le pourcentage varie selon le phonogène. La parole interprétée a le plus haut pourcentage (14,17%), suivie de la parole lue (10,65%) et de la parole spontanée (6,59%).

	parole spontanée	parole lue	parole interprétée	corpus entier
pourcentage de syllabes focalisées	6,59%	10,65%	14,17%	11,27%

Tableau 11. Pourcentage de syllabes focalisées pour chaque phonogène et dans le corpus entier.

Nous avons employé une régression logistique afin de déterminer la significativité des différences de pourcentage de syllabes focalisées entre les phonogènes. La régression logistique a été effectuée sous le logiciel R, à l'aide de la fonction *glmer* de la bibliothèque « lme4 » (Bates et Maechler 2009). La variable dépendante est la présence de focalisation sur la syllabe, et la variable fixe est le phonogène. Nous avons inclus le locuteur en tant que variable aléatoire²⁵. Les valeurs-p ont été obtenues en effectuant un test du rapport des vraisemblances entre le modèle ainsi qu'un modèle ne contenant pas la variable du phonogène. Les résultats révèlent un effet significatif du phonogène sur la présence de focalisation ($\chi^2(2) = 22,7, p < 0,001$). Nous indiquons ci-dessous (cf Tableau 12) les résultats de la régression. On voit, d'après les scores-z et les valeurs-p²⁶, que les différences entre la parole interprétée et les deux autres phonogènes sont significatives. La différence entre parole spontanée et parole lue est, quant à elle, marginale.

²⁵ Idéalement, le groupe d'experts aurait également dû être inclus en tant que variable aléatoire. Cependant, cela n'est pas possible car cette variable est en partie confondue avec la variable du locuteur. En effet, chaque groupe d'experts n'a pas annoté des enregistrements de tous les locuteurs (cf § 4.2.2). Dans cette régression logistique ainsi que dans les modèles suivants, nous avons donc soit uniquement inclus le locuteur en tant que variable aléatoire, soit employé des modèles sans variable aléatoire (cf § 5.3.2).

²⁶ Pour que la différence de pourcentage entre deux phonogènes soit significative à 95%, le score-z doit être supérieur à 1,96 en valeur absolue, ce qui correspond à une valeur-p d'un maximum de 0,05.

	estimation (et erreur standard)	score-z	valeur-p
parole spontanée/lue	-0,50 (0,27)	-1,837	0,0662
parole spontanée/ interprétée	-0,83 (0,27)	-3,083	0,00205
parole lue/interprétée	-0,33 (0,08)	-4,255	0,0000209

Tableau 12. Résultats de la régression logistique concernant l'effet du phonogène sur le pourcentage de syllabes focalisées.

5.3. Réalisation prosodique

5.3.1. Hauteur

On voit ci-dessous (cf Figure 23) que les syllabes focalisées présentent, numériquement, une hauteur moyenne nettement supérieure à celle des syllabes non-focalisées. Nous avons employé un modèle linéaire mixte afin de déterminer la significativité de cette différence. Le modèle a été effectué sous le logiciel R, à l'aide de la fonction *lmer* de la bibliothèque « lme4 ». La variable dépendante est la hauteur, la variable fixe est la présence de focalisation sur la syllabe, et la variable aléatoire est le locuteur (cf note 23). La valeur-p a été obtenue en effectuant un test du rapport des vraisemblances entre le modèle ainsi qu'un modèle ne contenant pas la variable de la présence de focalisation. L'effet est significatif ($\chi^2(1) = 521,3$, $p < 0,001$). La différence de hauteur entre les syllabes focalisées et les syllabes non-focalisées est bien significative²⁷ ($t = 23,25$, $p < 0,001$).

²⁷ Comme avec les scores-z (cf note 26), pour que la différence de hauteur entre syllabes focalisées et non-focalisées soit significative à 95%, la valeur-t doit être supérieure à 1,96 en valeur absolue, ce qui correspond à une valeur-p d'un maximum de 0,05.

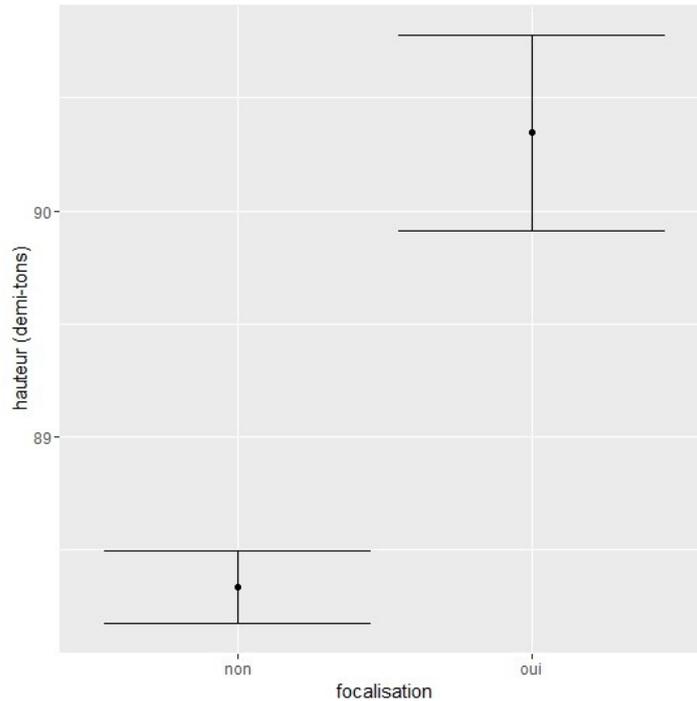


Figure 23. Hauteur moyenne des syllabes focalisées (« oui ») et des autres syllabes du corpus (« non »). Les barres d'erreur indiquent les intervalles de confiance à 95%.

Parmi les syllabes focalisées du corpus, on observe des différences de hauteur entre les trois phonogenres (cf Figure 24). On voit que les syllabes focalisées présentent, numériquement, une hauteur plus élevée en parole spontanée que dans les deux autres phonogenres. On n'observe en revanche pas de différence entre la parole lue et la parole interprétée. Afin de déterminer la significativité des différences, nous avons employé un modèle linéaire mixte avec la hauteur comme variable dépendante, le phonogenre comme variable fixe, et le locuteur comme variable aléatoire. L'effet n'est pas significatif ($\chi^2(2) = 0,3$, $p > 0,1$). Notons que le modèle révèle une forte variation inter-locuteurs. Cette variation justifie probablement l'absence de différence significative entre la parole spontanée et les deux autres phonogenres.

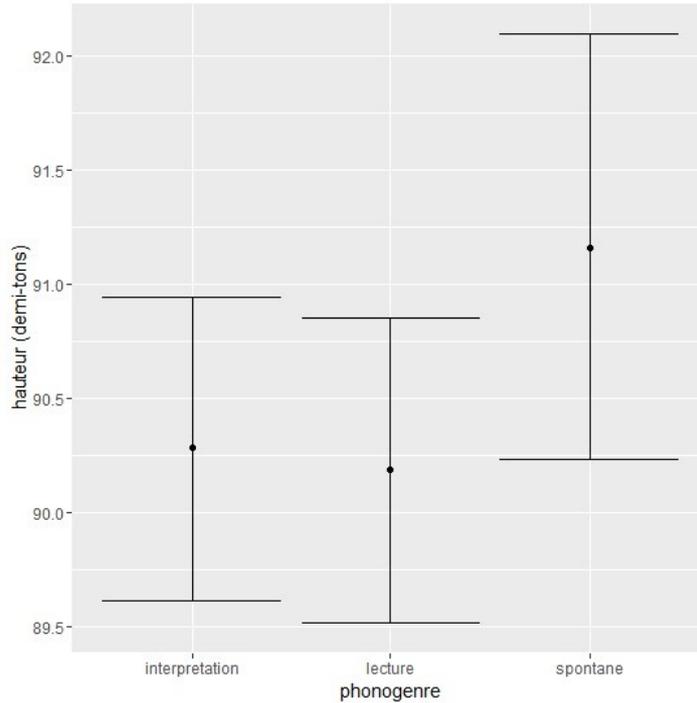


Figure 24. Hauteur moyenne des syllabes focalisées en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané »). Les barres d'erreur indiquent les intervalles de confiance à 95%.

5.3.2. Durée syllabique

On voit ci-dessous (cf Figure 25) que les syllabes focalisées présentent, numériquement, une durée nettement supérieure à celle des syllabes non-focalisées. Afin de déterminer la significativité de cette différence, nous avons employé un modèle linéaire mixte avec la durée comme variable dépendante et la présence de focalisation comme variable fixe. Nous n'avons pas inclus le locuteur comme variable aléatoire²⁸. En effet, la durée a déjà été normalisée par rapport aux locuteurs (cf § 4.3.2). L'effet est significatif ($\chi^2(1) = 71,8$, $p < 0,001$). La différence de hauteur entre les syllabes focalisées et les syllabes non-focalisées est bien significative ($t = 8,493$, $p < 0,001$).

²⁸ Etant donné que ce modèle linéaire mixte ne contient aucune variable aléatoire (cf note 25), nous ne l'avons pas effectué à l'aide de la fonction *lmer* (cf § 5.3.1) mais à l'aide de la fonction générique *glm* du logiciel R.

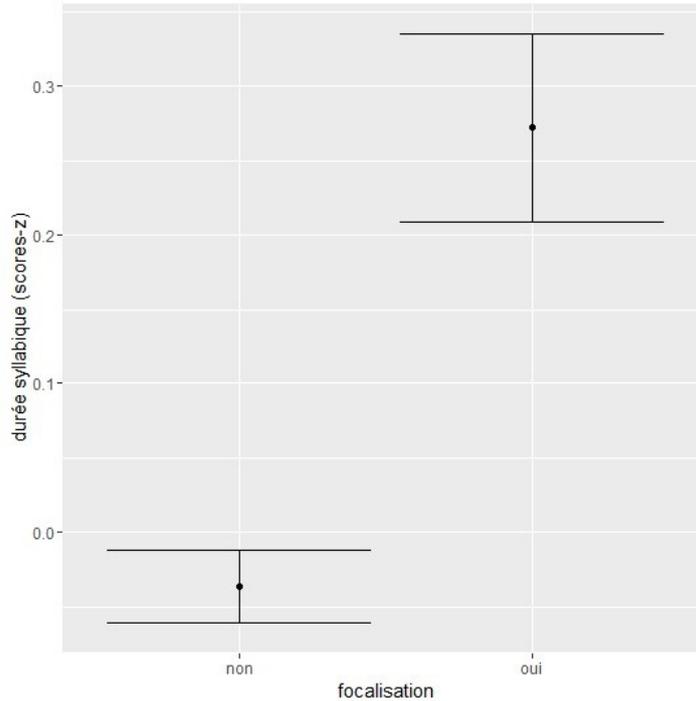


Figure 25. Durée moyenne des syllabes focalisées (« oui ») et des autres syllabes du corpus (« non »). Les barres d'erreur indiquent les intervalles de confiance à 95%. Étant donné que la durée syllabique a été normalisée par rapport aux locuteurs (cf § 4.3.2), son unité est le score-z.

Parmi les syllabes focalisées du corpus, on observe des différences de durée entre les trois phonogenres (cf Figure 26). On voit que les syllabes focalisées présentent, numériquement, une durée plus élevée en parole lue que dans les deux autres phonogenres. La différence entre la parole lue et la parole interprétée est particulièrement prononcée. Les syllabes focalisées présentent également une durée légèrement plus élevée en parole spontanée qu'en parole interprétée. D'après un modèle linéaire mixte (de nouveau sans variable aléatoire), l'effet du phonogène sur la durée syllabique est significatif ($\chi^2(2) = 9,1$, $p < 0,05$). Nous indiquons ci-dessous (cf Tableau 13) les résultats du modèle. On voit, d'après les valeurs-t et les valeurs-p, que la différence entre la parole lue et la parole interprétée est significative. Les autres différences entre phonogenres ne sont en revanche pas significatives.

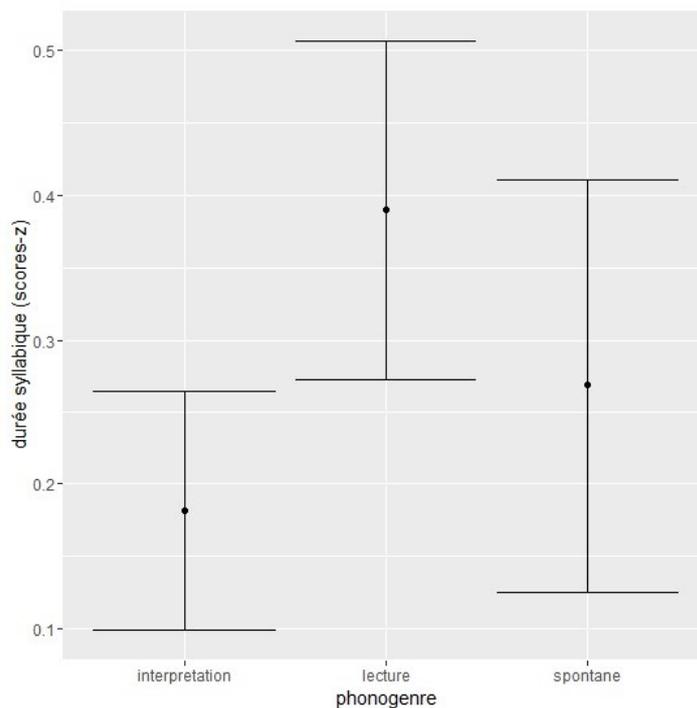


Figure 26. Durée moyenne des syllabes focalisées en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané »). Les barres d'erreur indiquent les intervalles de confiance à 95%. Etant donné que la durée syllabique a été normalisée par rapport aux locuteurs (cf § 4.3.2), son unité est le score-z.

	estimation (et erreur standard)	valeur-t	valeur-p
parole spontanée/lue	-0,12 (0,11)	-1,127	0,26000
parole spontanée/ interprétée	0,09 (0,11)	0,823	0,41053
parole lue/interprétée	0,21 (0,07)	3,023	0,00258

Tableau 13. Résultats du modèle linéaire mixte concernant l'effet du phonogène sur la durée des syllabes focalisées du corpus.

Nous indiquons ci-dessous (cf Figure 27) les différences globales de durée syllabique entre les trois phonogènes (cette fois toutes syllabes confondues). Nous avons mesuré ces différences car elles pourraient avoir une influence sur les différences de durée observées entre les

phonogenres sur les syllabes focalisées (cf Figure 26). On observe, numériquement, l'inverse de ce qu'on observe sur les syllabes focalisées. La parole interprétée présente des syllabes plus longues que les deux autres phonogenres (en particulier la parole lue). La parole spontanée présente également des syllabes plus longues que la parole lue. D'après un modèle linéaire mixte²⁹, l'effet du phonogène sur la durée syllabique est significatif ($\chi^2(2) = 9,0$, $p < 0,05$). Nous indiquons ci-dessous (cf Tableau 14) les résultats du modèle. On voit que la différence entre la parole lue et la parole interprétée est significative. Etant donné que cette différence est contradictoire avec la différence observée dans la figure précédente (où les syllabes focalisées présentent une durée plus élevée en parole lue qu'en parole interprétée), les deux différences ne peuvent pas être liées. Les autres différences entre phonogenres ne sont en revanche pas significatives.

²⁹ Contrairement au modèle précédent, ce modèle inclut le locuteur en tant que variable aléatoire. La variable dépendante est par conséquent la durée non-normalisée par rapport aux locuteurs. Le modèle a été effectué à l'aide de la fonction *lmer* sous le logiciel R (cf § 5.3.1). Ce choix est dû au fait que l'utilisation d'un modèle sans variable aléatoire (à l'aide de la fonction *glm*) a entraîné des valeurs-t toutes nulles. Ce résultat ne peut correspondre à la réalité et est forcément lié à une erreur du modèle.

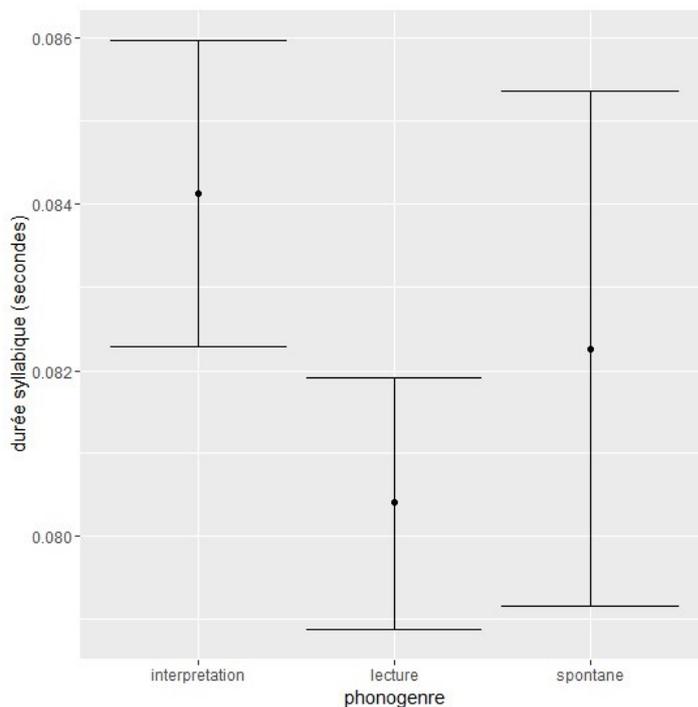


Figure 27. Durée moyenne des syllabes de parole interprétée (« interprétation »), de parole lue (« lecture ») et de parole spontanée (« spontané »). Les barres d'erreur indiquent les intervalles de confiance à 95%. La durée est indiquée en secondes car elle n'est pas normalisée par rapport aux locuteurs (cf note 29).

	estimation (et erreur standard)	valeur-t	valeur-p
parole spontanée/lue	0,001 (0,007)	0,121	0,9090
parole spontanée/ interprétée	-0,003 (0,007)	-0,418	0,6966
parole lue/interprétée	-0,004 (0,001)	-2,990	0,0028

Tableau 14. Résultats du modèle linéaire mixte concernant l'effet du phonogénre sur la durée de toutes les syllabes du corpus.

5.3.3. Contour intonatif

Nous indiquons ci-dessous (cf Figure 28) le pourcentage d'occurrence des différents contours intonatifs portés par les constituants focalisés (cf § 4.3.3). On peut faire les observations suivantes :

- les contours commencent très majoritairement (90,3%) par un ton haut
- les contours sont majoritairement (72,0%) de forme montante
- les cinq contours les plus fréquents sont, dans l'ordre :
 - H*H- (28,2%), un contour montant de fin de groupe intermédiaire (cf Figure 16)
 - H* (17,8%), un contour montant de fin de groupe accentuel (cf Figure 20)
 - Hi (14,2%), un ton haut initial sur un constituant de plusieurs syllabes (cf Figure 20)
 - H*L% (13,4%), un contour montant-descendant de fin de groupe intonatif (cf Figure 19)
 - H*H% (10,6%), un contour montant de fin de groupe intonatif (cf Figure 16)
- parmi les contours commençant par un ton L, le contour L*L% (contour descendant de fin de groupe intonatif, cf Figure 17) est plus fréquent que les autres (6,1%)
- les trois types de frontière prosodique (groupe accentuel, groupe intonatif ou groupe intermédiaire) sont répartis assez équitablement, et le groupe intermédiaire est le plus fréquent³⁰ :
 - groupe intermédiaire : 35,7%
 - groupe accentuel : 33,5%
 - groupe intonatif : 30,7%

³⁰ Notons que, étant donné qu'il existe une hiérarchie entre les niveaux de la structure prosodique (cf chap. 1, § 3), une frontière de groupe intonatif est également une frontière de groupe intermédiaire ou accentuel. De même, une frontière de groupe intermédiaire est également une frontière de groupe accentuel. Dans notre analyse, le type de frontière prosodique désigne la frontière la plus haute qui est réalisée sur le constituant focalisé. Par conséquent, une frontière de groupe accentuel désigne une frontière qui n'est pas également une frontière de groupe intermédiaire ou intonatif. De même, une frontière de groupe intermédiaire désigne une frontière qui n'est pas également une frontière de groupe intonatif.

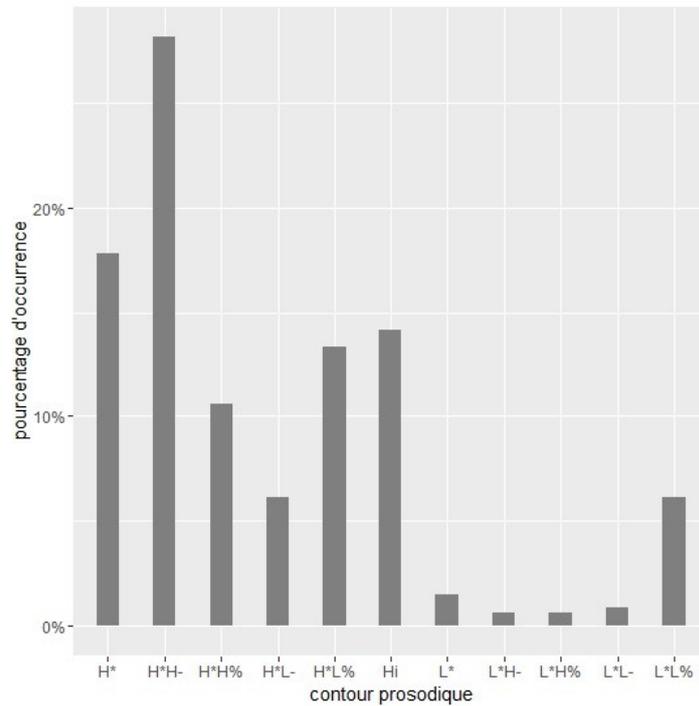


Figure 28. Pourcentage d'occurrence des différents contours intonatifs portés par les constituants focalisés.

Nous indiquons ci-dessous (cf Figure 29), pour chaque contour intonatif, les différences de pourcentage d'occurrence entre les trois phonogenres. Nous avons exclu les contours les moins fréquents (L*, L*H, L*H% et L*L-), pour lesquels la comparaison serait insignifiante. On voit qu'il y a, numériquement, très peu de différences entre les phonogenres. Le contour H*H% est plus fréquent en parole spontanée. Le contour L*L% est plus fréquent en parole interprétée, et il est absent en parole spontanée. Pour déterminer la significativité de l'effet du phonogenre sur le type de contour intonatif, nous avons effectué un test du khi carré de Pearson (cf Pearson 1900). Nous avons employé la fonction générique *chisq.test* du logiciel R. L'effet n'est pas significatif ($\chi^2(20) = 18,9$, $p > 0,5$). Il n'y a pas non plus d'effet significatif du phonogenre sur le type de frontière prosodique ($\chi^2(4) = 2,4$, $p > 0,5$).

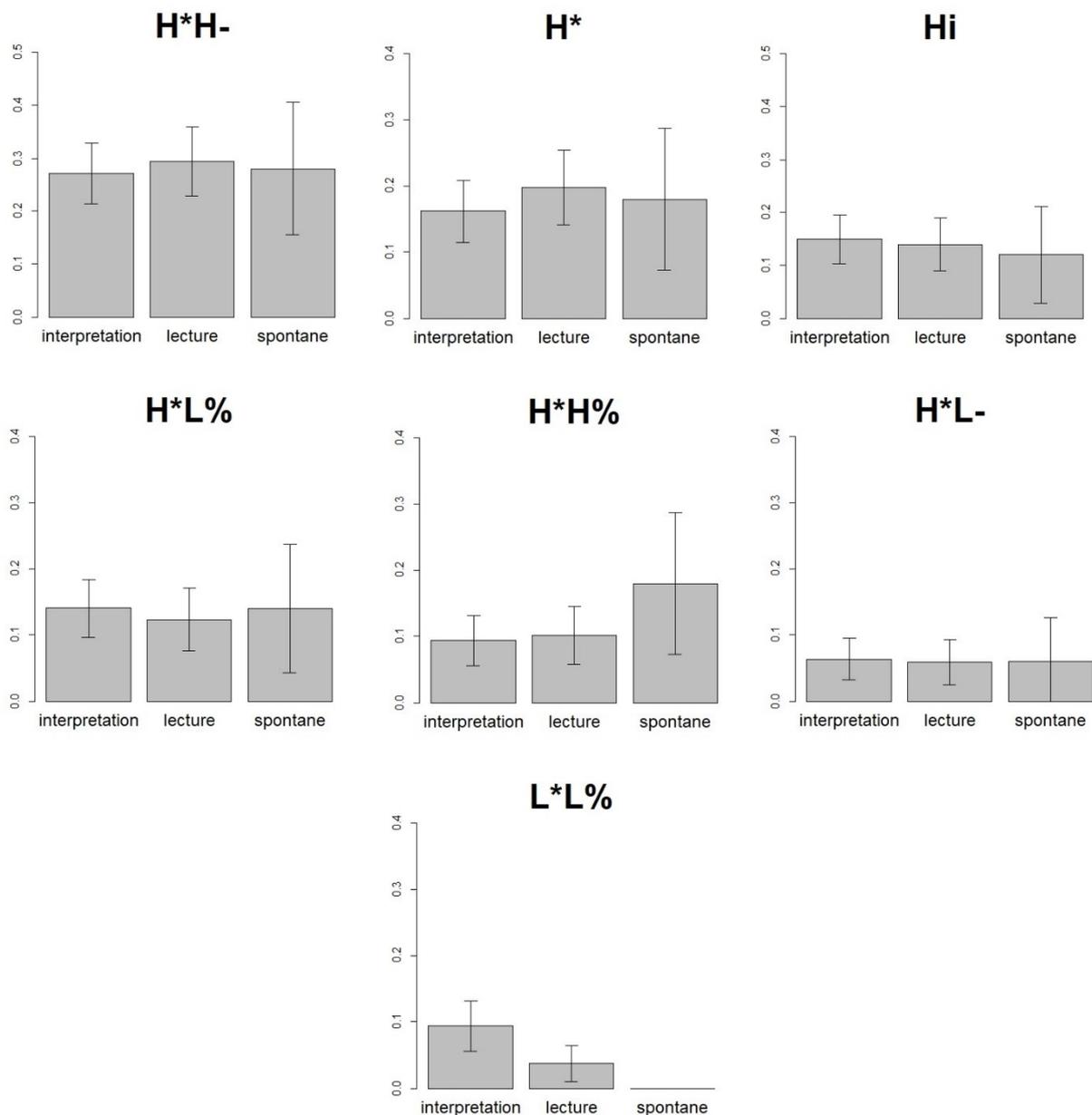


Figure 29. Pourcentages d'occurrence des différents contours intonatifs en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané »). Les barres d'erreur indiquent les intervalles de confiance à 95%.

5.3.4. Etendue du contour

Nous indiquons ci-dessous (cf Figure 30) le pourcentage d'occurrence des différents types d'étendue syllabique du contour intonatif (cf § 4.3.3). Seuls les constituants focalisés comportant plusieurs syllabes sont concernés. On voit que les contours intonatifs sont

majoritairement (72,7%) portés par la dernière syllabe du constituant focalisé. On observe une majorité encore plus grande (81,1%) de contours portés soit par la dernière syllabe, soit par les deux dernières syllabes. Les deux autres types d'étendue les plus fréquents sont le constituant entier (8,7%) et la première syllabe (8,0%). Les contours portant sur la deuxième syllabe ou sur les deux premières syllabes sont très minoritaires (2,3%).

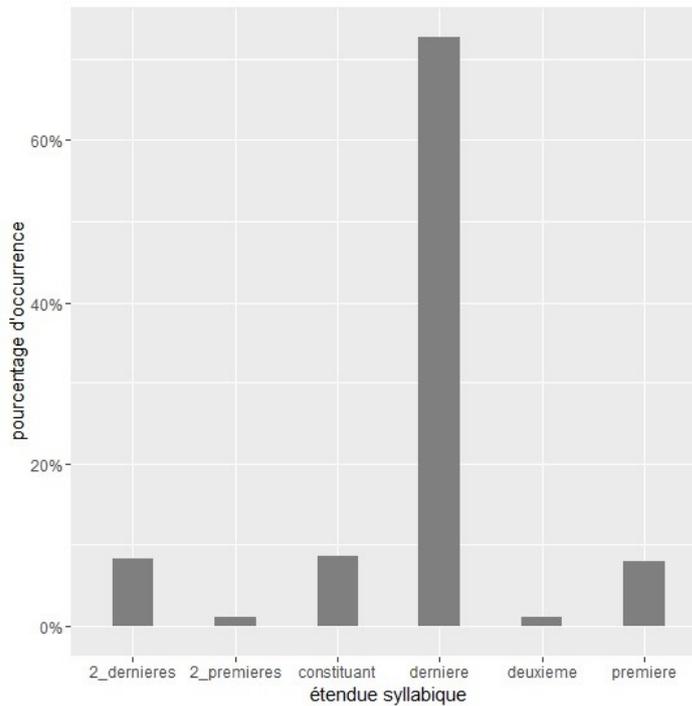


Figure 30. Pourcentage d'occurrence des différents types d'étendue syllabique du contour intonatif sur les constituants focalisés (« 2_dernières » : deux dernières syllabes, « 2_premières » : deux premières syllabes, « constituant » : constituant entier, « dernière » : dernière syllabe, « deuxième » : deuxième syllabe, « première » : première syllabe). Seuls les constituants de plusieurs syllabes sont concernés.

Nous indiquons ci-dessous (cf Figure 31), pour chaque type d'étendue syllabique du contour intonatif, les différences de pourcentage d'occurrence entre les trois phonogènes. Nous avons exclu les deux types d'étendue les moins fréquents (deuxième syllabe et deux premières syllabes). La seule différence numérique notable concerne la distribution des contours portant sur les deux dernières syllabes. Ces contours sont plus fréquents en parole interprétée qu'en parole lue, et ils sont absents de la parole spontanée. Comme pour le contour intonatif (cf § 4.3.3), nous avons employé un test du khi carré de Pearson afin de déterminer la

significativité de l'effet du phonogène sur l'étendue syllabique du contour. L'effet n'est pas significatif ($\chi^2(10) = 9,8, p = 0,46$).

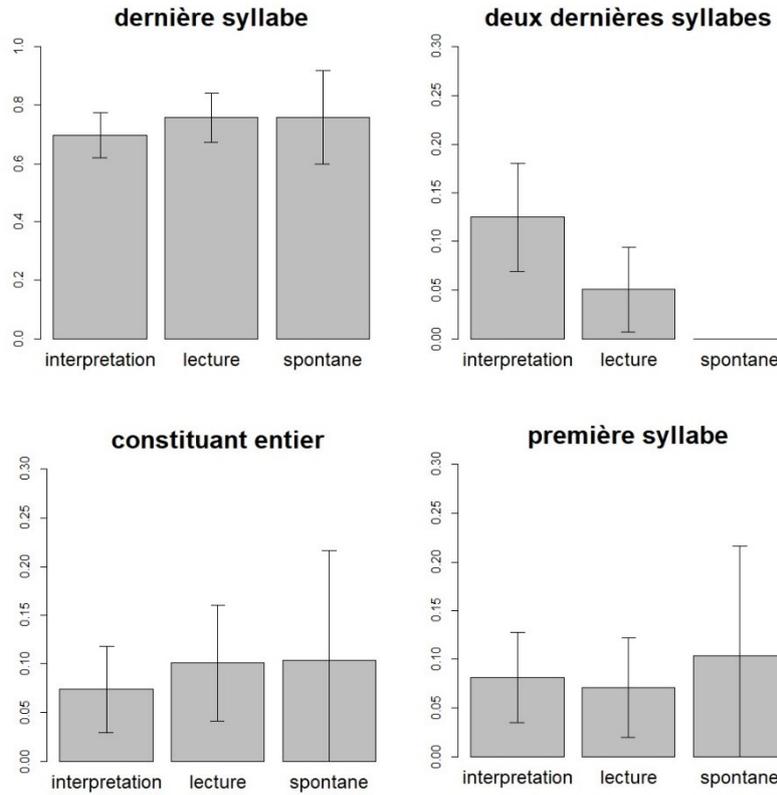


Figure 31. Pourcentage d'occurrence des différentes étendues syllabiques du contour intonatif en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané »). Les barres d'erreur indiquent les intervalles de confiance à 95%.

5.3.5. Accentuation initiale

Nous indiquons ci-dessous (cf Figure 32), pour les constituants de plusieurs syllabes uniquement, le pourcentage d'occurrence de focalisations comportant un accent initial (cf § 4.3.3). On voit que la présence d'accent initial est importante, bien que minoritaire (41,7%).

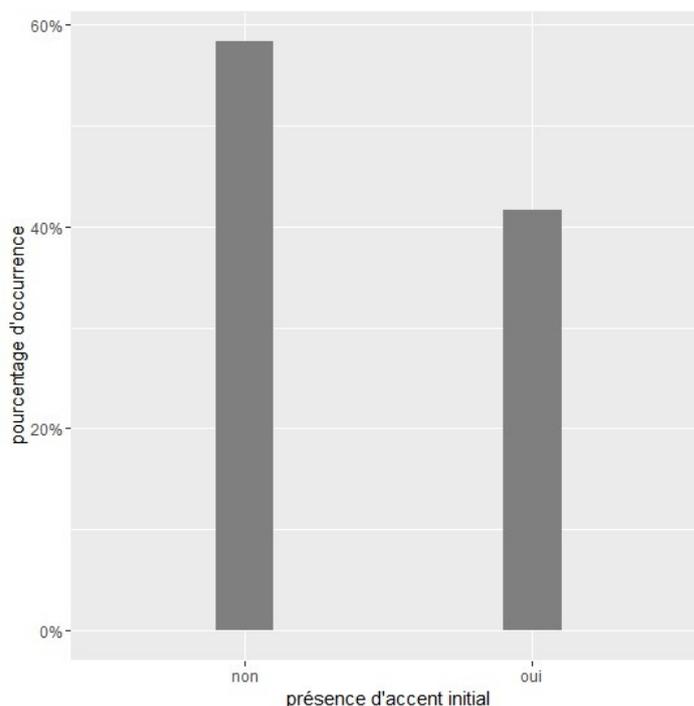


Figure 32. Pourcentage d'occurrence des constituants focalisés portant un accent initial (« oui ») et ne portant pas d'accent initial (« non »). Seuls les constituants de plusieurs syllabes sont concernés.

Nous indiquons ci-dessous (cf Figure 33) les différences de pourcentage d'occurrence d'accentuation initiale entre les trois phonogenres. On n'observe pas de différences numériques notables. Nous avons employé une régression logistique afin de tester la significativité de l'effet du phonogenre sur la présence d'accent initial. La régression logistique a été effectuée sous le logiciel R, à l'aide de la fonction *glmer* de la bibliothèque « lme4 » (Bates et Maechler 2009). La variable dépendante est la présence d'accent initial, et la variable fixe est le phonogenre. Nous avons inclus le locuteur en tant que variable aléatoire (cf note 23). L'effet n'est pas significatif ($\chi^2(2) = 0,7, p > 0,5$).

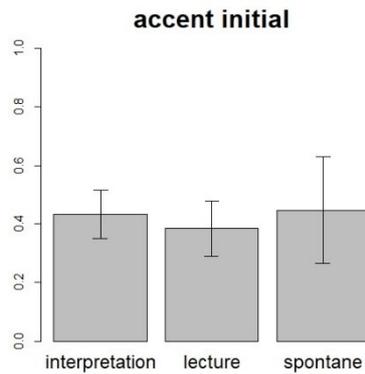


Figure 33. Pourcentage d'occurrence de la présence d'accent initial sur le constituant focalisé en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané »). Les barres d'erreur indiquent les intervalles de confiance à 95%.

5.4. Résultats de l'étude sur corpus réduit

Comme nous l'avons vu plus haut (cf § 5.1), le taux d'accord entre les experts en prosodie concernant le relevé des occurrences de focalisation est relativement bas. Dans le chapitre suivant (cf § 4.3.1), nous verrons que le taux d'accord concernant la classification fonctionnelle des occurrences de focalisation est encore plus bas. Cette faiblesse du taux d'accord pourrait expliquer pourquoi l'analyse n'a pas fourni plus de résultats significatifs. Par ailleurs, elle diminue la portée des résultats qui sont significatifs. Nous avons donc décidé de répéter l'étude en ne conservant que les enregistrements qui ont obtenu les meilleurs taux d'accord. Nous avons conservé les enregistrements pour lesquels le taux d'accord concernant le relevé des occurrences de focalisation est supérieur au taux d'accord pour le corpus entier (kappa de 0,284). Nous indiquons ci-dessous (cf Tableau 15) les enregistrements qui ont été retenus, leur taux d'accord, ainsi que les experts qui les ont annotés. Les enregistrements retenus sont les deux productions spontanées, ainsi que deux des quatre productions interprétées (celles réalisées par un acteur semi-professionnel et une actrice professionnelle, cf § 4.1.2). Par conséquent, le corpus réduit ne contient pas d'enregistrements de parole lue,

ce qui limite la comparaison des résultats avec les résultats de l'étude menée sur le corpus entier.

phonogène	locuteur	taux d'accord (kappa)	groupe d'experts
parole spontanée	locuteur A	0,381	experts 1, 2, 5 et 6
	locutrice B	0,362	
parole interprétée	locuteur A2	0,338	experts 5, 6, 9 et 10
	locutrice B2	0,438	

Tableau 15. Informations sur les enregistrements retenus pour l'étude sur corpus réduit.

Nous présentons ci-dessous les résultats qui diffèrent de manière importante des résultats de l'étude menée sur le corpus entier. Les résultats similaires aux résultats précédents sont disponibles dans l'annexe B.

5.4.1. Taux d'accord entre les experts

Comme on peut le voir ci-dessous (cf Tableau 16), le corpus réduit présente tout d'abord des différences avec le corpus entier concernant le taux d'accord entre les experts en prosodie (cf Tableau 10). Le taux d'accord pour le corpus entier (0,406) augmente de manière importante par rapport à précédemment (0,284). Cela est logique puisque nous avons uniquement conservé, dans le corpus réduit, les enregistrements ayant obtenu le meilleur taux d'accord. On voit aussi que la différence de taux d'accord entre les deux phonogènes n'est pas la même que précédemment. Dans le corpus entier, la parole spontanée présente un taux d'accord supérieur à la parole interprétée dans le corpus entier (0,374 contre 0,285). Dans le corpus réduit, au contraire, la parole interprétée présente un taux d'accord légèrement supérieur à la parole spontanée (0,396 contre 0,374). Tous les kappas sont significatifs d'après leur score-z (cf note 23).

	parole spontanée	parole interprétée	corpus entier
taux d'accord (kappa de Fleiss)	0,374 (z = 36,1)	0,396 (z = 38,8)	0,406 (z = 55,8)

Tableau 16. Taux d'accord entre les experts en prosodie concernant le relevé des occurrences de focalisation dans le corpus réduit. Les scores-z sont indiqués sous les kappas. Tous les scores-z sont significatifs à $p < 0,001$.

5.4.2. Contour intonatif

La seule autre différence notable entre le corpus réduit et le corpus entier concerne les contours intonatifs portés par les occurrences de focalisation prosodique. Comme on peut le voir ci-dessous (cf Figure 34), la distribution est globalement la même que précédemment (cf Figure 28). Cependant, la répartition des trois types de frontière prosodique (groupe accentuel, groupe intonatif et groupe intermédiaire) connaît un changement. Dans le corpus entier (cf § 5.3.3), la frontière la plus fréquente est celle du groupe intermédiaire (35,7%), suivi du groupe accentuel (33,5%) et du groupe intonatif (30,7%). Dans le corpus réduit, la frontière la plus fréquente est celle du groupe intonatif, et la moins fréquente est celle du groupe accentuel :

- groupe intonatif : 37,6%
- groupe intermédiaire : 31,7%
- groupe accentuel : 30,7%

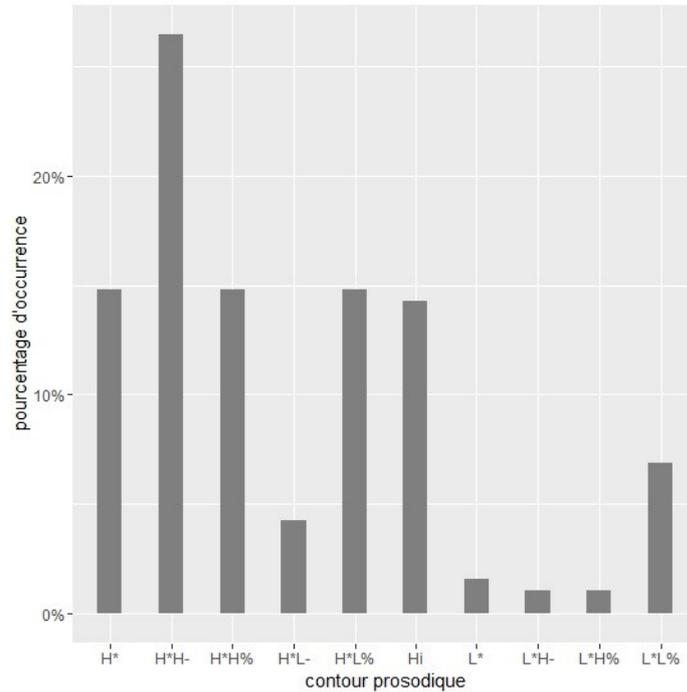


Figure 34. Pourcentage d'occurrence des différents contours intonatifs portés par les constituants focalisés dans le corpus réduit.

6. Discussion

6.1. Relevé des occurrences de focalisation

Avant de reprendre les questions de recherche posées au début de ce chapitre (cf § 1.1), on peut tout d'abord s'interroger sur le taux d'accord qui a été mesuré entre les experts en prosodie concernant le relevé des occurrences de focalisation prosodique (cf § 5.1). Comme nous l'avons vu, le taux d'accord concernant le corpus entier (κ de 0,284) est relativement faible par rapport aux taux d'accord obtenus par deux études récentes (Cole et al 2010, Avanzi et al 2011). Le fait qu'il n'y ait pas plus de concordance entre les annotations des experts est problématique, car cela réduit la portée des résultats de l'analyse de la fréquence d'occurrence et de la réalisation prosodique de la focalisation. Cependant, on peut aussi considérer qu'il s'agit d'un résultat intéressant en lui-même d'un point de vue méthodologique. Au moins deux raisons peuvent être avancées pour expliquer ce résultat :

- Il est possible qu'il soit difficile pour les experts de détecter les occurrences de focalisation de manière purement auditive, sans l'aide d'un logiciel d'analyse acoustique. Les critères acoustiques que nous avons donnés aux experts (cf § 4.2.2) ne sont en effet pas nécessairement faciles à identifier, même pour des experts possédant une expérience de l'étude de la prosodie.
- Il est également possible que la faiblesse du taux d'accord soit due à une disparité dans la compétence des experts. Comme nous l'avons vu (cf § 4.2.1), deux experts sont spécialisés en pragmatique et non en phonétique ou en phonologie. De plus, trois experts (autres que les deux précédents) sont des étudiants de niveau Master uniquement. L'expérience pourrait peut-être être améliorée en obtenant l'annotation du corpus en deux fois. Après un premier essai par les experts, l'expérimentateur vérifierait, auprès des experts dont l'annotation est très différente des autres (ce qui affaiblit le taux d'accord), qu'ils ont bien compris les consignes ainsi que les critères à utiliser. Une autre possibilité serait de retenir uniquement, après ce premier essai, les experts ayant obtenu le meilleur taux d'accord entre eux. Cependant, nous avons vu que les résultats de l'étude que nous avons menée sur un corpus réduit (constitué des enregistrements ayant obtenu le meilleur taux d'accord, cf § 5.4) sont globalement similaires aux résultats obtenus sur le corpus entier. Cela suggère que le fait de ne retenir que les meilleurs experts ne modifie pas les résultats de l'analyse de manière importante.

Les différences de taux d'accord selon le phonogène constituent également un résultat intéressant. Comme nous l'avons vu, c'est pour la parole spontanée que le taux d'accord est le plus élevé (kappa de 0,374), suivie de la parole interprétée (kappa de 0,285) et de la parole lue (kappa de 0,230). Ce résultat peut sembler contradictoire. En effet, on aurait pu s'attendre à ce que la parole spontanée soit plus difficile à annoter, en raison de ses nombreuses disfluences. Inversement, on aurait pu s'attendre à ce que le caractère normatif de la parole lue en fasse le phonogène le plus facile à annoter, car le plus régulier du point de vue prosodique. Les raisons suivantes peuvent être avancées pour expliquer les différences de taux d'accord entre les phonogènes :

- Comme nous l'avons mentionné plus haut (cf § 4.1.2), les locuteurs de parole lue et de parole interprétée possèdent une expérience de l'art dramatique. Il est possible que leurs

productions contiennent des traits prosodiques qui relèvent d'un phonostyle professionnel (celui du comédien). Cela a peut-être eu pour conséquence que ces productions diffèrent de manière importante de la parole spontanée, ou de la parole de laboratoire habituelle (c'est-à-dire de la parole lue à haute voix, mais par des locuteurs lambda). Cela expliquerait donc que les experts aient eu plus de mal à annoter ces productions. Il est également possible que la présence d'un phonostyle du comédien ait introduit plus de variation prosodique dans les productions lues et interprétées, par rapport aux productions spontanées. Cela a pu entraîner une plus grande difficulté chez les experts pour annoter ces deux phonogenres³¹.

- En plus de la variation phonostylistique, une différence de naturalité existe peut-être entre les trois phonogenres. La lecture à haute voix implique une double tâche cognitive (lire le texte et produire de la parole), ce qui a pu influencer la prosodie des locuteurs de parole lue. De plus, il n'a pas seulement été demandé à ces locuteurs de lire de façon neutre, mais aussi d'imaginer qu'ils étaient le personnage et participaient réellement à la conversation (cf § 4.1.2). La parole interprétée, quant à elle, contient souvent un aspect artistique, ce qui a pu rendre la prosodie des locuteurs moins naturelle. Cela est d'autant plus possible que, comme nous l'avons expliqué, les locuteurs de parole interprétée possèdent tous une expérience de l'art dramatique.
- Le résultat est peut-être simplement dû à la répartition des enregistrements du corpus parmi les experts en prosodie (cf § 4.2.2). En effet, les productions spontanées ont toutes été annotées par les mêmes experts, ce qui n'est pas le cas pour les productions lues et interprétées. Ceci est dû au fait que le corpus contient moins de productions spontanées que de productions lues et interprétées (cf Tableau 8). Le fait que les experts varient pour la parole lue et la parole interprétée pourrait expliquer que ces deux phonogenres présentent des taux d'accord inférieurs à celui de la parole spontanée. Cependant, dans le corpus réduit, nous avons vu (cf § 5.4.1) que la parole interprétée présente au contraire

³¹ Comme expliqué plus haut (cf § 4.1.2), le choix d'enregistrer des locuteurs possédant une expérience de l'art dramatique s'explique par la nécessité que les locuteurs soient suffisamment capables de mémoriser le texte (qui est relativement long).

un taux d'accord supérieur à celui de la parole spontanée. Or, les productions interprétées du corpus réduit sont toutes annotées par les mêmes experts (cf Tableau 15). Cela suggère que les variations de taux d'accord entre phonogenres dans le corpus entier ne sont pas dues à la répartition des enregistrements parmi les experts.

- La supériorité du taux d'accord en parole interprétée par rapport à la parole lue est peut-être due au protocole expérimental. Comme expliqué plus haut (cf § 4.1.2), pour la tâche d'interprétation, les locuteurs ont mémorisé le texte et ont « préparé » leur interprétation afin qu'elle paraisse spontanée et expressive. Par opposition, lors de la tâche de lecture à haute voix, les locuteurs ont découvert le texte pour la première fois, et ont seulement pu le lire de manière silencieuse avant l'enregistrement. Le fait que les locuteurs de parole interprétée aient préparé leur production a peut-être eu une influence au niveau prosodique. Il est par exemple possible que les focalisations prosodiques soient plus marquées dans ce phonogène, ou bien plus similaires les unes aux autres. Cela expliquerait que l'annotation ait été plus facile pour les experts dans ce phonogène.

6.2. Fréquence d'occurrence de la focalisation

La première question de recherche concerne la fréquence d'occurrence de la focalisation prosodique en français, indépendamment du phonogène. A notre connaissance, aucune étude jusqu'ici n'a mesuré cette fréquence. Dans notre étude, le pourcentage de syllabes focalisées du corpus est de 11,3%. Il est intéressant de comparer ce pourcentage à la fréquence de syllabes proéminentes détectées par le logiciel Analor dans le corpus. Comme nous l'avons expliqué (cf § 4.3.3), nous avons utilisé la fonction de détection automatique de proéminences proposée par ce logiciel afin de pouvoir déterminer la présence d'accent initial sur les constituants focalisés. D'après Analor, 35,8% des syllabes du corpus sont proéminentes. Pour rappel, la notion de proéminence d'Analor est plus large que la notion de focalisation que nous avons adoptée (cf § 2.1). En effet, elle comprend les accents rythmiques (ayant pour fonction de marquer la frontière initiale ou finale d'un groupe prosodique). Dans notre étude, nous avons demandé aux experts de ne pas prendre en compte les accents rythmiques dans leur annotation (cf § 4.2.2). Nos résultats montrent donc que, parmi toutes

les proéminences du français, un peu moins d'un tiers sont des accents de focalisation remplissant une fonction sémantico-pragmatique ou emphatique³².

6.3. Réalisation prosodique de la focalisation

La deuxième question de recherche concerne la réalisation prosodique de la focalisation, toujours indépendamment du phonogène. Au niveau phonétique, nous avons effectué des mesures de hauteur et de durée sur toutes les syllabes du corpus (cf § 4.3.2). Les résultats montrent que les syllabes focalisées ont une hauteur et une durée significativement plus élevées que celles des syllabes non-focalisées (cf Figure 23 et Figure 25). Notre étude confirme donc plusieurs études précédentes (e.g. Di Cristo 1999b, Rossi 1999 : 128, Astésano 2001 : 53-57) qui attribuent une augmentation de hauteur et de durée à la focalisation.

Nous avons également effectué une classification phonologique des occurrences de focalisation (cf § 4.3.3). La classification comporte trois catégories : le type de contour intonatif, l'étendue syllabique du contour intonatif, et la présence d'accent initial. A notre connaissance, c'est la première fois qu'une telle analyse de la focalisation est réalisée. En effet, les études précédentes s'intéressent plutôt au marquage phonologique des fonctions attribuées à la focalisation, et non au marquage de la focalisation en tant que phénomène global (comme c'est le cas dans ce chapitre). Nous discutons successivement ci-dessous les résultats observés :

- Les contours commencent très majoritairement (90,3%) par un ton haut, et ils sont majoritairement (72,0%) de forme montante (cf Figure 28). Ces deux résultats sont logiques, étant donné que la littérature attribue généralement une augmentation de hauteur à la focalisation, et que nous avons observé une telle augmentation dans notre corpus.

³² D'après Simon et al (2010), le pourcentage de syllabes proéminentes en français est de 29,9%. Cela signifierait qu'un peu plus d'un tiers des proéminences du français sont des accents de focalisation. Ce pourcentage a été obtenu sur le corpus C-PROM (Avanzi et al 2010), qui contient lui aussi plusieurs phonogènes. La détection a été effectuée de manière automatisée à l'aide de l'outil ProsoProm contenu dans Prosogram (Mertens 2004 ; cf § 4.3.2). Le processus de détection de ProsoProm diffère de celui d'Analor (cf Lacheret et al 2008).

- Le contour le plus fréquent (28,2%) est H*H-, un contour montant de fin de groupe intermédiaire (cf Figure 16). Ce contour est moins ample que le contour H*H% (qui marque la fin d'un groupe intonatif) et s'accompagne d'un allongement moins important. Dans notre corpus, le contour H*H- se trouve généralement en fin d'énoncé et sert à exprimer une continuation. Nous présentons ci-dessous (cf Figure 35) deux exemples de ce type de contexte.

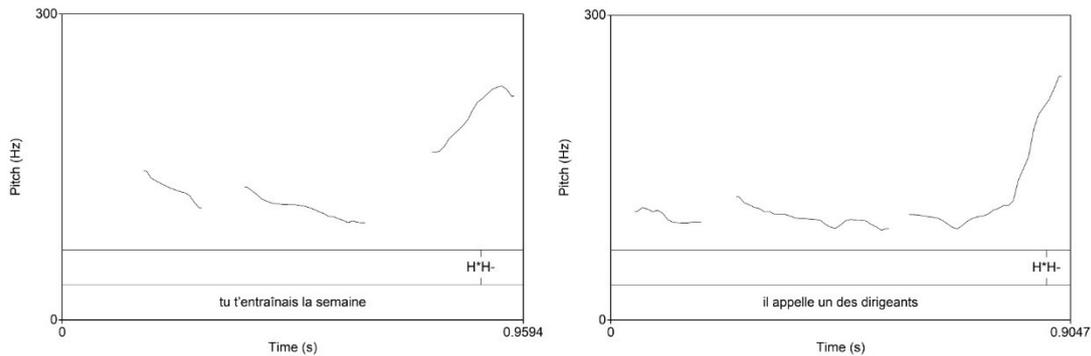


Figure 35. Courbes de fréquence fondamentale de deux énoncés issus du corpus de l'expérience se terminant par un contour intonatif H*H-.

- Les deux contours suivants dans l'ordre de fréquence sont H* (17,8%) et Hi (14,2%). Ces deux contours se situent au niveau du groupe accentuel dans la structure prosodique. H* marque la frontière finale du groupe, et Hi marque la frontière initiale (cf Figure 20). H* correspond à une augmentation de hauteur, ce qui est là aussi cohérent avec la littérature, ainsi qu'avec nos observations concernant ce paramètre. Hi correspond à une accentuation initiale sur un constituant de plusieurs syllabes. Cela est également cohérent avec la littérature, qui décrit l'accent initial comme un trait fréquent de la focalisation (e.g. Di Cristo 1999b, Rossi 1999 : 128, Astésano 2001 : 202). De plus, comme nous allons le voir, une proportion importante (bien que minoritaire) de constituants polysyllabiques comporte un accent initial dans notre étude.
- Les trois contours suivants dans l'ordre de fréquence sont H*L% (13,4%), H*H% (10,6%) et L*L% (6,1%). Ces contours marquent la fin d'un groupe intonatif et sont respectivement de forme montante-descendante, montante et descendante (cf Figure 19, Figure 16 et

Figure 17). Dans le premier chapitre (cf § 4), nous avons vu que ces contours sont soit des contours terminaux qui indiquent la modalité de l'énoncé, soit des contours continuatifs (en particulier H*H%). Ils peuvent également remplir des fonctions expressives, notamment indiquer l'attitude du locuteur (ex : évidence, doute, surprise). On retrouve effectivement ces différents contextes dans notre corpus, comme on peut le voir ci-dessous (cf Figure 36). La présence d'un contour terminal est cohérente avec la littérature, qui attribue souvent ce trait à la focalisation (e.g. Rossi 1999 : 84-87, Post 2002, Martin 2009 : 124-130). Le fait que les contours remplissent une fonction expressive est également attendu, étant donné que la littérature attribue ce type de fonction à la focalisation, comme nous le verrons dans le chapitre suivant (cf § 3.1.2).

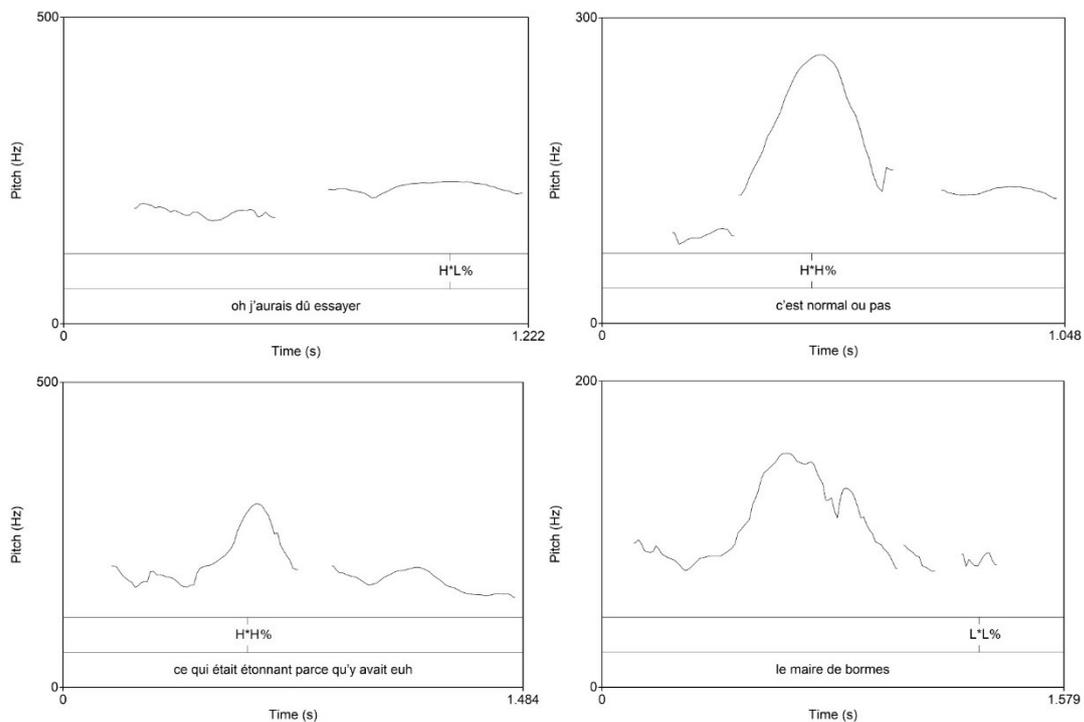


Figure 36. Courbes de fréquence fondamentale de quatre contours H*L%, H*H% et L*L% issus du corpus de l'expérience. En haut à gauche, le contour H*L% indique la modalité assertive (ou exclamative) de l'énoncé, et exprime probablement une attitude d'ironie de la part de la locutrice. En haut à droite, le contour H*H% (sur normal) indique la modalité interrogative de l'énoncé (il est suivi d'une séquence désaccentuée), et exprime probablement une attitude de surprise de la part du locuteur. En bas à gauche, le contour H*H% (sur étonnant) exprime une continuation. En bas à droite, le contour L*L% indique la modalité assertive de l'énoncé.

- Le type de frontière prosodique le plus fréquent sur les contours intonatifs est le groupe intermédiaire (35,7%), suivi du groupe accentuel (33,5%) et du groupe intonatif (30,7%).

Dans le corpus réduit, cette répartition est différente (cf § 5.4.2) : la frontière la plus fréquente est celle du groupe intonatif, suivi du groupe intermédiaire et du groupe accentuel. Dans les deux cas, on voit que la focalisation n'a pas lieu à un niveau particulier de la structure prosodique. Cela suggère que la focalisation peut affecter différents éléments : un mot, un groupe de mots, ou l'énoncé tout entier. Ce résultat est cohérent avec les différentes fonctions que la littérature attribue à la focalisation. Comme nous le verrons dans le chapitre 3 (cf §2.1 et § 3.1), la fonction de marquage de focus informationnel concerne l'énoncé entier, tandis que les autres fonctions peuvent s'appliquer à un constituant plus petit. On s'attendrait donc à ce que le marquage de focus informationnel soit réalisé au niveau du groupe intonatif (où se trouvent les contours terminaux indiquant la modalité de l'énoncé), et les autres fonctions au niveau du groupe accentuel ou intermédiaire. Les résultats de l'étude que nous présentons dans le chapitre 3 (cf § 4.3) ne permettent cependant pas de répondre entièrement à cette question. Nous verrons que la fonction globale de marquage de focus (informationnel, contrastif, associatif ou *verum*) est majoritairement réalisée au niveau du groupe intermédiaire, et que (de manière attendue) la fonction d'insistance est majoritairement réalisée au niveau du groupe accentuel.

- Les contours intonatifs sont majoritairement (72,7%) portés par la dernière syllabe du constituant focalisé (cf Figure 30). Contrairement à la catégorie du contour intonatif, pour laquelle une variété de formes est observée, on voit ici que la focalisation est associée à un type d'étendue syllabique en particulier. Nous avons analysé la distribution des contours intonatifs selon l'étendue syllabique. Comme on peut le voir ci-dessous (cf Tableau 17), les contours portés par la dernière syllabe sont en grande majorité (83,3%) montants ou descendants. Par contraste, les contours portés par les deux dernières syllabes ou par le constituant entier sont majoritairement (64,4%) des contours montants-descendants ou descendants-montants. Cette distribution est logique, car ces derniers contours présentent un mouvement tonal plus complexe que les précédents, ce qui justifie qu'ils s'étendent sur un plus grand nombre de syllabes. Enfin, le ton haut initial H_i est logiquement le seul contour qui est porté par les syllabes initiales du constituant focalisé.

	dernière syllabe	deux dernières syllabes	constituant entier	première syllabe, deuxième syllabe ou deux premières syllabes
contours montants ou descendants (H*, L*, H*H-, H*H%, L*L-, L*L%)	159	8	8	0
contours montants-descendants ou descendants-montants (H*L-, H*L%, L*H-, L*H%)	32	14	15	0
ton haut initial (Hi)	0	0	0	27

Tableau 17. Distribution des principales formes de contour intonatif selon l'étendue syllabique du contour.

- Parmi les constituants comportant plusieurs syllabes, une proportion importante bien que minoritaire (41,7%) comporte un accent initial. Comme nous l'avons mentionné plus haut, ce résultat est cohérent avec la littérature, qui attribue souvent ce trait à la focalisation (e.g. Di Cristo 1999b, Rossi 1999 : 128, Astésano 2001 : 202). Notre étude montre qu'il ne s'agit toutefois pas d'un trait obligatoire. Comme pour l'étendue syllabique, nous avons cherché à savoir si certains contours intonatifs sont corrélés à la présence d'un accent initial (hormis le contour Hi, qui l'est forcément). Les résultats montrent uniquement qu'une minorité de contours H*H- et H*H% (10,8% et 24,1% respectivement) sont accompagnés d'un accent initial. Comme nous l'avons vu plus haut, ces contours ont généralement pour fonction d'exprimer une continuation ou d'indiquer une modalité interrogative. Il est donc possible que le trait d'accentuation initiale ne soit pas relié à ces deux fonctions.

6.4. Influence du phonogène sur la fréquence d'occurrence

Concernant cette question de recherche, nous nous attendions (cf § 1.2) à observer une plus forte fréquence d'occurrence de la focalisation en parole interprétée par rapport aux deux autres phonogènes. Cette prédiction se base sur deux études. La première est l'étude de

Goldman et al (2011), qui observe une relation entre le degré de « préparation » du phonogène et le pourcentage d'accents initiaux. La seconde étude est celle de Godement-Berline (2014), qui relève un pourcentage de syllabes proéminentes relativement élevé en parole interprétée par rapport aux phonogènes du corpus C-PROM. Les résultats de l'analyse vont dans le sens de cette prédiction :

- Le pourcentage de syllabes focalisées est significativement plus élevé en parole interprétée (14,2%) qu'en parole lue (10,7%) et qu'en parole spontanée (6,6%). Ce résultat est cohérent avec l'intuition courante selon laquelle la parole interprétée se caractérise par une forte expressivité. D'un point de vue méthodologique, cela signifie également que la parole interprétée est un phonogène favorable à l'étude de la focalisation prosodique.
- La différence de pourcentage entre la parole lue et la parole spontanée approche la significativité. Ce résultat, ainsi que la différence significative entre parole interprétée et parole lue, est cohérent avec l'étude de Goldman et al (2011) : plus le degré de préparation est élevé, plus le nombre de focalisations augmente. La parole lue implique en effet un degré de préparation plus important que la parole spontanée (puisque le locuteur oralise un texte écrit) et moins important que la parole interprétée (où le locuteur mémorise le texte).

Notons toutefois que deux facteurs ont pu influencer les résultats :

- Comme nous l'avons mentionné plus haut (cf § 4.1.2), les locuteurs de parole lue et interprétée possèdent tous une expérience de l'art dramatique. Cela a pu introduire dans leurs productions des traits prosodiques qui relèvent d'un phonostyle du comédien. Ce phonostyle pourrait être à l'origine des différences de pourcentage entre la parole spontanée et les deux autres phonogènes.
- La différence de fréquence entre la parole interprétée et la parole lue pourrait être due à la différence de débit entre ces deux phonogènes. En effet, nous avons vu (cf Figure 27) que les syllabes ont globalement une durée moyenne significativement plus longue en parole interprétée qu'en parole lue. Il est possible qu'il soit plus facile pour les experts de repérer les occurrences de focalisations lorsque le débit est plus lent. Cela expliquerait qu'ils aient repéré plus d'occurrences en parole interprétée.

6.5. Influence du phonogène sur la réalisation prosodique

Concernant cette question de recherche, nous nous attendions (cf § 1.2) à ce que la focalisation soit plus « marquée » en parole lue et en parole interprétée, selon les différents paramètres que nous avons étudiés (ex : hauteur, durée, accentuation initiale). Cette prédiction se base sur les études précédentes sur les différences phonétiques globales entre ces phonogènes (cf § 3.2). Les résultats de l'analyse ne vont pas dans le sens de cette prédiction :

- La hauteur moyenne des syllabes focalisées ne varie pas de manière significative entre les trois phonogènes (cf Figure 24). Ce résultat est contradictoire avec les études qui observent, en parole interprétée, un registre tonal plus élevé et des variations de hauteur plus importantes qu'en parole spontanée (Audibert et al 2010, Jürgens et al 2011). Notons qu'il est possible que le contexte expérimental ait eu un effet inhibiteur sur les locuteurs de parole interprétée. Cela expliquerait qu'on ne retrouve pas certains des traits associés à ce phonogène. En revanche, l'absence de différence de hauteur entre la parole lue et la parole spontanée n'est pas forcément surprenante. En effet, nous avons vu plus haut (cf § 3.2) que la littérature est souvent contradictoire concernant les différences entre ces deux phonogènes (e.g. Lieberman et al 1985, Laan 1997, Simon et al 2010, Wagner et Windmann 2016).
- La durée moyenne des syllabes focalisées est significativement plus élevée en parole lue qu'en parole interprétée (cf Figure 26). De plus, ce résultat ne peut être attribué à une différence globale de durée syllabique entre ces deux phonogènes. En effet, nous avons vu (cf Figure 27) qu'on observe au contraire que le débit est plus lent en parole interprétée qu'en parole lue. Ce résultat est cohérent avec l'étude d'Astésano (2001 : 192) qui observe que les accents finaux sont plus allongés en parole lue qu'en parole spontanée. Bien que notre analyse ne porte pas sur tous les accents finaux du corpus (mais seulement sur les accents de focalisation), il semble que ce trait soit effectivement une des caractéristiques du phonogène de la lecture à haute voix.
- Au niveau phonologique, nous n'avons pas observé de différences significatives entre les phonogènes concernant le type de contour intonatif des constituants focalisés (cf Figure

29), l'étendue syllabique des contours (cf Figure 31) et la présence d'accent initial (cf Figure 33). Bien que la focalisation connaisse globalement une grande variété de contours (cf Figure 28), aucun de ces contours n'est associé à un phonogène plus qu'à un autre. L'absence de différences concernant l'étendue du contour est moins surprenante. En effet, nous avons vu que, dans le corpus entier, les contours des constituants focalisés sont très majoritairement portés par la dernière syllabe (cf Figure 30). L'absence de différence concernant l'accentuation initiale ne va pas dans le sens de notre prédiction selon laquelle la focalisation serait plus « marquée » prosodiquement en parole lue et interprétée. De plus, elle est contradictoire avec l'étude de Goldman et al (2011) qui observe au contraire une relation entre la fréquence d'accents initiaux et le degré de préparation du phonogène.

7. Conclusions

L'étude décrite dans ce chapitre présente deux contributions principales. La première est une description de la focalisation prosodique en français :

- La fréquence d'occurrence de la focalisation est de 11,3% dans notre corpus. Ce résultat inédit est à comparer à la fréquence d'occurrence de la proéminence en français, qui se situe entre 30 et 35%. Cela signifie qu'environ un tiers des proéminences du français remplissent une fonction sémantico-pragmatique (marquage de focus) ou emphatique (insistance ou expressivité). L'étude de ces différentes fonctions de la focalisation est abordée dans le chapitre suivant.
- Le rôle de l'augmentation de hauteur et de durée syllabique dans la focalisation est confirmé. Comme dans plusieurs études précédentes, on observe dans notre corpus une forte augmentation sur les constituants focalisés concernant ces deux paramètres.
- La focalisation peut être réalisée par une variété de contours intonatifs. La majorité de ces contours est toutefois de forme montante et commence par un ton haut. La focalisation peut avoir lieu à tout niveau de la structure prosodique (groupe accentuel, intermédiaire ou intonatif). Les contours intonatifs sont majoritairement portés par la dernière syllabe

du constituant focalisé. Toutefois, les contours montants-descendants ou descendants-montants ont tendance à présenter une étendue syllabique plus large.

- Le rôle de l'accent initial dans la focalisation est en partie confirmé. Une proportion importante, bien que minoritaire, de constituants focalisés comporte une telle accentuation.

La deuxième contribution de cette étude est une analyse de l'influence du phonogène sur la focalisation prosodique :

- La fréquence d'occurrence de la focalisation varie selon le phonogène. Conformément à notre prédiction, elle est significativement plus élevée en parole interprétée qu'en parole lue et qu'en parole spontanée. De façon marginale, elle est également plus élevée en parole lue qu'en parole spontanée. La fréquence d'occurrence semble donc augmenter selon le degré de préparation du phonogène, comme dans l'étude de Goldman et al (2011). Cependant, la supériorité de la fréquence d'occurrence en parole interprétée pourrait également être due à l'aspect expressif de ce phonogène (et non au processus de mémorisation du texte). Afin de distinguer le facteur responsable de l'augmentation de fréquence, il serait possible de comparer les productions interprétées avec les restitutions neutres du texte que nous avons également enregistrées (cf § 4.1.2). Pour rappel, dans ces restitutions, le locuteur a mémorisé le texte mais le prononce sans simulation de spontanéité ni d'expressivité.
- La réalisation prosodique de la focalisation n'est globalement pas influencée par le phonogène. La seule exception concerne la durée syllabique des constituants focalisés, qui est significativement plus élevée en parole lue qu'en parole interprétée. De plus, cette différence n'est pas due à une différence de débit entre ces deux phonogènes. Il est donc possible que la réalisation de la focalisation au moyen d'une forte augmentation de durée soit un trait typique du phonogène de la parole lue.

Cette étude peut par ailleurs être renforcée en améliorant le protocole expérimental et en prenant en compte d'autres traits prosodiques :

- Le corpus comprend seulement deux enregistrements de parole spontanée (contre quatre enregistrements de parole lue et quatre enregistrements de parole interprétée). Par ailleurs, les locuteurs de la tâche de reproduction possèdent tous une expérience de l'art dramatique. Cela a pu introduire dans leur production des traits relevant d'un phonostyle professionnel. Il serait peut-être préférable de n'enregistrer que des locuteurs lambda, voire de demander aux locuteurs de parole spontanée de reproduire eux-mêmes leurs productions, comme c'est le cas dans certaines études (e.g. Laan 1997, Mixdorff et Pfitzinger 2005, Wagner et Windmann 2016).
- Il existe une disparité dans la compétence des experts en prosodie qui ont relevé les occurrences de focalisation. Certains d'entre eux sont de niveau Master, ou bien ne sont pas spécialisés en phonétique-phonologie. L'annotation pourrait par ailleurs être améliorée en étant effectuée en deux étapes. Après un premier essai, seuls les experts ayant obtenu le meilleur taux d'accord entre eux seraient sélectionnés. Il serait également possible, après ce premier essai, de vérifier la compréhension des consignes par les experts. De plus, l'interdiction faite aux experts d'utiliser un logiciel d'analyse acoustique pour effectuer l'annotation a pu jouer un rôle dans la faiblesse du taux d'accord. Enfin, l'annotation pourrait être améliorée en faisant annoter tous les enregistrements par les mêmes experts (cela n'a pas été possible dans cette étude pour des raisons pratiques).
- Au niveau phonétique, des mesures d'intensité pourraient être effectuées (et éventuellement de qualité vocalique). Il serait également possible d'analyser, comme dans l'étude d'Astésano (2001), la configuration tonale des focalisations ainsi que leur constituance syllabique. Comme nous l'avons vu dans le chapitre 1 (cf § 2.4), la configuration tonale d'un accent désigne le caractère symétrique ou asymétrique de son pic de hauteur. L'analyse de la constituance syllabique consiste quant à elle à mesurer les durées respectives des différents éléments de la syllabe (attaque, noyau, coda et rime). Enfin, au niveau phonologique, notre étude n'a pas pris en compte un des traits souvent attribués à la focalisation, celui de la désaccentuation post-focale ou pré-focale (cf § 2.2).

8. Résumé

L'analyse présentée dans ce chapitre concerne l'influence du phonogène sur le phénomène de la focalisation prosodique. Deux aspects de la focalisation sont analysés : sa fréquence d'occurrence et sa réalisation prosodique. Le corpus d'étude comprend trois phonogènes : la parole interprétée (l'oralisation d'un texte écrit préalablement mémorisé par le locuteur), la parole lue et la parole spontanée. Les questions de recherche suivantes sont posées : 1) Quelle est la fréquence d'occurrence de la focalisation en français ? 2) Quelle est la réalisation prosodique de la focalisation en français ? 3) La focalisation présente-t-elle une fréquence d'occurrence supérieure en parole interprétée ? 4) La focalisation présente-t-elle des réalisations prosodiques spécifiques en parole interprétée ?

Une définition de la focalisation est adoptée : il s'agit du soulignement d'un constituant dans un énoncé, pour toute raison sémantico-pragmatique ou emphatique. Un rappel des principaux traits prosodiques de la focalisation est effectué : augmentation de hauteur, de durée ou d'intensité, accentuation initiale, présence d'un contour terminal, et désaccentuation post-focale ou pré-focale.

La parole spontanée, la parole lue et la parole interprétée sont ensuite situées au sein du domaine de la *phonostylistique*, ou l'étude des caractéristiques phonétiques des styles de parole. Dans ce domaine, une distinction est souvent faite entre deux catégories de styles de parole : les *phonostyles* et les *phonogènes*. Les phonostyles sont reliés à des catégories sociales, professionnelles ou culturelles, ou encore à des individus particuliers. Les phonogènes sont reliés à des conditions de production spécifiques. La parole spontanée, la parole lue et la parole interprétée constituent des phonogènes qui diffèrent notamment selon le degré de préparation du discours. Nous présentons ensuite l'état de l'art concernant les différences prosodiques entre la parole spontanée et les deux autres phonogènes.

La méthodologie de l'analyse est ensuite décrite. Pour constituer le corpus, nous avons employé un protocole semblable au protocole RepTask (Laurens, Marandin, Patin et Yoo 2011). Deux dialogues spontanés ont tout d'abord été enregistrés. Nous avons retenu des extraits dans lesquels un des locuteurs raconte une anecdote insolite. Ces extraits ont ensuite été retranscrits orthographiquement, de manière légèrement simplifiée. Puis ils ont été reproduits en parole lue et en parole interprétée par quatre locuteurs possédant tous une

expérience de l'art dramatique. Les locuteurs ont d'abord effectué la tâche de lecture à haute voix. Puis, une semaine plus tard, ils ont effectué la tâche d'interprétation, après avoir mémorisé le texte. L'étape suivante a consisté à relever les occurrences de focalisation dans le corpus. Nous avons sollicité pour cela un groupe de dix experts en prosodie. Les experts ont reçu pour consigne d'annoter les enregistrements de manière auditive en se référant tous aux mêmes critères acoustiques. Chaque enregistrement du corpus a été annoté par quatre experts. Seules les occurrences de focalisation relevées par au moins trois experts sur quatre ont été sélectionnées pour l'analyse prosodique. Pour réaliser cette analyse, le corpus a d'abord été aligné de manière semi-automatisée en mots, en syllabes et en phones. Des mesures de hauteur et de durée ont été effectuées de manière automatisée sur toutes les syllabes du corpus. Les occurrences de focalisation ont également été classées selon trois catégories phonologiques. La première catégorie est le type de contour intonatif, qui a été déterminé de façon manuelle à l'aide du système de transcription de la prosodie ToBI (dans sa version adaptée au français). La deuxième catégorie est l'étendue syllabique du contour intonatif sur le constituant focalisé, également déterminée de façon manuelle. La troisième catégorie est la présence d'accent initial, qui a été déterminée de manière automatisée à l'aide d'un algorithme de détection des proéminences.

Les résultats de l'étude sont ensuite présentés, à commencer par le taux d'accord entre les experts concernant le relevé des occurrences de focalisation. Le taux d'accord pour le corpus entier est relativement bas. On observe des différences entre les phonogenres : la parole spontanée a le taux d'accord le plus élevé, suivie de la parole interprétée. La fréquence d'occurrence de la focalisation (pourcentage de syllabes focalisées) est de 11,3% dans tout le corpus. Des différences significatives sont observées selon le phonogenre : la parole interprétée présente la fréquence d'occurrence la plus élevée (14,2%), suivie de la parole lue (10,7%). Concernant l'analyse prosodique, on observe tout d'abord que la focalisation est réalisée par une augmentation significative de hauteur et de durée syllabique. L'augmentation de hauteur n'est pas influencée par le phonogenre. En revanche, on observe que la focalisation est significativement plus marquée par la durée en parole lue qu'en parole interprétée. Au niveau phonologique, la focalisation est réalisée par une variété de contours intonatifs. De manière majoritaire, ces contours sont de forme montante et commencent par un ton haut. Concernant l'étendue syllabique, la majorité des contours sont portés par la dernière syllabe du constituant focalisé. Parmi les constituants focalisés polysyllabiques, une

proportion importante bien que minoritaire comporte un accent initial. Nous présentons également les résultats d'une analyse menée sur un corpus réduit, constitué des enregistrements ayant obtenu les meilleurs taux d'accord entre les experts. Les résultats sont globalement similaires aux résultats de l'analyse menée sur le corpus entier. Les exceptions concernent les différences de taux d'accord entre les phonogenres, ainsi que la distribution des frontières de groupes prosodiques au niveau des constituants focalisés.

Une interprétation méthodologique est faite du taux d'accord entre les experts. La faiblesse du taux d'accord dans le corpus entier limite la portée des résultats de l'analyse. Elle pourrait être due à des défauts dans le protocole d'annotation du corpus par les experts. Plusieurs facteurs ont pu causer les différences de taux d'accord selon le phonogenre : le protocole d'annotation, le protocole d'enregistrement des locuteurs de parole lue et interprétée, la présence d'un phonostyle professionnel chez ces locuteurs, ou bien la présence d'une différence de naturalité par rapport aux locuteurs de parole spontanée. Les questions de recherche 1 et 2 sont ensuite abordées. La fréquence d'occurrence de la focalisation dans le corpus entier (11,3%) est comparée au pourcentage de proéminences détecté de manière automatisée dans le corpus (35,8%). Cette comparaison montre qu'environ un tiers des proéminences du français sont des accents de focalisation. Concernant la réalisation prosodique de la focalisation, l'augmentation significative de hauteur et de durée syllabique confirme plusieurs études précédentes. L'analyse phonologique est, elle aussi, globalement cohérente avec la littérature. Nous analysons le contexte syntaxique et sémantique dans lequel apparaissent les contours intonatifs les plus fréquents. Nous analysons également la distribution des contours intonatifs selon leur étendue syllabique et selon la présence d'un accent initial sur le constituant focalisé. Les questions de recherche 3 et 4 sont ensuite abordées. L'étude confirme que la fréquence d'occurrence de la focalisation augmente avec le degré de « préparation » du phonogenre. Cependant, il est possible que d'autres facteurs aient influencé le résultat. L'absence globale d'effet du phonogenre sur la réalisation prosodique de la focalisation contredit en partie la littérature. Le fait que la focalisation soit plus marquée par la durée en parole lue qu'en parole interprétée est un résultat d'autant plus intéressant qu'il n'est pas dû à une différence de débit entre ces deux phonogenres. Il est donc possible que cet allongement soit caractéristique de la lecture à haute voix. Pour conclure, des propositions sont faites pour renforcer cette étude, en améliorant le protocole expérimental et en prenant en compte plus de traits dans l'analyse prosodique.

Chapitre 3 : Réalisation prosodique des fonctions de la focalisation

1. Introduction

1.1. Questions de recherche

Dans le chapitre précédent, nous nous sommes intéressés à la focalisation prosodique en tant que phénomène global de soulignement d'un constituant dans un énoncé. L'analyse qui a été proposée de la fréquence d'occurrence et de la réalisation prosodique de la focalisation n'a pas pris en compte les différentes fonctions que la littérature attribue à ce phénomène. Comme nous avons déjà eu l'occasion de le mentionner, on peut diviser ces fonctions en deux groupes : le marquage de focus et l'emphase. Le premier groupe concerne les fonctions de nature sémantico-pragmatique de la focalisation. Le second groupe concerne, comme son nom l'indique, les fonctions de nature emphatique de la focalisation. Dans ce chapitre, nous souhaitons savoir quelle est la réalisation prosodique de chacune de ces fonctions. Nous souhaitons également savoir si le phonogène (parole interprétée, lue ou spontanée) a un effet sur cette réalisation. Les deux questions de recherche suivantes sont posées :

Question 1 : Les fonctions attribuées à la focalisation présentent-elles des réalisations différentes ?

Question 2 : La réalisation des fonctions de la focalisation est-elle influencée par le phonogène ?

Dans les parties suivantes (cf § 2 et § 3), un état de l'art est consacré à la définition théorique et à la réalisation prosodique de chacune des fonctions. Nous présentons ensuite les deux études que nous avons menées afin de répondre aux questions de recherche. La première étude (cf § 4) porte sur le même corpus multi-genres que l'étude présentée dans le chapitre précédent. Cette étude a consisté à établir une classification fonctionnelle des occurrences de focalisation ayant été relevées. Une analyse prosodique, identique à celle que nous avons déjà présentée (cf chap. 2, § 4.3), a ensuite été effectuée. La seconde étude (cf § 5) est un test de perception. Ce test a consisté à obtenir des jugements de naturalité sur des énoncés à la

prosodie manipulée. Contrairement à la première, cette étude ne prend pas en compte l'effet du phonogène sur la réalisation prosodique des fonctions de la focalisation.

1.2. Arrière-plan et prédictions

De nombreuses études ont déjà été consacrées à la réalisation prosodique des différentes fonctions attribuées à la focalisation (cf § 2.2 et § 3.2). Cependant, aucune étude, à notre connaissance, n'a déjà effectué une analyse de toutes les fonctions à la fois dans le même corpus. La question 1 a donc pour but d'établir, dans un cadre contrastif, des généralisations descriptives concernant la réalisation de chacune des fonctions. Cela nous permettra de savoir si ces fonctions, qui sont distinctes au niveau sémantico-pragmatique, le sont également au niveau prosodique. La réponse permettra en retour de savoir si la focalisation est uniquement un ensemble de marques prosodiques fonctionnellement sous-spécifiées, ou s'il s'agit d'un terme descriptif qui regroupe de véritables catégories phonologiques.

La question 2 est inspirée de Godement-Berline (2014), une étude-pilote sur la réalisation des fonctions de la focalisation dans un corpus de parole interprétée. Dans cette étude, une proportion significative des occurrences de focalisation remplissant la fonction emphatique d'insistance (cf § 3.1.1) est réalisée au moyen d'une désaccentuation post-focale. Or, nous verrons plus bas (cf § 3.2.1) que cette réalisation n'est pas décrite dans la littérature pour cette fonction. Il est donc possible que le phonogène ait une influence sur la réalisation de l'insistance, ainsi que sur celle des autres fonctions de la focalisation. Nous nous attendons en particulier à observer des réalisations spécifiques en parole interprétée.

2. Le marquage de focus

Le marquage de focus désigne l'ensemble des fonctions sémantiques ou pragmatiques de la focalisation. Il existe plusieurs fonctions de ce type car, comme nous allons le voir dans la section suivante (cf § 2.1), il existe plusieurs catégories de focus. Notons que nous employons l'expression « marquage de focus », et non le terme « focus », pour désigner cet ensemble de fonctions. Ce choix est motivé par le fait que, comme nous allons le voir plus bas (cf § 2.1.1), Lambrecht (1994 : 206-218) montre que ces deux notions sont distinctes en ce qui concerne la catégorie du focus informationnel. En effet, le focus est une relation entre une information

et une proposition, tandis que le marquage de focus (*focus marking*) est l'indication de cette relation.

2.1. Catégories de focus

Avant de présenter les principales catégories de focus employées en linguistique, notons que la notion globale de focus trouve son origine dans un autre domaine scientifique. En effet, elle est vraisemblablement inspirée des théories psychologiques de la perception et de l'attention de la fin du XIX^{ème} siècle. On peut notamment citer le gestaltisme (e.g. Koffka 1922) et sa notion de « figure » se détachant sur un « fond », ou bien les travaux de William James sur le « courant de la conscience » (cf James 1890). En linguistique, la notion de focus connaît des emplois différents selon les domaines et les théories :

- **focus informationnel** : théorie de la structure informationnelle
- **focus contrastif** : sémantique formelle, pragmatique
- **focus associatif ou quantificationnel** : sémantique formelle
- **verum focus** : sémantique formelle
- mais aussi :
 - **focus de l'intérêt**³³ : théorie de l'intonation de Bolinger (cf Bolinger 1986, 1989)
 - **focus de l'empathie**³⁴ : syntaxe fonctionnelle de Kuno (cf Kuno et Kaburaki 1977, Kuno 1987)

³³ Dans la théorie de Bolinger, le focus de l'intérêt (*focus of interest*) désigne, comme son nom l'indique, une information contenue dans l'énoncé pour laquelle le locuteur manifeste de l'intérêt (cf Bolinger 1986 : 89-136). « *The accent [...] shows something that the speaker is keyed up about, something of keen interest, real or pretended* » (ibid., p. 90). Le focus de l'intérêt est indiqué au moyen de l'accentuation. Il s'agit de *fish* dans l'énoncé (i) ci-dessous, où l'accent est indiqué par des majuscules. La dimension de l'intérêt est indépendante de la dimension informationnelle. En effet, il est possible d'accentuer un mot qui a déjà été prononcé et qui a donc un statut informationnel « ancien » ou « donné » (cf § 2.1.1). C'est le cas de *fish* dans l'énoncé (ii). De plus, le locuteur peut marquer à la fois un focus informationnel et un focus de l'intérêt dans le même énoncé. C'est le cas en (iii), où l'accent sur *why* marque le focus informationnel, et l'accent sur *leave* le focus de l'intérêt.

- | | | |
|-------|--|----------------|
| (i) | Who likes raw FISH? | (ibid., p. 90) |
| (ii) | Raw fish is good for you, but after all, who likes raw FISH? | (ibid., p. 90) |
| (iii) | – John left early.
– WHY did he LEAVE? | (ibid., p. 96) |

³⁴ Dans la théorie de Kuno, le focus de l'empathie (*focus of empathy*) désigne, dans un énoncé, le participant dont le locuteur adopte le point de vue. « *Empathy is the speaker's identification, which may vary in degree, with a person/thing that participates in the event or state that he describes in a sentence* » (Kuno 1987 : 206). Le focus

Nous présentons ci-dessous les catégories du focus informationnel, du focus contrastif, du focus associatif et du verum focus. Il s'agit des catégories qui ont été le plus largement reprises dans la littérature. Nous nous attardons sur la catégorie du focus informationnel, qui a fait l'objet du plus grand nombre de travaux.

2.1.1. Focus informationnel

Au sein de la théorie de la structure informationnelle, la notion de focus informationnel a reçu plusieurs définitions. Avant de les présenter, notons que ces définitions postulent toutes une partition similaire de l'énoncé. Cette partition consiste globalement à distinguer d'une part ce dont on parle, et d'autre part ce que l'on dit à propos de ce dont on parle³⁵. Le focus informationnel désigne le second terme de cette partition : la nouvelle information exprimée par le locuteur concernant une information déjà connue de l'auditeur. Notons par ailleurs qu'on distingue généralement deux types de focus informationnel :

- **le focus informationnel « large »** : Il s'agit des cas où l'énoncé entier constitue une nouvelle information. Autrement dit, la totalité du contenu de l'énoncé est assertée par le locuteur.
- **le focus informationnel « étroit »** : il s'agit des cas où seule une partie de l'énoncé constitue une nouvelle information. Autrement dit, seule une partie du contenu de l'énoncé est assertée par le locuteur.

de l'empathie est indiqué au moyen de différentes constructions syntaxiques. Dans l'énoncé (i) ci-dessous, l'emploi du groupe nominal possessif *John's brother* indique que le locuteur adopte le point de vue de John. Dans l'énoncé (ii), l'emploi d'une construction passive indique que le locuteur adopte le point de vue de Bill. Il existe une règle (*ban on conflicting empathy foci*) qui interdit d'adopter deux points de vue incompatibles dans un même énoncé. Cette règle explique la relative inacceptabilité de l'énoncé (iii).

- | | | |
|-------|---------------------------------------|-----------------|
| (i) | John's brother hit him. | (ibid., p. 205) |
| (ii) | Then Bill was hit by John. | (ibid., p. 205) |
| (iii) | ? Then John's brother was hit by him. | (ibid., p. 206) |

³⁵ Notons que cette partition remonte à l'Antiquité. Elle correspond à la distinction faite par Aristote (*Organon*) entre le « sujet » et le « prédicat » d'une proposition (respectivement *hupokeimenon* : « ce qui est placé à la base », et *katēgórēma* : ce qui est dit « à propos de » quelque chose). Elle correspond également à une distinction faite par les grammairiens arabes classiques (*mubtada* : « initial », et *ḥabar* : « nouvelle information/ce qu'on dit de quelqu'un ou quelque chose »).

Il est courant d'illustrer cette distinction au moyen de paires question-réponse, comme dans les exemples (1) à (3) ci-dessous (traduits de Lambrecht 1994 : 121). En (1), la réponse contient un marquage de focus informationnel large, car aucune des informations contenues dans l'énoncé n'a déjà été formulée dans la question qui précède. En (2) et (3), la réponse contient un marquage de focus informationnel étroit, car seule une information (respectivement *les enfants* et *l'école*) n'a pas déjà été formulée dans la question qui précède. Nous verrons plus bas (cf § 2.2.1) comment ces deux types de focus informationnel se distinguent au niveau prosodique.

(1) – Que s'est-il passé ?

– Les enfants sont partis à l'école. (focus informationnel large)

(2) – Qui est parti à l'école ?

– Les enfants sont partis à l'école. (focus informationnel étroit)

(3) – Où sont partis les enfants ?

– Les enfants sont partis à l'école. (focus informationnel étroit)

La notion de focus informationnel a reçu, dans l'ordre chronologique, les définitions suivantes :

- **définition psychologique** : La distinction entre sujet et prédicat (cf note 35) a été redéfinie à un niveau cognitif par von der Gabelentz (1869), puis par Paul (1880). Selon ce dernier, tout énoncé contient un « sujet psychologique » et un « prédicat psychologique ». C'est à cette seconde catégorie que correspond le focus informationnel. Le sujet et le prédicat psychologiques ne correspondent pas toujours à leurs équivalents grammaticaux. Ils remplissent des fonctions psychologiques différentes : « *Le sujet psychologique est le groupe d'idées d'abord présent dans la conscience du locuteur ou du penseur ; [...] ce à quoi le locuteur souhaite faire penser l'auditeur, et vers quoi il aimerait diriger son attention ; le prédicat psychologique, ce qu'il souhaite qu'il pense à ce propos* » (ibid., § 170). Les mêmes notions ont ensuite reçu d'autres appellations : « thème » et « rhème »

(Amman 1928, Mathesius 1961), « thème » et « propos » (Bally 1932), et « *topic* » et « *comment* » (Hockett 1958).

- **définition informationnelle** : Les notions de sujet et de prédicat psychologiques ont ensuite été redéfinies au niveau de l'énoncé, en tant que message vecteur d'information. Les expressions « structure informationnelle » et « focus informationnel » sont introduites par Halliday (1967). Le focus informationnel est selon lui un procédé servant à indiquer l'information apportée par l'énoncé : « *Information focus is one kind of emphasis, that whereby the speaker marks out a part (which may be the whole) of a message block as that which he wishes to be interpreted as informative.* » Halliday introduit également les notions de « nouveauté » et d'« ancienneté » d'une information, qui seront reprises par la suite (e.g. Chafe 1974, Lambrecht 1994). Une information est dite nouvelle lorsque le locuteur la présente comme « non-récupérable » à partir du discours qui précède³⁶.
- **définition sémantique** : La notion de « présupposition » est introduite par Chomsky (1970). Il s'agit de la partie de l'énoncé dont le contenu est présupposé sémantiquement par le locuteur. Par opposition, le focus est la partie de l'énoncé dont le contenu est asserté par le locuteur. Pour déterminer le focus de l'énoncé, Jackendoff (1972) postule la présence d'un marqueur syntaxique F dans la structure de surface. Il formule une règle

³⁶ L'approche informationnelle est également représentée par la théorie de la FSP (*functional sentence perspective*) de l'école de Prague (e.g. Mathesius 1961, Vachek 1966, Daneš 1974, Firbas 1992). Contrairement aux autres théories informationnelles, cette théorie ne postule pas de partition du contenu de l'énoncé. Le concept central est celui du « dynamisme communicatif ». Il s'agit du degré d'importance des différents constituants de l'énoncé dans la réalisation du but communicatif du locuteur. Le dynamisme communicatif augmente par défaut du début à la fin de l'énoncé. Il peut être influencé par trois facteurs (cf Firbas 1992) :

- **le facteur « contextuel »** : Il s'agit de la possibilité ou non pour une information d'être « retrouvée » à partir du « contexte d'intérêt immédiat » (ibid., p. 21). Cette distinction est similaire à la distinction entre ancienneté et nouveauté d'une information. Une information nouvelle possède un degré de dynamisme communicatif plus fort qu'une information ancienne.
- **le facteur « sémantique »** : Il s'agit d'une classification des constituants de l'énoncé selon différentes dimensions : leur « fonction dynamique » de « présentation » ou de « qualité », leur rôle informationnel (thème, rhème ou « transition »), ainsi que les spécifications du rôle informationnel (« propre » ou « orienté »). Un constituant possède un degré de dynamisme communicatif plus ou moins grand selon les catégories auxquelles il appartient dans cette classification.
- **l'intonation** : Le degré de dynamisme communicatif d'un constituant augmente lorsque celui-ci est distingué prosodiquement.

d'assignement du focus : « *The semantic material associated with surface structure nodes dominated by F is the FOCUS of the sentence. To derive the Presupposition, substitute appropriate semantic variables for the focused material* » (ibid., p. 240). Jackendoff donne en exemple l'énoncé « *Is it JOHN who writes poetry?* », où *John* est le focus. La présupposition peut être exprimée de la manière suivante : « quelqu'un écrit de la poésie », ou encore « x écrit de la poésie ». Le contenu sémantique de l'énoncé est partitionné entre un terme non saturé (la présupposition) et un terme saturé (le focus). Ceci est représenté par Jackendoff en notation lambda : la présupposition est la fonction $\lambda x \text{Presupp}_s(x)$, et $\text{Focus} \in \lambda x \text{Presupp}_s(x)$. La formalisation sémantique de la structure informationnelle a ensuite été approfondie par Krifka (1992) et Büring (1997).

- **définition pragmatique** : Plusieurs auteurs ont tout d'abord explicité les notions de nouveauté et d'ancienneté d'une information au niveau pragmatique. Selon Clark et Haviland (1977), la communication repose sur un accord entre les locuteurs concernant le statut informationnel des constituants de l'énoncé (*given-new contract*). Cet accord contient une « maxime d'antécédence » (similaire aux maximes de la communication de Grice 1975) qui assure la possibilité de retrouver l'antécédent d'une information ancienne : « *Try to construct your utterance such that the listener has one and only [one] direct antecedent for any given information and that it is the intended antecedent.* » Selon Chafe (1974), le statut informationnel est basé sur les présuppositions du locuteur concernant le contenu de la conscience de l'auditeur au moment où il s'adresse à lui : « *Given information is suggested to be that which the speaker assumes to be already present in the addressee's consciousness at the time of an utterance.* » Prince (1981) montre qu'il existe une distinction entre ce type de statut informationnel (*discourse new/old information*), et le fait qu'une information soit présente ou pas dans les connaissances de l'auditeur (*hearer new/old information*). Elle donne en exemple l'énoncé « *Where were your grandparents born?* », où *your grandparents* est un référent de discours identifiable par l'auditeur, même s'il n'est pas présent dans sa conscience au moment de l'échange. Enfin, comme nous l'avons évoqué plus haut (cf § 2), une distinction importante est établie par Lambrecht (1994 : 206-218). Le focus ne consiste pas en un constituant de l'énoncé, ni même en son référent de discours. Il s'agit de la relation entre ce référent et une proposition qui exprime la présupposition de l'énoncé.

Lambrecht donne en exemple l'échange « – Where did you go last night? – I went to the MOVIES » (ibid., p. 209). L'information exprimée par la réponse n'est pas *movies* mais la proposition abstraite « L'endroit où je suis allé hier soir est le cinéma ». Le constituant qui reçoit l'accentuation exprimant la relation de focus est le « domaine focal » de l'énoncé. Le focus est défini comme l'information par laquelle l'assertion diffère de la présupposition.

Parmi les différentes approches présentées ci-dessus, la définition que nous avons retenue du marquage de focus informationnel pour notre analyse se base sur l'approche sémantique et sur l'approche pragmatique. Elle repose sur les notions d'assertion et de statut informationnel nouveau :

Marquage de focus informationnel : Le locuteur distingue ce qu'il asserte spécifiquement, c'est-à-dire l'information qu'il présente à l'interlocuteur comme nouvelle dans la conversation en cours. Il s'agit de la totalité du contenu dans le cas d'un focus informationnel large, et d'une partie seulement du contenu dans le cas d'un focus informationnel étroit.

2.1.2. Focus contrastif

Cette catégorie de focus concerne les cas où le constituant focalisé entre dans une relation de contraste avec une information qui précède. Dans l'exemple (4) ci-dessous, la focalisation sur *head* marque ainsi un contraste avec *back*. Cependant, selon Bolinger (ibid. : 91-95) et Lambrecht (1994 : 286-291), on peut considérer qu'il s'agit là d'un cas de marquage de focus informationnel étroit. L'information nouvelle exprimée par le locuteur peut être paraphrasée par la proposition « ce qui me fait mal est ma tête ». L'interprétation contrastive est uniquement due au contexte qui précède, c'est-à-dire le fait que le mot *back* entre en relation avec la proposition « quelque chose te fait mal ». Par opposition, la même focalisation dans l'exemple (5) n'a pas d'interprétation contrastive, car le contexte qui précède ne contient pas la même présupposition. Bolinger (1989 : 342) considère par ailleurs qu'il n'existe pas de marques prosodiques spécifiques au marquage de focus contrastif.

(4) – I suppose your back aches.

– You’re wrong. My HEAD aches. (Bolinger 1986 : 91)

(5) – You’re looking depressed. What’s the matter?

– My HEAD aches. (ibid.)

Cependant, il existe plusieurs arguments en faveur de l’existence de la catégorie du focus contrastif :

- Chafe (1976) considère qu’il existe bien une différence prosodique entre le marquage de focus contrastif et le marquage de focus informationnel étroit. L’augmentation de hauteur et d’intensité serait plus importante dans le cas du marquage de focus contrastif.
- Chafe (1974) montre qu’un référent de discours peut entrer dans une relation de contraste même lorsqu’il n’exprime pas d’information nouvelle. C’est le cas avec la focalisation sur le pronom *he* dans l’énoncé « HE killed Cock Robin ». En effet, d’après Chafe, un référent de discours au statut informationnel nouveau ne peut pas être pronominalisé.
- Dans l’exemple (4) ci-dessus, la partie qui suit le constituant focalisé (*aches*) a déjà été formulée dans le tour précédent, d’où la ressemblance avec les cas de marquage de focus informationnel étroit. Cependant, dans l’exemple (6) ci-dessous, la focalisation sur *play* est suivie d’informations nouvelles. Il ne peut donc pas s’agir d’un cas de marquage de focus informationnel étroit. Ce genre de cas est qualifié de « *topic* contrastif » par Lambrecht (1994 : 291-295).

(6) The PLAY, John saw yesterday. (Chafe 1976)

Afin de résoudre ces divergences, Vallduví et Vilkuña (1998) ont développé une notion qu’ils nomment « *kontrast* ». Une information est dite « *kontrastive* » dès lors que sa formulation génère un ensemble d’alternatives dont elle fait partie. Cette définition rend compte à la fois des cas comme (4), où le constituant focalisé est la seule information nouvelle (*identificational kontrast*), et des cas comme (6), où le constituant focalisé est le *topic* (*thematic kontrast*). Le *kontrast* est une dimension indépendante de la structure informationnelle de l’énoncé. Il peut

en effet affecter n'importe quel constituant, quel que soit son statut informationnel. C'est sur cette notion de *kontrast* que se base la définition que nous adoptons du marquage de focus contrastif pour notre analyse :

Marquage de focus contrastif : Le locuteur distingue un élément appartenant à un ensemble d'alternatives présent dans le contexte de discours immédiat.

2.1.3. Focus associatif

Cette catégorie de focus concerne les cas où le constituant focalisé est l'argument d'un adverbe tel que *seulement*. C'est le cas de *Bill* et de *Sue* dans les exemples (7) et (8) ci-dessous. Selon beaucoup d'auteurs (e.g. Jackendoff 1972, Jacobs 1983, von Stechow 1990, Rooth 1992, Krifka 1992, Krifka 2006), ce genre d'exemple constitue lui aussi un cas de marquage de focus informationnel étroit. Les adverbes *seulement*, *aussi* et *même* (ainsi que certaines expressions négatives et exclamatives) ont été décrites comme des « particules sensibles au focus » (PSF). Le phénomène illustré en (7) et (8) a été nommé « association avec le focus ». Le contraste entre (7) et (8) montre, selon ces auteurs, que l'interprétation sémantique d'un énoncé contenant une PSF dépend du focus informationnel de l'énoncé.

(7) John only introduced BILL to Sue. (Krifka 1992)

(8) John only introduced Bill to SUE. (ibid.)

Cependant, Vallduví et Zacharski (1994) montrent qu'il existe des cas où l'argument d'une PSF n'est pas le domaine focal de l'énoncé :

- Dans l'exemple (9) ci-dessous, *John* est associé à *even*, et *water* est associé à *only*. Or, un énoncé ne peut contenir par définition qu'un seul focus informationnel (puisqu'il ne peut contenir plusieurs assertions). Un des deux constituants accentués n'est donc pas le domaine focal de l'énoncé.

(9) Even JOHN drank only WATER.

- Dans l'exemple (10), la partie de la réponse qui suit *Petr* est constituée d'informations anciennes (puisqu'elles ont été formulées dans le tour précédent). Par conséquent, l'argument de *only* dans la réponse (soit *Xerox copies*, soit *the graduate students*) ne peut pas être le domaine focal de l'énoncé. Seul *Petr* peut être le domaine focal.

(10) – Eva only gave Xerox copies to the graduate students.

– No, PETR only gave Xerox copies to the graduate students.

- Dans l'exemple (11), *only* et son argument *yours* font tous deux partie du domaine focal de l'énoncé. Le domaine focal ne peut pas être uniquement constitué de *yours*.

(11) I will go to bed with my man. But you'll only dream about YOURS.

Ces exemples montrent que le marquage de focus associatif n'est pas de nature informationnelle. Il s'agit de la distinction d'un constituant qui est l'argument d'un opérateur sémantique réalisé à distance de cet argument. La définition du focus associatif que nous avons retenue pour notre analyse est donc la suivante :

Marquage de focus associatif : Le locuteur distingue l'argument d'un opérateur sémantique tel que *seulement* ou *aussi*.

2.1.4. Verum focus

Cette catégorie de focus concerne les cas où le constituant focalisé est un auxiliaire, un verbe ou un complémenteur. Gutzmann et Castroviejo Miró (2011) donnent les exemples suivants :

(12) Karl HAT sein Buch beendet. « Karl a bel et bien fini son livre »

(13) Karl SCHREIBT ein Drehbuch. « Karl écrit bel et bien un scénario »

(14) Ich denke, DASS er ein Philosoph ist. « Je pense qu'il est bel et bien philosophe »

Dans ces trois énoncés, ce qui est souligné n'est pas la dénotation du constituant focalisé mais la vérité de la proposition entière. Selon Höhle (1992), qui est à l'origine de l'expression *verum focus*, la focalisation réalise un opérateur sémantique sous-jacent. Tout comme pour le marquage de focus contrastif et le marquage de focus associatif, ce genre de cas ne peut pas être assimilé au marquage de focus informationnel étroit. La définition que nous retenons du *verum focus* pour notre analyse est donc la suivante :

Marquage de *verum focus* : Le locuteur insiste sur la vérité de la proposition.

2.2. Réalisation prosodique

Nous présentons ci-dessous les résultats des études sur la réalisation prosodique du marquage des différentes catégories de focus en français. Nous évaluons ensuite le poids des méthodes expérimentales de ces études dans les résultats observés.

2.2.1. Traits du marquage de focus informationnel

Comme nous l'avons vu plus haut (cf § 2.1.1), on distingue deux types de marquage de focus informationnel : le marquage de focus informationnel large, et le marquage de focus informationnel étroit. Dans le premier cas, la totalité du contenu de l'énoncé est assertée par le locuteur. Dans le second cas, seule une partie du contenu de l'énoncé est assertée par le locuteur. D'après la plupart des études, ces deux types de marquage de focus informationnel connaissent une réalisation prosodique différente. Dans le cas du marquage de focus informationnel large, tous les constituants majeurs de l'énoncé reçoivent une accentuation. Dans le cas du marquage de focus informationnel étroit, le constituant dont le référent de discours entre en relation de focus est focalisé. De plus, une désaccentuation a lieu sur la séquence qui suit ou qui précède le constituant focalisé (cf chap. 1, § 4.3). Afin d'illustrer cette différence de réalisation, nous reprenons ci-dessous les exemples (1) à (3) que nous avons vus plus haut. En (15), les trois constituants majeurs sont accentués dans la réponse, car l'énoncé n'est constitué que d'informations nouvelles par rapport à la question. En (16) et (17), au contraire, seul le constituant qui constitue une information nouvelle (respectivement *enfants* et *école*) est accentué.

(15) – Que s'est-il passé ?

– Les ENFANTS sont PARTIS à L'ÉCOLE. (focus informationnel large)

(16) – Qui est parti à l'école ?

– Les ENFANTS sont partis à l'école. (focus informationnel étroit)

(17) – Où sont partis les enfants ?

– Les enfants sont partis à L'ÉCOLE. (focus informationnel étroit)

En plus de la différence globale de réalisation illustrée par les exemples (15) à (17), les résultats suivants sont observés :

- **marquage de focus informationnel large** : On observe généralement une réalisation prosodique « par défaut » du point de vue de l'accentuation et du phrasé. D'après Féry (2001), les constituants majeurs sont tous accentués, et sont phrasés dans des groupes prosodiques séparés. Toutefois, les constituants très courts (par exemple un sujet monosyllabique) sont parfois dépourvus d'accent, et sont phrasés dans le même groupe que le constituant adjacent. Du point de vue de l'intonation, Dohen et Løevenbruck (2004) observent les formes suivantes sur des énoncés contenant un sujet, un verbe et un objet : contour montant de fin de groupe accentuel H* sur le sujet et le verbe, et contour terminal descendant de fin de groupe intonatif L% sur l'objet. Un ton haut initial Hi est parfois également présent sur le sujet et le verbe. Cependant, dans l'étude de Beyssade et al (2009), le contour terminal ne tombe à la fin de l'énoncé que dans 69% des cas. Dans 31% des cas, il tombe au milieu de l'énoncé (une réalisation qui est généralement associée au marquage de focus informationnel étroit).
- **marquage de focus informationnel étroit**³⁷ : On observe généralement la présence d'un contour intonatif terminal sur le constituant focalisé, et d'une désaccentuation sur la

³⁷ Notons qu'en français, le marquage de focus informationnel étroit peut également faire appel à des traits syntaxiques. Le constituant dont le référent de discours entre en relation de focus peut en effet se trouver au centre d'une construction clivée (ex : « – Qui est parti à l'école ? – Ce sont les ENFANTS qui sont partis à l'école »). Selon Lambrecht (1994 : 13-25), cette construction accompagne forcément les traits prosodiques en français. Cependant, de nombreuses études montrent que les traits prosodiques peuvent suffire à réaliser le marquage de focus informationnel étroit en français (e.g. Jun et Fougeron 2000, Dohen et Løevenbruck 2004, Beyssade et al 2009). Par ailleurs, le marquage de focus informationnel étroit n'est pas la seule fonction possible de la construction clivée. Cette dernière peut en effet servir à introduire une information, sans la faire entrer dans une

séquence qui suit ou qui précède (e.g. Delattre 1966, Rossi 1999 : 84-87, Martin 2009 : 124-130). Cependant, plusieurs études montrent que la désaccentuation n'est pas systématique (e.g. Di Cristo 1999b, Féry 2001, Beyssade et al 2009). Dans l'étude de Beyssade et al, le contour terminal ne tombe sur le constituant focalisé que dans 60% des cas. Dans 40% des cas, il tombe à la fin de l'énoncé (une réalisation qui, comme nous l'avons vu, est généralement associée au marquage de focus informationnel large). Du point de vue de l'accentuation, le constituant focalisé peut porter un accent initial, qui forme un arc accentuel avec l'accent final du constituant (Di Cristo 1999b, Beyssade et al 2009). Du point de vue du phrasé, de même que pour le marquage de focus informationnel large, le constituant focalisé et le constituant pré-focal sont généralement phrasés dans des groupes séparés (Féry 2001). Cependant, les constituants courts peuvent être phrasés dans le même groupe qu'un constituant adjacent.

2.2.2. Traits du marquage de focus contrastif

Comme nous l'avons vu plus haut (cf § 2.1.2), on peut distinguer deux types de marquage de focus contrastif. Dans le premier cas, le constituant focalisé est suivi de constituants au statut informationnel ancien. Le second cas est celui que Lambrecht (1994 : 291-285) qualifie de « *topic* contrastif ». Ces deux types de marquage de focus contrastif, que nous avons illustrés plus haut par les exemples (4) et (6), connaissent des réalisations différentes :

- Dans le cas où le constituant focalisé est suivi d'informations anciennes, on observe une réalisation similaire à celle généralement observée pour le marquage de focus informationnel étroit (cf § 2.2.1). Le constituant focalisé porte un contour intonatif terminal, et la séquence qui précède ou qui suit est désaccentuée (Di Cristo 1999b, Jun et Fougeron 2000, Dohen et Lœvenbruck 2004). D'après Jun et Fougeron, le contour terminal est un ton haut. Il est porté par la syllabe initiale ou finale du constituant focalisé, ou bien par le constituant entier. D'après Dohen et Lœvenbruck, une augmentation de durée a également lieu sur le constituant focalisé (surtout sur le premier segment), ainsi que sur

relation d'assertion (Doetjes et al 2003). C'est le cas dans l'énoncé suivant : « C'est avec plaisir que je vous invite à ce séminaire ».

la dernière syllabe de la séquence pré-focale. Enfin, d'après Di Cristo, le constituant focalisé est marqué par un accent initial rythmique ou emphatique (cf chap. 1, Tableau 1).

- Dans le cas du *topic* contrastif, Beyssade et al (2004) observent la présence d'un accent initial spécifique (« accent C ») sur le constituant focalisé. Cet accent porte sur la première syllabe. Il est réalisé par une forte augmentation de hauteur, un allongement du segment initial de la syllabe accentuée, et une augmentation d'intensité.

2.2.3. Traits du marquage de focus associatif

Il existe, à notre connaissance, seulement deux études sur la réalisation prosodique de cette fonction en français. Elles portent toutes deux sur la distinction prosodique de l'argument de l'adverbe *seulement*. D'après Gendrot et Raynal (2004), une augmentation de hauteur a lieu sur le constituant focalisé. Cette augmentation affecte plus spécifiquement la dernière voyelle. D'après Beyssade et al (2008), la réalisation a lieu à deux niveaux :

- Le « domaine d'association » dans lequel se trouve le constituant focalisé est délimité par une frontière finale plus forte que celle qui délimite le syntagme où apparaît l'adverbe *seulement*. Les constituants qui se trouvent après cette frontière ne peuvent être associés à *seulement*. Les auteurs donnent en exemple l'énoncé (18) ci-dessous. Une frontière finale de groupe accentuel apparaît après *seulement*. Le domaine d'association est *remanié le roman*. Ce domaine est délimité par une frontière finale de groupe intonatif, plus forte que la frontière de groupe accentuel (cf chap. 1, § 3). Le constituant *le mois dernier* est exclu du domaine d'association de *seulement*.

(18) J'ai seulement]_{GA} remanié le roman]_{GI} le mois dernier.

- A l'intérieur du domaine d'association, l'« associé étroit » de *seulement* doit être distingué prosodiquement. Cette distinction est réalisée par un accent initial ou bien des variations de registre tonal et de tempo. Dans l'exemple (18), l'associé étroit de *seulement* est soit *remanié*, soit *roman*.

2.2.4. Traits du marquage de verum focus

L'unique étude sur la réalisation prosodique de cette fonction en français est, à notre connaissance, celle de Turco et al (2013). Cette étude compare des énoncés réalisés avec ou sans marquage de verum focus. Dans l'exemple (19) ci-dessous, le deuxième énoncé contient un marquage de verum focus sur le groupe verbal *a déchiré*.

(19) – Sur mon image, l'enfant n'a pas déchiré le billet.

– Sur mon image, l'enfant a déchiré le billet.

Les résultats suivants sont observés :

- Dans 49% des énoncés contenant un marquage de verum focus, on observe sur le groupe verbal des réalisations qu'on n'observe pas dans les énoncés sans marquage de verum focus. Il s'agit de la présence d'un ton haut initial Hi sur l'auxiliaire ou sur le participe passé. Ce dernier est par ailleurs délimité à droite par une frontière de groupe accentuel dans la majorité des cas.
- Dans 94% des énoncés contenant un marquage de verum focus, on observe une désaccentuation post-focale après le groupe verbal. Par contraste, seuls 4% des énoncés sans marquage de verum focus présentent une désaccentuation post-focale.

2.2.5. Poids des méthodes expérimentales

Les études que nous avons présentées n'emploient pas toutes les mêmes méthodes expérimentales. Ces différences doivent être prises en compte dans la comparaison des résultats observés. Nous décrivons ci-dessous (cf Tableau 18) le protocole de chaque étude selon trois aspects : les phonogenres étudiés, la taille du corpus, et le cadre théorique utilisé. Une case vide signifie que l'information n'est pas rendue explicite par les auteurs.

	phonogenres	taille du corpus	cadre théorique
Delattre (1966)	parole spontanée, parole interprétée, conférences		modèle de Delattre
Rossi (1999)			modèle de Rossi
Di Cristo (1999b)			modèle autosegmental- métrique
Jun et Fougeron (2000)	parole lue	225 énoncés	modèle autosegmental- métrique
Féry (2001)	parole lue	400 énoncés	modèle autosegmental- métrique
Dohen et Lœvenbruck (2004)	parole lue	64 énoncés	modèle autosegmental- métrique
Gendrot et Raynal (2004)	parole lue	420 énoncés	modèle autosegmental- métrique
Beysade et al (2004)			modèle autosegmental- métrique
Beysade et al (2008)	parole lue	77 énoncés	modèle autosegmental- métrique
Beysade et al (2009)	parole lue	112 énoncés	modèle autosegmental- métrique
Martin (2009)	parole lue, parole spontanée	plusieurs heures	modèle de Martin
Turco et al (2013)	parole semi- spontanée	103 énoncés	modèle autosegmental- métrique

Tableau 18. Informations sur le protocole expérimental des études sur la réalisation prosodique du marquage des différentes catégories de focus.

3. L'emphase

L'emphase désigne l'ensemble des fonctions expressives de la focalisation prosodique. Nous empruntons le terme d'emphase à Fónagy (1980) et Di Cristo (1999a). Dans les études sur la focalisation, on trouve généralement deux types de fonctions emphatiques. Le premier type

est celui qui a été qualifié d'« insistance » par le courant de la phonostylistique en français (e.g. Carton 1973, Séguinot 1976, Fónagy 1983). Le second type est l'expression d'émotions et d'attitudes (e.g. Fónagy 1983, Léon 1993, Rossi 1999, Di Cristo 1999a). Nous qualifions ce second type d'emphase d'« expressivité ». Nous donnons ci-dessous une définition de chacune de ces deux fonctions. Nous décrivons ensuite les résultats observés concernant leur réalisation prosodique³⁸.

3.1. Catégories d'emphase

3.1.1. Insistance

Comme nous venons de le mentionner, la notion d'insistance a été introduite par le courant de la phonostylistique. Dans ce courant, l'insistance n'est en fait pas définie en tant que catégorie mais est associée à un accent initial de mot (« accent d'insistance »). Cet accent peut

³⁸ Notons que la notion d'« accent de puissance » (*accent of power*) de Bolinger (1986 : 74-88) concerne également les fonctions emphatiques de la focalisation prosodique. Nous n'avons pas retenu cette notion, car elle a été développée pour l'anglais et n'a pas été reprise par les études sur la focalisation en français. Selon Bolinger, la fonction des accents de puissance est d'« impressionner » ou d'« émouvoir » l'auditeur. Ces accents, qui portent sur l'énoncé entier, se divisent en quatre types :

- ***climax*** : Il s'agit du déplacement d'un accent lexical ou nucléaire vers la droite afin de renforcer le message. Dans l'exemple (i) ci-dessous, l'accent lexical de *absolute* est déplacé sur la dernière syllabe du mot. Dans l'exemple (ii), l'accent nucléaire de l'énoncé, qui devrait tomber par défaut sur *helped*, est déplacé sur la dernière syllabe de l'énoncé.
 - (i) His rule is absoLUTE.
 - (ii) When the roast came by, he helped himSELF.
- ***anticlimax*** : Il s'agit, comme son nom l'indique, de l'inverse de l'accentuation de *climax*. L'accent lexical ou nucléaire est déplacé vers la gauche afin de tempérer le message. Dans l'exemple (iii) ci-dessous, l'accent lexical de *absolutely* est déplacé sur la première syllabe du mot. Dans l'exemple (iv), l'accent nucléaire de l'énoncé, qui devrait tomber par défaut sur *ingenious*, est déplacé sur *most*.
 - (iii) ABSolutely.
 - (iv) I thought that that was MOST ingenious.
- ***exclamatory early accent*** : Il s'agit d'une accentuation au début de l'énoncé, en raison d'un « afflux d'émotion », comme illustré par l'exemple (v) ci-dessous.
 - (v) WHY oh why did it have to happen?
- ***cumulative accents*** : Il s'agit de la présence d'accents additionnels avant un accent de *climax* à la fin de l'énoncé, comme illustré par les exemples (vi) et (vii) ci-dessous.
 - (vi) ExCUSE ME.
 - (vii) IT MAKES a MESS.

remplir de nombreuses fonctions. D'après Carton (1973), ces fonctions sont d'ordre emphatique mais aussi sémantico-pragmatique : « *Depuis un demi-siècle, la terminologie relative à l'insistance est considérable, mais parfois confuse : on parle d'une part d'accent pathétique, emphatique, affectif, expressif, d'autre part d'accent antithétique, distinctif, oppositif, logique, psychologique, différentiel.* » Séguinot (1976) ajoute : « *Enfin, l'insistance remplit des fonctions identificatrices (émotions) et impressives (démarcatives, contrastives et oppositionnelles).* » La division que fait Carton entre deux ensembles de fonctions correspond à la fameuse distinction faite par Marouzeau (1923) entre « accent affectif » et « accent intellectuel »³⁹. Les fonctions « contrastives » et « oppositionnelles » mentionnées par Séguinot correspondent au marquage de focus contrastif. Il est difficile de trouver dans la littérature une définition plus précise de l'insistance. Dans le cadre de sa théorie des « bases pulsionnelles de la phonation », Fónagy (1983 : 107-115) propose une définition de l'accent « emphatique » (qui contient pour lui l'accent d'insistance). Toutefois, cette définition n'est pas de nature linguistique et ne peut donc être retenue pour notre analyse. Les seules définitions linguistiques de l'insistance sont celles proposées par Rossi (1999 : 116-129) et Di Cristo (1999b) :

- Selon Rossi, la focalisation prosodique possède une fonction de « proéminence sémantique », qui consiste à « *mettre en surbrillance les éléments sur lesquels l'allocutaire doit porter prioritairement son attention* ».
- Selon Di Cristo, l'insistance correspond à une « focalisation d'intensification », ou encore « focalisation emphatique étroite », qui consiste à « *promouvoir un item de la chaîne* ».

La définition que nous avons retenue de l'insistance pour notre analyse est basée sur ces deux dernières définitions :

³⁹ D'après cette distinction, l'accent affectif sert à exprimer une émotion, tandis que l'accent intellectuel sert à mettre en valeur un mot (cf Milner et Regnault 1987 : 133-135). Au niveau phonétique, l'accent affectif porterait sur la première consonne du mot (ex : « c'est divin », « c'est horrible »), tandis que l'accent intellectuel porterait sur la première syllabe (« c'est divin », « c'est horrible »). Cependant, d'après l'étude de Séguinot (1976), il n'existe pas de différence phonétique entre ces deux types d'accent.

Insistance : Le locuteur souligne, pour l'interlocuteur, l'importance d'une partie ou de la totalité du contenu de son énoncé.

3.1.2. Expressivité

Il est difficile de trouver dans la littérature une définition précise de la fonction d'expressivité. Les travaux de Fónagy (1983) sur les émotions concernent (de même que pour l'insistance) leurs bases pulsionnelles, et ne sont donc pas pertinents d'un point de vue formel. Comme nous l'avons mentionné dans le chapitre 1 (cf § 1.2), une distinction est établie par Léon (1993 : 21-22) entre les émotions et les attitudes. Les émotions (ex : colère, joie, tristesse) sont exprimées sans contrôle conscient de la part du locuteur. Les attitudes (ex : ironie, doute, évidence) sont au contraire exprimées de façon consciente par le locuteur⁴⁰. Selon Di Cristo (1999b), l'expressivité est une « emphase de cumul » portant sur l'énoncé entier. Cette notion est inspirée de la notion de *climax* de Bolinger (cf note 38). La tentative la plus approfondie de formalisation de l'expressivité est probablement la théorie de Potts (2007). Cette théorie, développée à partir d'une analyse de mots expressifs en anglais (ex : *damn* et *bastard*), identifie six propriétés du contenu expressif⁴¹. Cependant, Potts ne propose pas de définition

⁴⁰ Selon Scherer (2005), les émotions et les attitudes sont deux catégories appartenant à l'ensemble des « phénomènes affectifs » (*affective phenomena*). Cet ensemble inclut également d'autres catégories (*preferences, moods, affect dispositions, interpersonal stances*). Scherer distingue par ailleurs deux types d'émotions (*utilitarian* et *aesthetic*). Les phénomènes affectifs sont classés par Scherer selon un ensemble de traits (*design features*) pour lesquels la valeur du phénomène affectif varie de « très bas » à « très haut » : *event focus, intrinsic appraisal, transactional appraisal, synchronization, rapidity of change, behavioral impact, intensity, duration*. Cette description se base sur la théorie psychologique des émotions de Scherer (1986), dite modèle des « processus composants » (*component process model*). Dans ce modèle, les émotions sont envisagées comme des phénomènes dynamiques qui résultent de l'évaluation, par différents « sous-systèmes de l'organisme » (*organistic subsystems*), de stimuli externes ou internes, en fonction de leur importance pour la survie et le bien-être de l'organisme.

⁴¹ Les six propriétés du contenu expressif identifiées par Potts sont les suivantes :

- ***independence*** : Le contenu expressif est traité séparément du contenu descriptif par l'auditeur.
- ***nondisplaceability*** : Le contenu expressif ne peut s'appliquer à un contexte (temps, lieu, locuteur) autre que celui de l'énoncé.
- ***perspective dependence*** : Le contenu expressif est toujours associé à un certain point de vue, généralement celui du locuteur.
- ***descriptive ineffability*** : Le contenu expressif ne peut être paraphrasé au moyen de termes descriptifs.
- ***immediacy*** : Comme les énoncés performatifs, les mots expressifs réalisent leur intention communicative par le simple fait d'être prononcés.
- ***repeatability*** : La répétition d'un mot expressif crée un effet de renforcement plutôt que de redondance.

de l'expressivité en tant que fonction globale. Etant donné le manque de convergence de la littérature, la définition que nous avons retenue de l'expressivité pour notre analyse reste volontairement simple. Elle conserve uniquement la distinction entre émotions et attitudes :

Expressivité : Le locuteur exprime une émotion ou une attitude par rapport à une partie ou la totalité du contenu de son énoncé.

3.2. Réalisation prosodique

3.2.1. Traits de l'insistance

Notons que les études sur la réalisation prosodique de l'insistance ne s'intéressent qu'à un type de réalisation : l'accentuation initiale du constituant focalisé. Comme nous l'avons expliqué plus haut (cf § 3.1.1), c'est souvent cette accentuation qui définit elle-même la fonction d'insistance dans ces études. Un problème de circularité se présente donc, puisque les études ne se basent pas sur une définition purement théorique de l'insistance (pour analyser ensuite la réalisation de cette fonction). Par ailleurs, nous avons vu que les fonctions attribuées à l'« accent d'insistance » par la phonostylistique sont très nombreuses, et incluent en partie la dimension sémantico-pragmatique. Les résultats suivants peuvent néanmoins être cités :

- L'accent initial d'insistance est réalisé au moyen d'une augmentation de hauteur, de durée et d'intensité (Séguinot 1976, Di Cristo 1999b, Delais-Roussarie et Di Cristo, à paraître). L'augmentation de durée concerne surtout la consonne initiale. D'après Séguinot, la hauteur joue un rôle plus important que les deux autres paramètres.
- Lorsque la syllabe commence par une voyelle, un coup de glotte est généralement réalisé avant l'accent (Séguinot 1976, Di Cristo 1999b). Une augmentation de la pression sous-glottique est effectivement mesurée dans l'étude de Fónagy (1980), alors qu'elle est absente chez les accents finaux.
- Une analyse de la constituance syllabique et de la configuration tonale de l'accent d'insistance (« accent initial emphatique ») a été effectuée par Astésano (2001), comme nous l'avons vu dans le chapitre 1 (cf § 2.4). Pour rappel, la syllabe accentuée se

caractérise par une attaque plus longue que la rime, ainsi que par un pic de hauteur de forme saillante et asymétrique.

- Enfin, notons que la fréquence d'occurrence de l'accent initial d'insistance est influencée par plusieurs facteurs. Le premier facteur est la structure syllabique du constituant focalisé. D'après Fónagy (1980), dont l'étude porte sur des constituants de deux syllabes, la probabilité de réalisation d'un accent initial est élevée quand la première syllabe est fermée, ou quand la seconde syllabe est ouverte. La probabilité est en revanche faible quand la seconde syllabe est fermée (en particulier quand elle se termine par une consonne fricative ou occlusive). D'après Lucci (1980), la probabilité de réalisation est plus élevée avec les mots commençant par une consonne (en particulier occlusive). Cela est d'autant plus le cas quand ces mots commencent par un groupe consonantique. Lucci montre aussi qu'un rôle est joué par la position, la taille et la rareté du mot. Un mot qui se trouve en début ou en fin de groupe accentuel a plus de chances de recevoir une accentuation initiale. Il en va de même pour les mots longs et rares⁴².

3.2.2. Traits de l'expressivité

Il existe de nombreuses études sur la réalisation prosodique de l'expressivité. Cependant, la plupart de ces études ne portent pas sur le français. De plus, beaucoup d'études ne concernent pas la réalisation de l'expressivité en tant que fonction globale mais la réalisation d'émotions ou d'attitudes spécifiques (e.g. Lieberman et Michaels 1962, Uldall 1964, Fónagy 1983, Léon 1993). Enfin, certaines études ne s'intéressent pas explicitement à la focalisation prosodique. Nous présentons néanmoins les résultats de ce dernier groupe d'études (déjà présenté dans le chapitre 1, cf § 4.4), car ils peuvent s'appliquer en partie à la focalisation :

- L'étude de Selting (1994) montre que les attitudes sont globalement marquées par une forte densité de syllabes accentuées, ainsi que par des groupes prosodiques courts et

⁴² L'étude de Lucci ne porte pas explicitement sur l'accent d'insistance, mais sur un accent qu'il qualifie de « didactique ». Cet accent « caractérise en synchronie le parler d'une couche socio-culturelle ou professionnelle (intellectuels, politiciens, annonceurs, enseignants) ayant l'habitude de s'adresser à un auditoire ». Nous assimilons cet accent à l'accent d'insistance, car Lucci le décrit comme un accent initial, et lui attribue une fonction de « mise en valeur intellectuelle ».

isochroniques. Les accents se caractérisent par une augmentation de hauteur et d'intensité nettement supérieure aux accents des unités voisines.

- Certaines études observent des corrélations entre des traits prosodiques et des « dimensions » expressives servant à décrire les émotions et attitudes. Au niveau de la qualité vocalique, la « tension » de la voix augmente avec le degré d'*activation*, une dimension liée notamment au degré d'importance et de nouveauté du stimulus (Scherer 1986). Le registre tonal s'élève quant à lui avec le degré d'*intensité*, c'est-à-dire le niveau d'excitation physiologique de l'organisme (Scherer et al 1984, Bänziger et Scherer 2005, Audibert et al 2010).

La réalisation de l'expressivité en français au moyen de la focalisation prosodique a été étudiée par Rossi (1999 : 116) et Di Cristo (1999b). Une forte augmentation de hauteur et de durée est observée sur la syllabe qui porte le contour terminal de l'énoncé. Cela a pour effet d'augmenter l'amplitude du contour terminal. On observe, d'après Di Cristo, « *une chute mélodique ample (de l'ordre d'une octave), issue d'un registre aigu* ».

3.2.3. Poids des méthodes expérimentales

Nous présentons ci-dessous (cf Tableau 19) des informations concernant les méthodes expérimentales des études sur la réalisation des fonctions emphatiques de la focalisation. Ces informations ont pour but de pouvoir comparer les résultats de ces études, malgré les différences concernant les méthodes employées. Comme dans le tableau concernant la réalisation prosodique du marquage de focus (cf § 2.2.5), une case vide signifie que l'information n'est pas rendue explicite par les auteurs.

	phonogenres	taille du corpus	cadre théorique
Séguinot (1976)			phonostylistique
Fónagy (1980)	parole lue, parole spontanée, journal radiophonique, conférences, lecture de comptines		phonostylistique
Lucci (1980)	parole lue, conférences	601 énoncés	phonostylistique
Scherer et al (1984)	interviews	66 énoncés	théorie de l'affect vocal (<i>vocal affect</i>)
Scherer (1986)			modèle des « processus composants »
Selting (1994)	parole spontanée		analyse conversationnelle
Rossi (1999)			modèle de Rossi
Di Cristo (1999b)			modèle autosegmental-métrique
Astésano (2001)	parole lue, journal radiophonique, interviews	plusieurs heures	modèle autosegmental-métrique
Bänziger et Scherer (2005)	parole interprétée	144 énoncés	théorie de l'affect vocal (<i>vocal affect</i>)
Audibert et al (2010)	parole spontanée, parole interprétée	193 énoncés	théorie de l'affect vocal (<i>vocal affect</i>)
Delais-Roussarie et Di Cristo (à paraître)			modèle autosegmental-métrique

Tableau 19. Informations sur le protocole expérimental des études sur la réalisation prosodique des différentes catégories d'emphase.

4. Etude de production

4.1. Introduction

Cette première étude a pour but de répondre aux deux questions de recherche de cette analyse (cf § 1.1). Nous rappelons ces questions ci-dessous :

Question 1 : Les fonctions attribuées à la focalisation présentent-elles des réalisations différentes ?

Question 2 : La réalisation des fonctions de la focalisation est-elle influencée par le phonogène ?

L'étude consiste à obtenir une classification fonctionnelle d'occurrences de focalisation prosodique, puis à effectuer une analyse prosodique des ces occurrences. Le croisement entre la classification fonctionnelle et l'analyse prosodique permet de répondre à la question 1. Le corpus d'étude est le même que celui de l'étude précédente (cf chap. 2, § 4.1). Pour rappel, ce corpus comprend trois phonogènes (parole spontanée, parole lue et parole interprétée). La prise en compte de l'effet du phonogène dans l'analyse prosodique permet de répondre à la question 2. La classification fonctionnelle a été effectuée par le même groupe d'experts qui a relevé les occurrences de focalisation dans le corpus (cf chap. 2, § 4.2). L'analyse prosodique est la même que celle de l'étude précédente (cf chap. 2, § 4.3).

4.2. Méthodologie

4.2.1. Constitution du corpus

Nous rappelons ci-dessous les principales étapes de la constitution du corpus, déjà décrites dans le chapitre précédent (cf § 4.1) :

- Deux dialogues spontanés (un dialogue entre deux hommes, et l'autre entre deux femmes) ont été enregistrés. Pour chaque dialogue, nous avons retenu un extrait dans lequel un des locuteurs raconte une anecdote insolite qui lui est arrivée.
- Les extraits ont été transcrits orthographiquement, de manière légèrement simplifiée. Des marques de ponctuation ont été ajoutées.

- Les extraits ont ensuite été reproduits par quatre locuteurs (deux femmes et deux hommes). Ces locuteurs sont différents des locuteurs originaux. Ils possèdent tous, à des degrés divers, une expérience de l'art dramatique. Les locuteurs ont uniquement reproduit le texte du locuteur qui raconte l'anecdote dans l'extrait original. La tâche de reproduction s'est déroulée en deux parties :
 - Les locuteurs ont d'abord lu le texte à voix haute. Ils ont découvert le texte à cette occasion. Ils ont reçu pour consigne de prononcer le texte de manière naturelle, et de respecter autant que possible les mots exacts du texte.
 - Les locuteurs ont ensuite interprété le texte de mémoire. Cette seconde partie a eu lieu environ une semaine après la première partie (afin que les locuteurs aient le temps de mémoriser le texte). Les locuteurs ont reçu les mêmes consignes que pour la tâche de lecture à voix haute.

4.2.2. Classification des occurrences de focalisation

Comme nous l'avons vu dans le chapitre précédent (cf § 4.2), un relevé des occurrences de focalisation a été effectué dans le corpus par un groupe d'experts en prosodie. Nous rappelons ci-dessous les principales étapes de ce relevé :

- Nous avons sélectionné dix experts pour annoter le corpus. Les experts sont de niveaux différents et n'ont pas tous le même domaine de spécialisation (cf chap. 2, Tableau 7). Ils possèdent tous une expérience de l'étude de la prosodie.
- Le corpus, qui comprend dix enregistrements, a été réparti parmi les experts. Chaque enregistrement a été annoté par quatre experts (cf chap. 2, Tableau 8).
- Les experts ont effectué leur annotation de manière purement auditive (sans s'aider d'un logiciel d'analyse acoustique). Ils se sont tous référés aux mêmes critères acoustiques pour relever les occurrences de focalisation. Les critères acoustiques sont constitués des principaux traits attribués à la focalisation par la littérature (cf chap. 2, § 2.2). Les experts

ont également reçu pour consigne de ne pas relever les accents purement rythmiques (marquant la frontière initiale ou finale d'un groupe prosodique).

En plus de ce relevé, les experts ont eu pour tâche d'établir une classification fonctionnelle des occurrences de focalisation. Nous leur avons remis une liste des différentes fonctions attribuées à la focalisation par la littérature (cf § 2.1 et § 3.1). Nous leur avons demandé d'assigner une de ces fonctions à chaque occurrence, ou bien d'assigner plusieurs fonctions à la même occurrence. Les experts ont également eu la possibilité de choisir des fonctions absentes de la liste. Nous leur avons demandé de préciser dans ce cas la nature de la fonction choisie. Afin de s'assurer que la classification fonctionnelle soit cohérente, les experts se sont tous référés aux mêmes définitions théoriques des fonctions de la focalisation. Il s'agit des définitions que nous avons présentées plus haut (cf § 2.1 et § 3.1). Nous rappelons ces définitions ci-dessous :

- **marquage de focus informationnel** : Le locuteur distingue ce qu'il affirme spécifiquement, c'est-à-dire l'information qu'il présente à l'interlocuteur comme nouvelle dans la conversation en cours. Il s'agit de la totalité du contenu dans le cas d'un focus informationnel large, et d'une partie seulement du contenu dans le cas d'un focus informationnel étroit.
- **marquage de focus contrastif** : Le locuteur distingue un élément appartenant à un ensemble d'alternatives présent dans le contexte de discours immédiat.
- **marquage de focus associatif** : Le locuteur distingue l'argument d'un opérateur sémantique tel que *seulement* ou *aussi*.
- **marquage de verum focus** : Le locuteur insiste sur la vérité de la proposition.
- **insistance** : Le locuteur souligne, pour l'interlocuteur, l'importance d'une partie ou de la totalité du contenu de son énoncé.
- **expressivité** : Le locuteur exprime une émotion ou une attitude par rapport à une partie ou la totalité du contenu de son énoncé.

Nous avons ensuite recoupé les annotations des experts entre elles⁴³. Nous avons considéré qu'il y avait accord entre les experts concernant une occurrence de focalisation lorsqu'au moins trois experts sur quatre lui ont attribué la même fonction. Seules ces occurrences ont été conservées pour l'analyse prosodique. On peut voir ci-dessous (cf Figure 37) les annotations des experts pour un énoncé du corpus. Concernant les cas où un expert a attribué plusieurs fonctions à la même occurrence, nous avons déterminé l'accord entre experts fonction par fonction. Autrement dit, même si plusieurs experts n'ont pas attribué exactement le même ensemble de fonctions à une occurrence, nous avons considéré qu'il pouvait y avoir accord concernant une ou plusieurs fonctions appartenant à cet ensemble. C'est le cas avec le mot *telle* dans la figure ci-dessous : deux experts ont attribué la fonction d'insistance à ce mot, et un troisième expert lui a attribué à la fois la fonction d'insistance et la fonction de focus informationnel étroit (il y a donc accord concernant la fonction d'insistance). Cette décision s'explique par le fait que l'adoption d'un critère plus strict (si l'accord entre experts devait concerner l'ensemble exact des fonctions attribuées) aurait trop fait diminuer le taux d'accord entre les experts (cf § 4.3.1).

⁴³ Nous avons choisi la syllabe comme unité de recoupement des annotations. Pour rappel (cf chap. 2, note 15), cette décision s'explique par le fait qu'il s'agit de l'unité que nous avons utilisée dans l'analyse prosodique. Nous avons converti les annotations des experts à cet effet.

				30.709724	0.186753	30.896477		
1				i	i	i	expert 5 (793)	
2				i fie			expert 6 (793)	
3				i			expert 9 (793)	
4					a	a	expert 10 (793)	
5	puR	Zwe	ko~	tR@	tE	le	kip	syll (115/793)
6	pour jouer contre telle équipe							ortho (228)

Figure 37. Capture d'écran du logiciel Praat contenant l'alignement texte-parole d'un énoncé, ainsi que les annotations des experts. En bas, la tire orthographique contient la transcription orthographique de l'énoncé. La tire du dessus contient la segmentation syllabique de l'énoncé (les syllabes sont transcrites phonétiquement dans l'alphabet SAMPA). Les quatre tires du dessus contiennent les annotations des experts. Les lettres indiquent la fonction assignée par les experts aux occurrences de focalisation (« i » : insistance, « fie » : focus informationnel étroit, « a » : autre fonction). On voit que trois experts ont assigné la même fonction (insistance) au mot telle.

4.2.3. Analyse prosodique

Nous rappelons ci-dessous les principales étapes de l'analyse prosodique, déjà décrites dans le chapitre précédent (cf § 4.3) :

- Un alignement texte-parole du corpus a été réalisé de manière semi-automatisée à l'aide d'EasyAlign (cf Goldman 2011). Le corpus a été segmenté en mots, en syllabes et en phones.
- Nous avons effectué des mesures de hauteur et de durée sur toutes les syllabes du corpus. Ces mesures ont été effectuées de manière automatisée à partir de l'alignement texte-parole, à l'aide de Prosogram (cf Mertens 2004). La hauteur a été normalisée par rapport aux locuteurs. La durée a été normalisée par rapport à la structure syllabique, ainsi que par rapport au débit de parole des locuteurs.
- Nous avons également classé les occurrences de focalisation selon trois catégories phonologiques. Les deux premières catégories sont le type de contour intonatif du constituant focalisé et l'étendue syllabique du contour intonatif. Ces catégories ont été

déterminées de manière auditive, ainsi qu'à l'aide d'un logiciel d'analyse acoustique. Pour le type de contour intonatif, nous nous sommes basés sur le système de transcription de la prosodie ToBI, dans sa version adaptée au français (F-ToBI, cf Delais-Roussarie et al 2015). La troisième catégorie est la présence d'accent initial sur le constituant focalisé. Cette catégorie a été déterminée de manière automatisée, à l'aide de la fonction de détection des proéminences du logiciel Analor (cf Avanzi et al 2008).

4.3. Résultats

Pour rappel, nous présentons dans le chapitre précédent (cf Tableau 9) des informations globales sur le corpus (nombre de mots, nombre de syllabes, et fréquence des différents types de structure syllabique). Ces informations sont indiquées à la fois pour le corpus entier, ainsi que pour chaque phonogène.

4.3.1. Taux d'accord entre les experts

Nous avons mesuré le taux d'accord entre les experts en prosodie concernant la classification fonctionnelle des occurrences de focalisation. Comme pour le taux d'accord concernant le relevé des occurrences (cf chap. 2, § 5.1), nous avons employé le test statistique du kappa de Fleiss (cf Fleiss 1971). Nous avons obtenu les kappas sous le logiciel R, à l'aide de la fonction *kappam.fleiss* de la bibliothèque « irr » (Gamer et al 2012). Nous indiquons ci-dessous (cf Tableau 20) le taux d'accord pour chaque fonction dans le corpus entier⁴⁴. On voit, d'après les scores- z ⁴⁵, que tous les kappas sont significatifs, sauf celui du marquage de focus associatif.

⁴⁴ Pour rappel (cf chap. 2, note 22), le test du kappa de Fleiss permet de calculer le taux d'accord pour plusieurs enregistrements à la fois, même si les experts varient d'un enregistrement à l'autre (ce qui est le cas ici, cf § 4.2.2). Notons par ailleurs que le kappa de chaque fonction a été obtenu en incluant toutes les syllabes du corpus. Une autre option aurait consisté à n'inclure que les syllabes annotées par au moins un expert sur quatre. Cependant, il s'est avéré que cette option rendait tous les kappas non-significatifs.

⁴⁵ Pour rappel, pour que le kappa soit significatif à un degré de confiance de 95%, le score- z doit être supérieur à 1,96 en valeur absolue (ce qui correspond à une valeur- p d'un maximum de 0,05).

	marquage de focus informationnel large	marquage de focus informationnel étroit	marquage de focus contrastif	marquage de focus associatif	marquage de verum focus
taux d'accord (kappa de Fleiss)	0,025 (z = 5,5)	0,046 (z = 10,1)	0,036 (z = 7,8)	-0,002 (z = -0,4)	0,051 (z = 11,1)
	marquage de focus (fonction globale)	insistance	expressivité	emphase (fonction globale)	autre fonction
	0,121 (z = 26,1)	0,166 (z = 35,9)	0,183 (z = 39,7)	0,213 (z = 46,2)	0,037 (z = 7,9)

Tableau 20. Taux d'accord entre les experts en prosodie concernant la classification fonctionnelle des occurrences de focalisation. Les scores-z sont indiqués sous les kappas. Tous les scores-z sont significatifs à $p < 0,05$, sauf celui du marquage de focus associatif ($p = 0,698$).

On voit que le taux d'accord est globalement bas pour toutes les fonctions. En effet, il est inférieur au taux d'accord concernant le relevé des occurrences de focalisation (0,284), que nous avons mesuré dans le chapitre précédent (cf § 5.1). Pour rappel, ce dernier taux d'accord est relativement bas par rapport aux taux d'accord observés dans deux études récentes sur la prosodie (Cole et al 2010, Avanzi et al 2011). Les taux d'accord les plus élevés sont ceux de l'emphase et du marquage de focus (envisagées en tant que fonctions globales), de l'insistance et de l'expressivité. Parmi ces fonctions, l'emphase présente le taux d'accord le plus élevé (0,213), suivie de l'expressivité (0,183), de l'insistance (0,166), et du marquage de focus (0,121). Concernant les autres fonctions (le marquage des différentes catégories de focus, ainsi que les fonctions annotées « autres » par les experts⁴⁶), le taux d'accord est trop

⁴⁶ Pour rappel (cf § 4.2.2), les experts avaient la possibilité de choisir des fonctions non-présentes dans la liste que nous leur avons donnée. Nous leur avons demandé, dans ce cas, de préciser la nature de la fonction choisie. Cependant, lors du calcul des taux d'accord, nous avons rangé ces fonctions dans la même catégorie (« autre fonction »). En effet, la disparité parmi les fonctions choisies par les experts est trop grande pour pouvoir calculer le taux d'accord de ces fonctions séparément.

bas pour pouvoir prendre en compte ces fonctions dans l'analyse prosodique⁴⁷. Cette dernière ne porte par conséquent que sur trois fonctions : l'insistance, l'expressivité, et le marquage de focus⁴⁸.

Nous présentons ci-dessous (cf Tableau 21) le nombre de syllabes analysées pour chacune des fonctions prises en compte dans l'analyse prosodique. Nous indiquons également la distribution de ces syllabes parmi les trois phonogenres. Le nombre total de syllabes analysées (371) est relativement bas. La distribution de ces syllabes parmi les trois fonctions est relativement hétérogène : l'insistance présente le plus grand nombre de syllabes (183), suivie du marquage de focus (102) et de l'expressivité (86). La distribution des syllabes parmi les phonogenres est également hétérogène. La parole spontanée contient peu d'occurrences de chaque fonction. Cela est logique, car le corpus comprend seulement deux enregistrements de parole spontanée (contre quatre enregistrements de parole lue et quatre enregistrements de parole interprétée, cf chap. 2, § 4.2.2). On voit que l'expressivité et l'insistance présentent une majorité de syllabes en parole interprétée (82,6% et 62,3% respectivement). Le marquage de focus présente, au contraire, une majorité de syllabes en parole lue (74,5%). Enfin, notons qu'on observe très peu de syllabes pour certaines configurations. C'est notamment le cas de l'expressivité en parole lue (5), ainsi que de l'insistance en parole spontanée (5).

⁴⁷ Nous n'avons également pas pris en compte les combinaisons de fonctions que les experts pouvaient attribuer aux occurrences de focalisation (cf § 4.2.2). En effet, le taux d'accord pour chaque combinaison présente un score-z non-significatif.

⁴⁸ Nous n'avons pas pris en compte la fonction globale d'emphase dans l'analyse prosodique, étant donné que les deux catégories qu'elle contient (l'insistance et l'expressivité) présentent elles-mêmes un taux d'accord suffisamment élevé pour être prises en compte.

	expressivité	marquage de focus	insistance
parole spontanée	10	12	5
parole lue	5	76	64
parole interprétée	71	14	114
nombre total	86	102	183

Tableau 21. Distribution des syllabes analysées selon la fonction de l'occurrence de focalisation ainsi que le phonogène de l'enregistrement.

4.3.2. Hauteur

Les effets suivants ont été évalués :

- **effet de la fonction sur la hauteur des syllabes focalisées** : On observe une gradation du point de vue numérique (cf Figure 38) : l'insistance présente une hauteur plus élevée que l'expressivité, et cette dernière présente une hauteur plus élevée que le marquage de focus. Nous avons employé un modèle linéaire mixte afin de tester la significativité de l'effet. Le modèle a été effectué sous le logiciel R, à l'aide de la fonction *lmer* de la bibliothèque « lme4 » (Bates et Maechler 2009). La variable dépendante est la hauteur, et la variable fixe est la fonction. Nous avons inclus le locuteur en tant que variable aléatoire⁴⁹. Les valeurs-p ont été obtenues en effectuant un test du rapport des vraisemblances entre le modèle ainsi qu'un modèle ne contenant pas la variable de la fonction. L'effet n'est pas significatif ($\chi^2(2) = 4,2, p > 0,1$). Notons que le modèle révèle une forte variation inter-locuteurs. Cette variation justifie probablement l'absence de

⁴⁹ Pour rappel (cf chap. 2, note 25), le groupe d'experts aurait également dû être inclus en tant que variable aléatoire. Cependant, cela n'est pas possible, car cette variable est en partie confondue avec la variable du locuteur. Dans ce modèle linéaire mixte ainsi que dans les modèles suivants, nous avons donc soit uniquement inclus le locuteur en tant que variable aléatoire, soit employé des modèles sans variable aléatoire (cf § 4.3.3).

différences significatives entre les trois fonctions (qui est contradictoire avec ce qui est observé dans la figure ci-dessous).

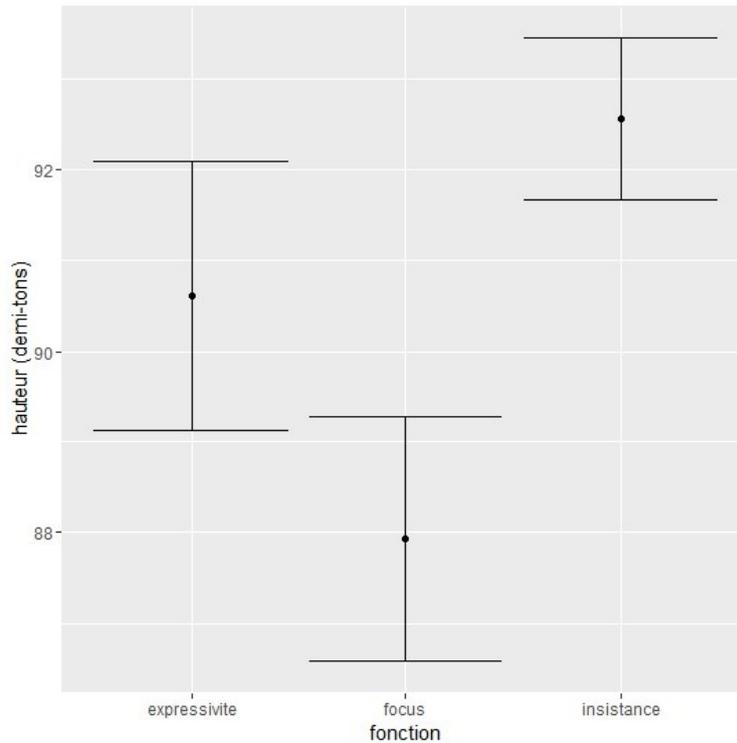


Figure 38. Hauteur moyenne des syllabes focalisées pour chaque fonction. Les barres d'erreur indiquent les intervalles de confiance à 95%.

- **effet du phonogène sur les différences de hauteur entre les fonctions** : Nous avons employé un modèle linéaire mixte avec la fonction et le phonogène comme variables fixes. Les valeurs-p ont été obtenues en effectuant un test du rapport des vraisemblances entre ce modèle et le modèle précédent (avec la fonction comme seule variable fixe). L'effet n'est pas significatif ($\chi^2(2) = 0,5$, $p > 0,1$). Notons que le modèle révèle, de même que le modèle précédent, une forte variation inter-locuteurs. Cette variation justifie probablement l'absence d'effet significatif.
- **effet de l'interaction entre fonction et phonogène sur la hauteur des syllabes focalisées** : On peut voir ci-dessous (cf Figure 39) la hauteur moyenne des syllabes focalisées selon la fonction et le phonogène. On observe certaines différences numériques. En parole spontanée, l'expressivité présente la hauteur la plus élevée, suivie

du marquage de focus et de l'insistance. En parole lue et en parole interprétée, l'insistance présente la hauteur la plus élevée, suivie de l'expressivité et du marquage de focus⁵⁰. Afin de tester la significativité de l'effet, nous avons employé un modèle linéaire mixte avec l'interaction entre fonction et phonogène comme variable fixe. Les valeurs-p ont été obtenues en effectuant un test du rapport des vraisemblances entre ce modèle et le modèle précédent (avec la fonction et le phonogène comme variables fixes). L'effet n'est pas significatif ($\chi^2(4) = 1,4, p > 0,1$). Comme les deux modèles précédents, l'absence d'effet significatif est probablement due à une forte variation inter-locuteurs.

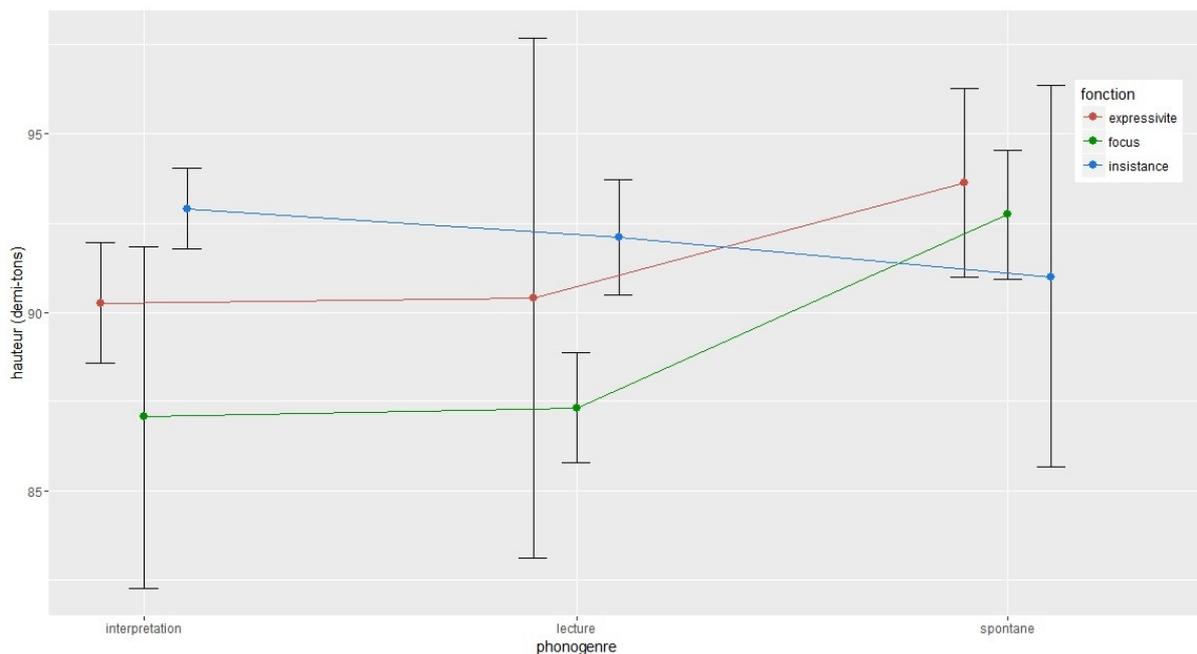


Figure 39. Hauteur moyenne des syllabes focalisées selon la fonction et le phonogène. Les barres d'erreur indiquent les intervalles de confiance à 95%.

4.3.3. Durée syllabique

Les effets suivants ont été évalués :

⁵⁰ On voit, dans la Figure 39, que certaines valeurs de hauteur présentent des intervalles de confiance très larges : c'est notamment le cas du marquage de focus en parole interprétée, de l'expressivité en parole lue, et de l'insistance en parole spontanée. Ceci est dû à la faiblesse du nombre de syllabes analysées pour ces configurations, comme nous l'avons vu plus haut (cf Tableau 21).

- **effet de la fonction sur la durée des syllabes focalisées** : On voit ci-dessous (cf Figure 40) que l'expressivité et l'insistance présentent, numériquement, une durée plus élevée que le marquage de focus. Afin de tester la significativité de l'effet, nous avons employé un modèle linéaire mixte avec la durée comme variable dépendante et la fonction comme variable fixe. Nous n'avons pas inclus le locuteur comme variable aléatoire⁵¹, car la durée a déjà été normalisée par rapport aux locuteurs (cf § 4.2.3). L'effet est significatif ($\chi^2(2) = 6,9$, $p < 0,05$). Nous présentons ci-dessous (cf Tableau 22) les résultats du modèle. On voit, d'après les valeurs-t et les valeurs-p⁵², que les différences entre le marquage de focus et les deux autres fonctions sont significatives. La différence entre l'expressivité et l'insistance n'est en revanche pas significative.

⁵¹ Etant donné que ce modèle linéaire mixte ne contient aucune variable aléatoire (cf note 49), nous ne l'avons pas effectué à l'aide de la fonction *lmer* (cf § 4.3.2), mais à l'aide de la fonction générique *glm* du logiciel R.

⁵² Comme avec les scores-z (cf note 45), pour que la différence de durée entre deux fonctions soit significative à 95%, la valeur-t doit être supérieure à 1,96 en valeur absolue, ce qui correspond à une valeur-p d'un maximum de 0,05.

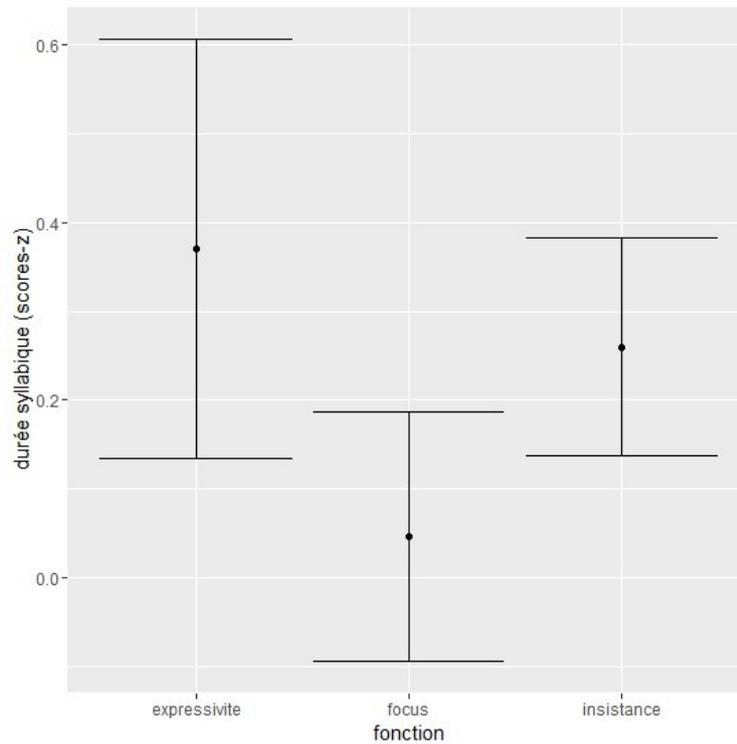


Figure 40. Durée moyenne des syllabes focalisées pour chaque fonction. Les barres d'erreur indiquent les intervalles de confiance à 95%.

	estimation (et erreur standard)	valeur-t	valeur-p
expressivité/ insistance	0,11 (0,12)	0,962	0,3366
expressivité/ marquage de focus	0,32 (0,13)	2,529	0,0119
insistance/ marquage de focus	0,21 (0,11)	1,978	0,0487

Tableau 22. Résultats du modèle linéaire mixte concernant l'effet de la fonction sur la durée des syllabes focalisées du corpus.

- **effet du phonogène sur les différences de durée syllabique entre les fonctions** : Nous avons employé un modèle linéaire mixte avec la fonction et le phonogène comme

variables fixes (et pas de variable aléatoire, comme dans le modèle précédent). L'effet est significatif ($\chi^2(2) = 7,3$, $p < 0,05$). Nous présentons ci-dessous (cf Tableau 23) les résultats du modèle. Concernant les phonogenres, on voit que seule la différence entre la parole lue et la parole interprétée est significative (cette différence a déjà été observée dans l'étude précédente, cf chap. 2, § 5.3.2). Toutefois, cette différence n'affecte pas les différences entre les fonctions. Comme dans le modèle précédent, on observe des différences significatives entre le marquage de focus et les deux autres fonctions. La différence entre l'expressivité et l'insistance est en revanche non-significative.

		estimation (et erreur standard)	valeur-t	valeur-p
différences entre phonogenres	parole spontanée/lue	-0,09 (0,19)	-0,484	0,628671
	parole spontanée/ interprétée	0,21 (0,19)	1,101	0,27151
	parole lue/interprétée	0,30 (0,11)	2,678	0,007748
différences entre fonctions	expressivité/ insistance	0,19 (0,12)	1,576	0,11594
	expressivité/ marquage de focus	0,53 (0,15)	3,572	0,000403
	insistance/ marquage de focus	0,34 (0,12)	2,915	0,003787

Tableau 23. Résultats du modèle linéaire mixte concernant l'effet de la fonction et du phonogène sur la durée des syllabes focalisées du corpus.

- **effet de l'interaction entre fonction et phonogène sur la durée des syllabes focalisées :**
On peut voir ci-dessous (cf Figure 41) la durée moyenne des syllabes focalisées selon la fonction et le phonogène. On observe certaines différences numériques. Le marquage de

focus présente, dans les trois phonogenres, une durée moins élevée que les deux autres fonctions. En parole spontanée, l'insistance présente une durée plus élevée que l'expressivité. En parole lue et en parole l'interprétée, l'expressivité présente au contraire une durée plus élevée que l'insistance⁵³. Afin de tester la significativité de l'effet, nous avons employé un modèle linéaire mixte avec l'interaction entre fonction et phonogenre comme variable fixe (et pas de variable aléatoire, comme dans les deux modèles précédents). L'effet est significatif ($\chi^2(4) = 10,6$, $p < 0,05$). Nous présentons ci-dessous (cf Tableau 24) les résultats du modèle. On observe les résultats suivants :

- La différence de durée entre l'expressivité et l'insistance varie significativement entre la parole spontanée et la parole lue ($t = 2,936$, $p < 0,01$), ainsi qu'entre la parole lue et la parole interprétée ($t = 2,333$, $p < 0,05$).
- La différence de durée entre l'expressivité et le marquage de focus varie significativement entre la parole spontanée et la parole lue ($t = 2,434$, $p < 0,05$), ainsi qu'entre la parole lue et la parole interprétée ($t = 2,597$, $p < 0,01$).

⁵³ Comme précédemment (cf note 19), on voit dans la figure 5 que certaines valeurs de durée présentent des intervalles de confiance très larges (expressivité en parole lue, et insistance en parole spontanée). Pour rappel, ceci est dû à la faiblesse du nombre de syllabes analysées pour les configurations concernées (cf tableau 4).

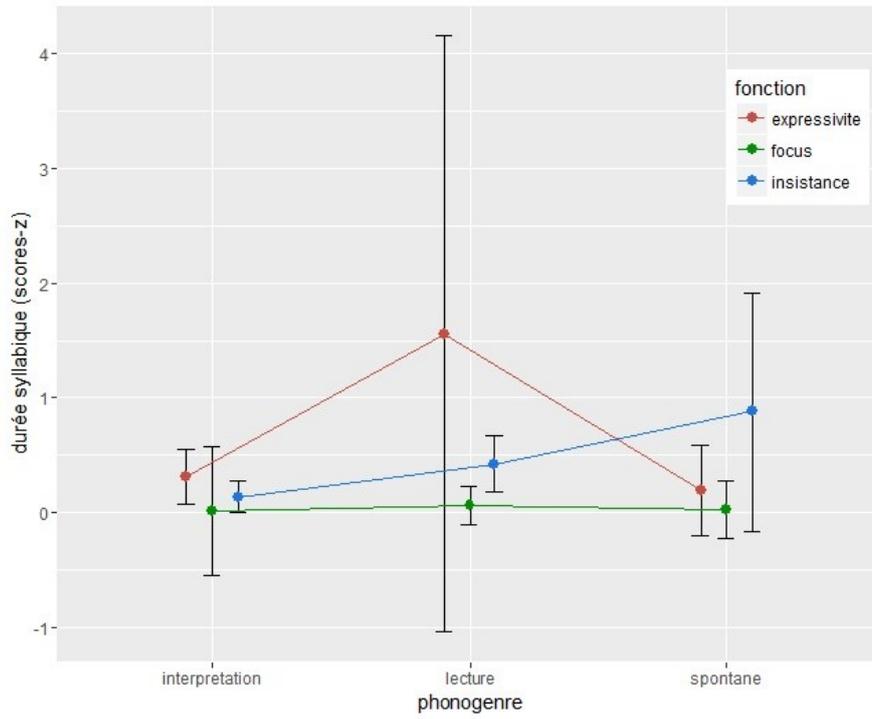


Figure 41. Durée moyenne des syllabes focalisées selon la fonction et le phonogenre. Les barres d'erreur indiquent les intervalles de confiance à 95%.

		différences entre phonogenres								
		parole spontanée/lue			parole spontanée/interprétée			parole lue/interprétée		
		estimation (et erreur standard)	valeur-t	valeur-p	estimation (et erreur standard)	valeur-t	valeur-p	estimation (et erreur standard)	valeur-t	valeur-p
différences entre fonctions	expressivité/ insistance	1,83 (0,62)	2,936	0,00354	0,86 (0,50)	1,721	0,08605	0,97 (0,41)	2,333	0,02020
	expressivité/ marquage de focus	1,33 (0,55)	2,434	0,0154	0,14 (0,46)	0,295	0,7681	1,20 (0,46)	2,597	0,009802
	insistance/ marquage de focus	0,49 (0,47)	1,034	0,30203	0,73 (0,51)	1,420	0,15656	0,24 (0,28)	0,841	0,400724

Tableau 24. Résultats du modèle linéaire mixte concernant l'effet de l'interaction entre fonction et phonogène sur la durée des syllabes focalisées du corpus.

4.3.4. Contour intonatif

Les effets suivants ont été évalués :

- **effet de la fonction sur le type de contour intonatif des constituants focalisés** : On peut voir ci-dessous (cf Figure 42), pour chaque contour intonatif, le pourcentage d'occurrence pour chaque fonction. Nous avons exclu les contours les moins fréquents (L*, L*H, L*H% et L*L-), pour lesquels la comparaison serait insignifiante. On observe plusieurs différences numériques. L'insistance présente plus de contours Hi que les deux autres fonctions, et moins de contours H*L% et H*H%. Le marquage de focus présente plus de contours H*H-, H*L- et L*L% que les deux autres fonctions, et moins de contours H*. Afin de tester la significativité de l'effet, nous avons effectué un test du khi carré de Pearson (cf Pearson 1900). Nous avons employé la fonction générique *chisq.test* du logiciel R. L'effet est significatif ($\chi^2(16) = 46,3$, $p < 0,001$). Afin de déterminer les contours intonatifs concernés, nous avons employé une régression logistique pour chaque contour. Les régressions logistiques ont été effectuées à l'aide de la fonction *glmer* de la bibliothèque « lme4 » (Bates et Maechler 2009) sous le logiciel R. La variable dépendante est le contour intonatif, la variable fixe est la fonction, et la variable aléatoire est le locuteur. Nous présentons ci-dessous (cf Tableau 25) les résultats des régressions. On observe les résultats suivants :
 - L'effet de la fonction est significatif pour les contours Hi, H*H% et H*L- uniquement.
 - Concernant le contour Hi, les différences entre l'insistance et les deux autres fonctions sont significatives. La différence entre l'expressivité et le marquage de focus est en revanche non-significative.
 - Concernant le contour H*H%, seule la différence entre l'expressivité et l'insistance est significative.

- Concernant le contour H*L-, on peut uniquement observer que la différence entre l'insistance et le marquage de focus n'est pas significative⁵⁴.

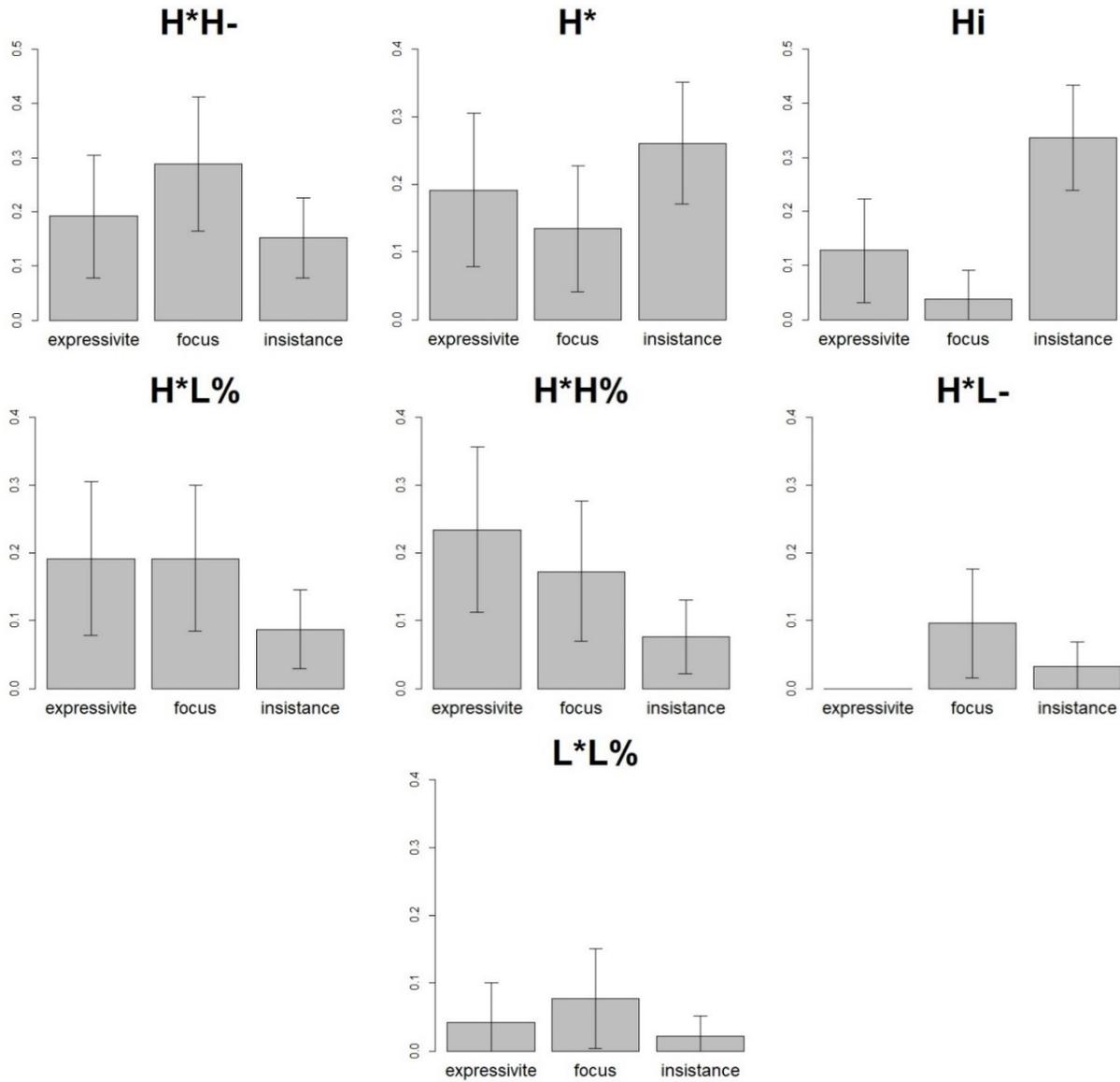


Figure 42. Pourcentages d'occurrence des différents types de contour intonatif sur les constituants focalisés pour chaque fonction. Les barres d'erreur indiquent les intervalles de confiance à 95%.

⁵⁴ Pour ce contour, concernant les deux autres différences (entre l'expressivité et les deux autres fonctions), la régression logistique donne en effet un score-z égal à 1 (et une valeur-p égale à 0).

	résultats de la régression logistique			
contour H*H-	$\chi^2(2) = 1,7, p > 0,1$			
contour H*	$\chi^2(2) = 1,0, p > 0,1$			
contour H*L%	$\chi^2(2) = 4,3, p > 0,1$			
contour L*L%	$\chi^2(2) = 1,7, p > 0,1$			
contour Hi	$\chi^2(2) = 22,9, p < 0,001$			
	différences entre fonctions	estimation (et erreur standard)	score-z	valeur-p
	expressivité/ insistance	-1,25 (0,49)	-2,543	0,010999
	expressivité/ marquage de focus	1,30 (0,84)	1,538	0,123999
	insistance/ marquage de focus	2,54 (0,75)	3,371	0,000749
contour H*H%	$\chi^2(2) = 7,9, p < 0,05$			
	différences entre fonctions	estimation (et erreur standard)	score-z	valeur-p
	expressivité/ insistance	1,46 (0,55)	2,651	0,00803
	expressivité/ marquage de focus	0,40 (0,52)	0,769	0,442162
	insistance/ marquage de focus	-1,06 (0,56)	-1,883	0,059658
contour H*L-	$\chi^2(2) = 7,1, p < 0,05$			
	différences entre fonctions	estimation (et erreur standard)	score-z	valeur-p
	expressivité/ insistance	–	–	–
	expressivité/ marquage de focus	–	–	–
	insistance/ marquage de focus	-1,12 (0,75)	-1,528	0,127

Tableau 25. Résultats des régressions logistiques concernant l'effet de la fonction sur le type de contour intonatif des constituants focalisés du corpus.

- **effet de la fonction sur le type de frontière prosodique des constituants focalisés** : Il s'agit du type de groupe prosodique (accentuel, intonatif ou intermédiaire) délimité par le contour intonatif⁵⁵ (cf chap. 2, § 5.3.3). On peut voir ci-dessous (cf Figure 43), pour chaque type de frontière, le pourcentage d'occurrence pour chaque fonction. On observe plusieurs différences numériques. L'insistance présente plus de frontières de groupe accentuel que les deux autres fonctions, et moins de frontières de groupe intonatif. Le marquage de focus présente plus de frontières de groupe intermédiaire que les deux autres fonctions. Nous avons employé un test du khi carré de Pearson afin de tester la significativité de l'effet. L'effet est significatif ($\chi^2(4) = 35,6$, $p < 0,001$). Afin de déterminer les types de frontière concernés, nous avons employé une régression logistique pour chaque type de frontière. Nous présentons ci-dessous (cf Tableau 26) les résultats des régressions logistiques. On observe les résultats suivants :
- L'effet de la fonction est significatif pour le groupe accentuel et le groupe intonatif. On observe un effet marginal pour le groupe intermédiaire.
 - Concernant le groupe accentuel, les différences entre l'insistance et les deux autres fonctions sont significatives. La différence entre l'expressivité et le marquage de focus est en revanche non-significative.
 - Concernant le groupe intonatif, les différences entre l'insistance et les deux autres fonctions sont significatives. La différence entre l'expressivité et le marquage de focus est en revanche non-significative.

⁵⁵ Pour rappel (cf chap. 2, note 28), dans notre analyse, le type de frontière prosodique désigne la frontière la plus haute qui est réalisée sur le constituant focalisé. Par conséquent, une frontière de groupe accentuel désigne une frontière qui n'est pas également une frontière de groupe intermédiaire ou intonatif. De même, une frontière de groupe intermédiaire désigne une frontière qui n'est pas également une frontière de groupe intonatif.

- Concernant le groupe intermédiaire, la différence entre le marquage de focus et l'expressivité est significative. La différence entre le marquage de focus et l'insistance approche quant à elle la significativité.

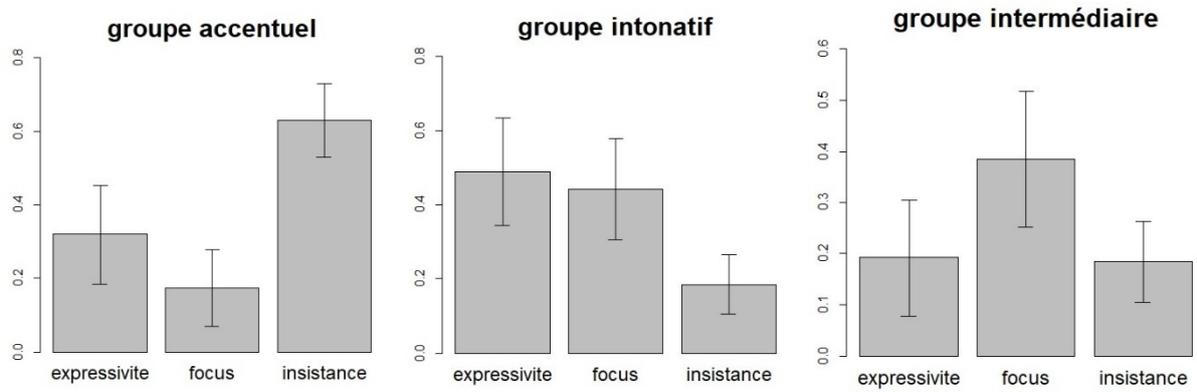


Figure 43. Pourcentages d'occurrence des différents types de frontière prosodique sur les constituants focalisés pour chaque fonction. Les barres d'erreur indiquent les intervalles de confiance à 95%.

	résultats de la régression logistique			
groupe accentuel	$\chi^2(2) = 22,1, p < 0,001$			
	fonctions	estimation (et erreur standard)	score-z	valeur-p
	expressivité/ insistance	-1,18 (0,40)	-2,912	0,00359
	expressivité/ marquage de focus	0,77 (0,49)	1,581	0,113929
	insistance/ marquage de focus	1,95 (0,45)	4,324	0,0000153
groupe intonatif	$\chi^2(2) = 15,1, p < 0,001$			
	fonctions	estimation (et erreur standard)	score-z	valeur-p
	expressivité/ insistance	1,44 (0,41)	3,511	0,000447
	expressivité/ marquage de focus	0,21 (0,41)	0,512	0,60895
	insistance/ marquage de focus	-1,23 (0,41)	-3,043	0,002341
groupe intermédiaire	$\chi^2(2) = 5,5, p = 0,06561$			
	fonctions	estimation (et erreur standard)	score-z	valeur-p
	expressivité/ insistance	-0,11 (0,49)	-0,232	0,8169
	expressivité/ marquage de focus	-0,96 (0,48)	-2,020	0,0434
	insistance/ marquage de focus	-0,85 (0,43)	-1,946	0,0516

Tableau 26. Résultats des régressions logistiques concernant l'effet de la fonction sur le type de frontière prosodique des constituants focalisés du corpus.

- **effet du phonogène sur les différences de type de contour intonatif et de frontière prosodique entre les fonctions** : Nous avons employé, pour chaque type de contour et de

frontière, une régression logistique avec le contour ou la frontière comme variable dépendante, la fonction et le phonogène comme variables fixes, et le locuteur comme variable aléatoire. Les valeurs-p ont été obtenues en effectuant un test du rapport des vraisemblances entre ces régressions et les régressions précédentes (avec la fonction comme seule variable fixe). Nous présentons ci-dessous (cf Tableau 27) les résultats des régressions. On voit que l'effet n'est significatif pour aucune des catégories.

	résultats de la régression logistique
contour H*H-	$\chi^2(2) = 1,2, p > 0,1$
contour H*	$\chi^2(2) = 1,5, p > 0,1$
contour Hi	$\chi^2(2) = 0,3, p > 0,1$
contour H*L%	$\chi^2(2) = 4,2, p > 0,1$
contour H*H%	$\chi^2(2) = 0,5, p > 0,1$
contour H*L-	$\chi^2(2) = 3,7, p > 0,1$
contour L*L%	$\chi^2(2) = 2,0, p > 0,1$
groupe accentuel	$\chi^2(2) = 0,9, p > 0,1$
groupe intonatif	$\chi^2(2) = 3,0, p > 0,1$
groupe intermédiaire	$\chi^2(2) = 0,9, p > 0,1$

Tableau 27. Résultats des régressions logistiques concernant l'effet de la fonction et du phonogène sur le type de contour intonatif et de frontière prosodique des constituants focalisés du corpus.

- **effet de l'interaction entre fonction et phonogène sur le type de contour intonatif et de frontière prosodique** : Nous avons employé, pour chaque type de contour et de frontière, une régression logistique avec le contour ou la frontière comme variable dépendante, l'interaction entre fonction et phonogène comme variable fixe, et le locuteur comme variable aléatoire. Les valeurs-p ont été obtenues en effectuant un test du rapport des vraisemblances entre ces régressions et les régressions précédentes (avec la fonction et le phonogène comme variables fixes). Nous présentons ci-dessous (cf Tableau 28) les résultats des régressions. On voit que l'effet n'est significatif pour aucune des catégories.

	résultats de la régression logistique
contour H*H-	$\chi^2(4) = 6,0, p > 0,1$
contour H*	$\chi^2(4) = 3,9, p > 0,1$
contour Hi	$\chi^2(4) = 6,0, p > 0,1$
contour H*L%	$\chi^2(4) = 6,4, p > 0,1$
contour H*H%	$\chi^2(4) = 4,1, p > 0,1$
contour H*L-	$\chi^2(4) = 6,4, p > 0,1$
contour L*L%	$\chi^2(4) = 0,7, p > 0,1$
groupe accentuel	$\chi^2(4) = 5,0, p > 0,1$
groupe intonatif	$\chi^2(4) = 2,9, p > 0,1$
groupe intermédiaire	$\chi^2(4) = 6,1, p > 0,1$

Tableau 28. Résultats des régressions logistiques concernant l'effet de l'interaction entre fonction et phonogène sur le type de contour intonatif et de frontière prosodique des constituants focalisés du corpus.

4.3.5. Etendue du contour

Les effets suivants ont été évalués :

- **effet de la fonction sur le type d'étendue syllabique du contour intonatif** : On peut voir ci-dessous (cf Figure 44), pour chaque type d'étendue, le pourcentage d'occurrence pour chaque fonction. Seuls les constituants focalisés comportant plusieurs syllabes sont concernés. Nous avons exclu les deux types d'étendue les moins fréquents (deuxième syllabe et deux premières syllabes). On observe quelques différences numériques. L'insistance présente moins de contours portant sur la dernière syllabe que les deux autres fonctions, et plus de contours portant sur la première syllabe. Le marquage de focus présente plus de contours portant sur les deux dernières syllabes, et aucun contour portant sur la première syllabe. Afin de déterminer la significativité de l'effet, nous avons effectué un test du khi carré de Pearson. L'effet n'est pas significatif ($\chi^2(10) = 13,2, p > 0,1$).

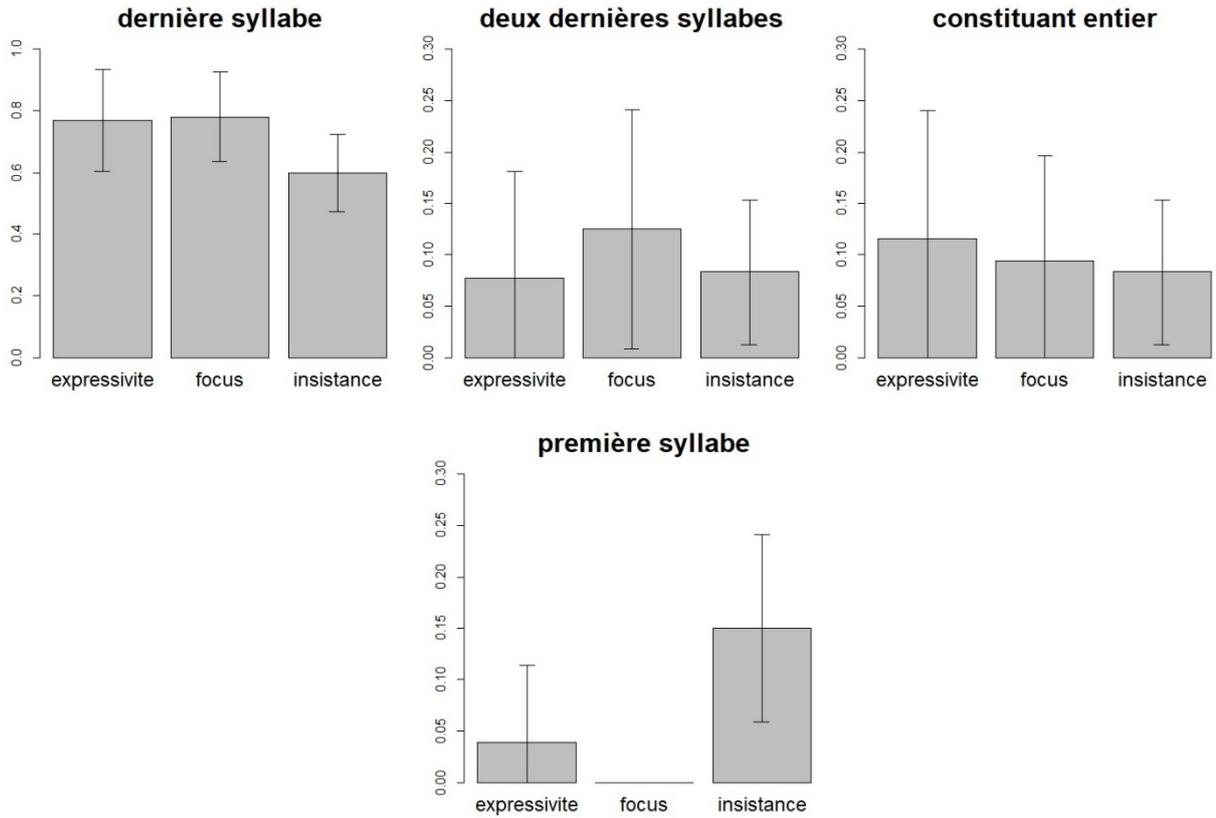


Figure 44. Pourcentages d'occurrence des différents types d'étendue syllabique du contour intonatif pour chaque fonction. Les barres d'erreur indiquent les intervalles de confiance à 95%.

- **effet du phonogène sur les différences de type d'étendue syllabique entre les fonctions** : Nous avons employé, pour chaque type d'étendue, une régression logistique avec l'étendue comme variable dépendante, la fonction et le phonogène comme variables fixes, et le locuteur comme variable aléatoire. Nous présentons ci-dessous (cf Tableau 29) les résultats des régressions. On voit que l'effet n'est significatif pour aucun type d'étendue.

	résultats de la régression logistique
dernière syllabe	$\chi^2(2) = 3,0, p > 0,1$
deux dernières syllabes	$\chi^2(2) = 5,2, p > 0,05$
constituant entier	$\chi^2(2) = 2,1, p > 0,1$
première syllabe	$\chi^2(2) = 2,0, p > 0,1$

Tableau 29. Résultats des régressions logistiques concernant l'effet de la fonction et du phonogène sur le pourcentage d'occurrences des différents types d'étendue syllabique du contour intonatif.

- **effet de l'interaction entre fonction et phonogène sur le type d'étendue syllabique :**
Nous avons employé, pour chaque type d'étendue, une régression logistique avec l'étendue comme variable dépendante, l'interaction entre fonction et phonogène comme variable fixe, et le locuteur comme variable aléatoire. Nous présentons ci-dessous (cf Tableau 30) les résultats des régressions. On voit que l'effet n'est significatif pour aucun type d'étendue.

	résultats de la régression logistique
dernière syllabe	$\chi^2(4) = 3,7, p > 0,1$
deux dernières syllabes	$\chi^2(4) = 1,0, p > 0,1$
constituant entier	$\chi^2(4) = 9,1, p > 0,05$
première syllabe	$\chi^2(4) = 3,0, p > 0,1$

Tableau 30. Résultats des régressions logistiques concernant l'effet de l'interaction entre fonction et phonogène sur le pourcentage d'occurrences des différents types d'étendue syllabique du contour intonatif.

4.3.6. Accentuation initiale

Les effets suivants ont été évalués :

- **effet de la fonction sur la présence d'un accent initial sur le constituant focalisé :** On peut voir ci-dessous (cf Figure 45) le pourcentage d'occurrence d'accentuation initiale pour

chaque fonction. Comme pour l'étendue syllabique du contour intonatif (cf § 4.3.4), seuls les constituants focalisés comportant plusieurs syllabes sont concernés. On voit que l'insistance présente, numériquement, plus d'accentuation initiale que les deux autres fonctions. Afin de déterminer la significativité de l'effet, nous avons employé une régression logistique. La variable dépendante est la présence d'accent initial, la variable fixe est la fonction, et la variable aléatoire est le locuteur. L'effet est significatif ($\chi^2(2) = 21,8, p < 0,001$). Nous présentons ci-dessous (cf Tableau 31) les résultats de la régression. On voit que les différences entre l'insistance et les deux autres fonctions sont significatives. La différence entre l'expressivité et le marquage de focus est en revanche non-significative.

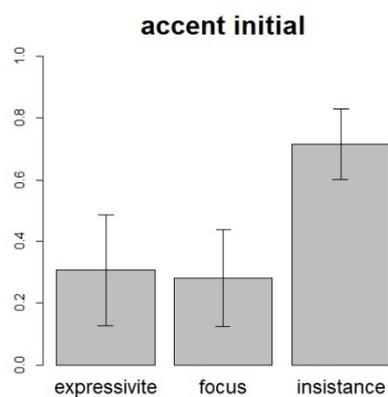


Figure 45. Pourcentage d'occurrence de la présence d'accent initial sur le constituant focalisé pour chaque fonction. Les barres d'erreur indiquent les intervalles de confiance à 95%.

	estimation (et erreur standard)	score-z	valeur-p
expressivité/ insistance	-1,74 (0,51)	-3,393	0,000691
expressivité/ marquage de focus	0,13 (0,58)	0,220	0,825901
insistance/ marquage de focus	1,87 (0,49)	3,836	0,000125

Tableau 31. Résultats de la régression logistique concernant l'effet de la fonction la présence d'un accent initial sur les constituants focalisés du corpus.

- **effet du phonogène sur les différences de présence d'accent initial entre les fonctions :** Nous avons employé une régression logistique avec la présence d'accent initial comme variable dépendante, la fonction et le phonogène comme variables fixes, et le locuteur comme variable aléatoire. L'effet n'est pas significatif ($\chi^2(2) = 3,5$, $p > 0,1$).
- **effet de l'interaction entre fonction et phonogène sur la présence d'accent initial :** Nous avons employé une régression logistique avec la présence d'accent initial comme variable dépendante, l'interaction entre fonction et phonogène comme variable fixe, et le locuteur comme variable aléatoire. L'effet n'est pas significatif ($\chi^2(4) = 8,0$, $p > 0,05$).

4.3.7. Résultats de l'étude sur corpus réduit

Comme pour l'étude précédente (cf chap. 2, § 5.4), nous avons répété cette étude sur un corpus réduit constitué des enregistrements ayant obtenu les meilleurs taux d'accord concernant le relevé des occurrences de focalisation. Pour rappel, les enregistrements retenus sont les deux productions spontanées, ainsi que deux productions interprétées. Par conséquent, le corpus réduit ne contient pas d'enregistrements de parole lue. Nous présentons ci-dessous les résultats qui diffèrent de manière importante des résultats obtenus sur le corpus entier (les résultats similaires aux résultats précédents sont disponibles dans l'annexe B) :

- **taux d'accord entre les experts :** On voit ci-dessous (cf Tableau 32) que tous les kappas sont significatifs, sauf ceux du marquage de focus informationnel large, du marquage de

focus contrastif et du marquage de focus associatif (cf note 45). Par rapport au corpus entier, on voit que le taux d'accord augmente pour la fonction d'emphase (0,314 contre 0,213) ainsi que ses deux sous-fonctions, l'insistance (0,228 contre 0,166) et l'expressivité (0,243 contre 0,183). Les taux d'accord pour ces fonctions restent toutefois relativement bas par rapport aux taux d'accord observés dans d'autres études sur la prosodie (cf § 4.3.1). Pour les autres fonctions, le taux d'accord reste similaire.

	marquage de focus informationnel large	marquage de focus informationnel étroit	marquage de focus contrastif	marquage de focus associatif	marquage de verum focus
	-0,003 (z = -0,4)	0,050 (z = 6,85)	0,010 (z = 1,4)	-0,002 (z = -0,3)	0,083 (z = 11,4)
taux d'accord (kappa de Fleiss)	marquage de focus (fonction globale)	insistance	expressivité	emphase (fonction globale)	autre fonction
	0,125 (z = 17,1)	0,228 (z = 31,4)	0,243 (z = 33,5)	0,314 (z = 43,2)	0,038 (z = 5,2)

Tableau 32. Taux d'accord entre les experts en prosodie concernant la classification fonctionnelle des occurrences de focalisation dans le corpus réduit. Les scores-z sont indiqués sous les kappas. Tous les scores-z sont significatifs à $p < 0,05$, sauf ceux du marquage de focus informationnel large ($p = 0,702$), du marquage de focus contrastif ($p = 0,17$) et du marquage de focus associatif ($p = 0,743$).

- **durée syllabique** : Nous avons observé, dans le corpus entier (cf § 4.3.3), une différence significative entre le marquage de focus et les deux autres fonctions concernant ce paramètre. Nous avons également vu que les différences entre fonctions sont influencées par le phonogène. Dans le corpus réduit, d'après des modèles linéaires mixtes similaires aux précédents, tous les effets évalués s'avèrent être non-significatifs :
 - effet de la fonction sur la durée des syllabes focalisées : $\chi^2(2) = 2,8$, $p > 0,1$

- effet du phonogène sur les différences de durée entre les fonctions : $\chi^2(1) = 2,9$, $p > 0,05$
 - effet de l'interaction entre fonction et phonogène sur la durée : $\chi^2(2) = 4,7$, $p > 0,05$
- **contour intonatif** : Nous avons observé, dans le corpus entier (cf § 4.3.3), des différences significatives entre les fonctions concernant deux contours (Hi et H*H%), ainsi que concernant les trois types de frontière prosodique (groupe accentuel, intonatif ou intermédiaire). Dans le corpus réduit, d'après des tests statistiques similaires aux précédents (tests du khi carré de Pearson et régressions logistiques), les résultats suivants sont observés :
- Contrairement à ce qui est observé dans le corpus entier, l'effet de la fonction sur le type de contour intonatif n'est pas significatif ($\chi^2(16) = 20,5$, $p > 0,1$).
 - Comme pour le corpus entier, l'effet de la fonction sur le type de frontière prosodique est significatif ($\chi^2(4) = 14,2$, $p < 0,01$). On observe des résultats similaires concernant le groupe accentuel et le groupe intonatif. Cependant, contrairement à précédemment, l'effet n'est pas significatif concernant le groupe intermédiaire ($\chi^2(2) = 0,004$, $p > 0,1$).

4.4. Discussion

4.4.1. Classification des occurrences de focalisation

Avant de reprendre les questions de recherche posées au début de ce chapitre (cf § 1.1), on peut tout d'abord s'interroger sur les taux d'accord qui ont été mesurés entre les experts en prosodie concernant la classification fonctionnelle des occurrences de focalisation prosodique. Comme nous l'avons vu (cf § 4.3.1), le taux d'accord est globalement bas pour toutes les fonctions. Comme dans l'étude précédente (cf chap. 2, § 6.1), le manque de concordance entre les experts réduit la portée des résultats de l'analyse prosodique. Seules trois fonctions ont pu être prises en compte dans cette dernière : l'insistance, l'expressivité, et le marquage de focus (en tant que fonction globale). Il s'agit en effet (avec l'emphase en tant que fonction globale) des seules fonctions pour lesquelles le taux d'accord n'est pas

quasiment nul. On observe par ailleurs que le taux d'accord de l'emphase (0,213) est plus élevé que celui du marquage de focus (0,121). Cette différence est encore plus prononcée dans le corpus réduit, où le taux d'accord augmente pour l'emphase (0,314). Cela signifie qu'il est plus facile pour les experts de relever les occurrences d'emphase que les occurrences de marquage de focus. Il est particulièrement difficile pour les experts de relever les occurrences de marquage des différentes catégories de focus. En effet, il s'agit des fonctions (avec les fonctions annotées « autres ») qui ont obtenu les taux d'accord les plus bas.

La faiblesse globale du taux d'accord soulève plusieurs questions méthodologiques et théoriques :

- Le résultat est inattendu, étant donné que nous avons demandé aux experts de se référer à la même définition théorique pour chaque fonction (cf § 4.2.2). Ces définitions sont explicites, et sont inspirées de la littérature la plus récente concernant chaque fonction. Cependant, il est possible qu'elles n'aient pas été bien assimilées par les experts. Notons que certaines notions (notamment les différentes catégories de focus) sont relativement complexes. La méthodologie de cette étude présente un risque de ce point de vue. Il est également possible que, malgré nos consignes, les experts se soient référés à d'autres définitions présentes dans la littérature. Or, nous avons vu (cf § 2.1 et § 3.1) que cette dernière est souvent confuse et contradictoire concernant la plupart des fonctions.
- Comme nous l'avons évoqué dans la discussion de l'étude précédente (cf chap. 2, § 6.1), il existe une disparité dans la compétence des experts. Cette disparité pourrait expliquer que les taux d'accord ne soient pas plus élevés. Cependant, nous avons vu (cf § 4.3.7) que les taux d'accord restent globalement bas dans l'étude menée sur le corpus réduit. Pour rappel, ce corpus réduit est constitué des enregistrements ayant obtenu les meilleurs taux d'accord concernant le relevé des occurrences de focalisation.
- Il est possible que le corpus ne présente pas suffisamment d'occurrences de certaines fonctions (ce qui expliquerait le faible taux d'accord concernant ces fonctions). Cela pourrait être dû au protocole d'élicitation que nous avons employé. Pour rappel (cf § 4.2.1), il s'agit du protocole RepTask (Laurens et al 2011), qui consiste à enregistrer une conversation spontanée, puis à demander à des locuteurs de reproduire cette

conversation. Le taux d'accord serait sûrement plus élevé avec un corpus de parole de laboratoire constitué d'énoncés indépendants (comme c'est le cas de nombreuses études sur la réalisation prosodique des fonctions de la focalisation, cf § 2.2.5 et § 3.2.3). Nous n'avons pas choisi un tel corpus, puisque nous souhaitons déterminer l'influence du phonogène (parole spontanée, lue ou interprétée) sur la réalisation des fonctions.

- D'un point de vue théorique, on peut enfin se demander s'il existe vraiment des distinctions entre les différentes fonctions attribuées à la focalisation par la littérature. Bien que ces distinctions soient valides d'un point de vue sémantique ou pragmatique, il est possible qu'elles ne se reflètent pas au niveau phonologique ou phonétique. Cette possibilité rejoint notamment les travaux de Marandin (2005) et de Trotzke (2017), qui considèrent qu'il existe une interaction entre la structure informationnelle (cf § 2.1.1) et la dimension expressive du langage (cf § 3.1). Cela signifierait que les différentes fonctions attribuées à la focalisation sont en réalité des sous-catégories d'une même fonction globale.
- La différence de taux d'accord entre l'emphase et le marquage de focus pourrait, quant à elle, s'expliquer par les deux facteurs suivants :
 - Comme nous l'avons vu (cf § 4.3.2 et § 4.3.3), ces deux ensembles de fonctions présentent des différences phonétiques. En effet, l'insistance et l'expressivité présentent une durée syllabique significativement supérieure à celle du marquage de focus. Une différence numérique de hauteur est également observée, bien qu'elle ne soit pas significative. Il est possible que, pour les experts, l'augmentation de hauteur et de durée soit particulièrement associée à l'emphase. Les experts auraient alors attribué des fonctions emphatiques à certaines occurrences de focalisation parce qu'elles présentent une telle augmentation, et non parce qu'elles correspondent aux définitions théoriques que nous avons données à ces fonctions.
 - La différence pourrait également être due au contenu sémantique des constituants focalisés. En effet, 50,4% des constituants focalisés du corpus consistent en des adjectifs ou adverbes intensifs (ex : *tellement*, *hyper*, *immonde*, *passionnant*), ou bien

des mots expressifs (ex : *putain*). Il est possible que ce type de contenu intensif ou expressif entraîne plus facilement l'attribution d'une fonction emphatique de la part des experts. On observe effectivement une différence de fréquence d'occurrence de ce type de contenu entre le marquage de focus (13,5%) et l'emphase (69,5%). Afin de tester la significativité de cette différence, nous avons employé une régression logistique avec la présence de contenu intensif ou expressif comme variable dépendante, la fonction comme variable fixe, et le locuteur comme variable aléatoire. L'effet est marginal ($\chi^2(1) = 3,3$, $p = 0,069$). La différence entre les deux fonctions approche la significativité ($z = 1,78$, $p = 0,075$).

4.4.2. Réalisation prosodique des fonctions

Avant de répondre à la première question de recherche, notons tout d'abord qu'on observe des différences de fréquence d'occurrence dans le corpus entre les trois fonctions analysées (cf Tableau 21). Nous discutons ces différences ci-dessous :

- Parmi les occurrences de focalisation retenues pour l'analyse prosodique, l'emphase (72,5%) est beaucoup plus représentée que le marquage de focus (27,5%). Cette différence est logique, puisque nous avons vu (cf § 4.3.1) que l'emphase présente un taux d'accord supérieur au marquage de focus, en particulier dans le corpus réduit (cf § 4.3.7). Comme nous l'avons expliqué ci-dessus (cf § 4.4.1), deux facteurs pourraient expliquer cette différence de taux d'accord : les différences phonétiques entre ces deux ensembles de fonctions, ainsi qu'une différence dans le contenu sémantique des constituants concernés. Cependant, la différence de fréquence d'occurrence entre l'emphase et le marquage de focus pourrait également être due à la présence de parole interprétée dans le corpus. En effet, nous avons vu qu'une majorité des occurrences d'emphase se trouvent dans ce phonogène. Nous discutons ce résultat dans la section suivante (cf § 4.4.3).
- Parmi les occurrences d'emphase, l'insistance (68,0%) est beaucoup plus représentée que l'expressivité (32,0%). Cette différence pourrait s'expliquer par la forte association qui est observée dans le corpus entre l'insistance et la présence d'accent initial (cf § 4.3.6). En effet, nous avons vu (cf § 3.2.1) que les études précédentes sur l'insistance considèrent que l'accent initial est un trait nécessaire à la réalisation de cette fonction. Il est possible

que les experts, se basant sur ces études plutôt que sur notre définition théorique (cf § 3.1.1), aient systématiquement attribué la fonction d'insistance aux constituants focalisés comportant ce trait. Cela expliquerait la forte fréquence d'occurrence de cette fonction parmi les constituants focalisés.

La première question de recherche de cette analyse concerne les différences de réalisation entre les fonctions attribuées à la focalisation. Au niveau phonétique, nous avons analysé les différences de hauteur et de durée syllabique entre les trois fonctions (cf § 4.2.3). Nous avons observé (cf Figure 40) que l'insistance et l'expressivité présentent une durée syllabique significativement supérieure à celle du marquage de focus. Cependant, cette différence n'est pas répétée dans l'étude menée sur le corpus réduit (cf § 4.3.7). L'insistance et l'expressivité présentent également une hauteur supérieure à celle du marquage de focus (cf Figure 38), bien que la différence ne soit pas significative. Ces résultats suggèrent néanmoins qu'il existe bien une distinction phonétique entre les fonctions emphatiques et les fonctions sémantico-pragmatiques de la focalisation. Concernant l'expressivité, la différence avec le marquage de focus est cohérente avec la littérature. En effet, nous avons vu (cf § 3.2.2) que celle-ci attribue une forte augmentation de hauteur et de durée à l'expressivité.

Nous avons également évalué l'effet de la fonction sur trois catégories phonologiques des constituants focalisés (cf § 4.2.3). Il s'agit du type de contour intonatif, de l'étendue syllabique du contour intonatif, et de la présence d'un accent initial. Nous discutons ci-dessous les résultats observés :

- L'insistance présente significativement plus de contours Hi que l'expressivité et que le marquage de focus (cf Figure 42). Pour rappel (cf chap. 2, § 4.3.3), le contour Hi est un ton haut initial sur le constituant focalisé, lorsque celui-ci se trouve au début d'un groupe accentuel. Cependant, cette différence ne s'observe pas dans le corpus réduit. L'insistance se distingue également des deux autres fonctions par une fréquence significativement plus élevée d'accentuation initiale⁵⁶ (cf Figure 45). Notre étude soutient donc l'existence de la

⁵⁶ Pour rappel (cf chap. 2, note 19), l'accentuation initiale et la présence d'un contour Hi sont deux traits distincts. En effet, un accent initial peut tomber sur un constituant portant un autre contour que Hi.

relation, souvent décrite dans la littérature (cf § 3.2.1), entre insistance et accentuation initiale.

- Le contour H*H% est significativement plus fréquent pour l'expressivité que pour l'insistance (cf Figure 42). Cependant, cette différence n'est pas répétée dans le corpus réduit. Le résultat est néanmoins cohérent avec la forte augmentation de hauteur que la littérature attribue à l'expressivité. En effet, le contour H*H% est un contour montant présentant une forte amplitude, ce qui implique une forte augmentation de hauteur. De plus, nous avons vu dans le premier chapitre (cf § 4.2) qu'une des fonctions possibles du contour H*H% est l'indication de la modalité exclamative de l'énoncé. Cette modalité est cohérente avec la fonction d'expressivité.

- On observe des différences entre les fonctions concernant le type de frontière prosodique sur les constituants focalisés (cf Figure 43) :
 - Les frontières de groupe accentuel sont significativement plus fréquentes pour l'insistance que pour les deux autres fonctions. Inversement, les frontières de groupe intonatif sont significativement moins fréquentes pour l'insistance. Ces résultats sont cohérents avec la littérature sur les différentes fonctions de la focalisation. En effet, nous avons vu (cf § 3.1.1) que l'insistance consiste à souligner l'importance d'un élément. Rossi (1999 : 116-129) parle de « proéminence sémantique », et Di Cristo (1999b) de « focalisation emphatique étroite ». Il est donc logique que cette fonction soit réalisée au niveau du groupe accentuel, qui est le niveau le plus bas dans la structure prosodique (cf chap. 1, § 3). Par contraste, le marquage de focus et l'expressivité peuvent affecter l'énoncé entier. C'est notamment le cas du marquage de focus informationnel large, qui désigne l'assertion par le locuteur de la totalité du contenu de l'énoncé (cf § 2.1.1). Nous avons également vu (cf § 3.1.2) que Di Cristo (1999b) définit l'expressivité comme une « emphase de cumul » portant sur l'énoncé entier. Il est donc logique que le marquage de focus et l'expressivité soient plus souvent réalisés au niveau du groupe intonatif, qui est le niveau le plus haut de la structure prosodique.

- Les frontières de groupe intermédiaire sont significativement plus fréquentes pour le marquage de focus que pour l'expressivité. De manière presque significative, elles sont également plus fréquentes pour le marquage de focus que pour l'insistance. Cependant, aucune de ces différences n'est observée dans le corpus réduit. Nous avons analysé les contours intonatifs concernés : le contour H*H- est majoritaire (75,0%). On observe effectivement une différence numérique (mais non-significative) entre le marquage de focus et les deux autres fonctions concernant ce contour (cf Figure 42). Pour rappel (cf chap. 2, § 4.3.3), le contour H*H- est un contour montant dont l'amplitude et l'allongement se situent entre ceux du contour H* et du contour H*H%. Nous avons vu (cf chap. 2, Figure 35) que ce contour sert généralement à exprimer une continuation dans notre corpus.
- On n'observe aucune différence entre les fonctions concernant l'étendue syllabique du contour intonatif (cf Figure 44). Ce résultat n'est pas surprenant, car nous avons vu dans l'étude précédente (cf chap. 2, Figure 30) qu'une majorité (72,7%) des contours sont portés par la dernière syllabe du constituant focalisé dans notre corpus.

4.4.3. Influence du phonogène sur la réalisation

Avant de répondre à la seconde question de recherche, nous reprenons tout d'abord les différences de fréquence d'occurrence concernant chaque fonction selon le phonogène (cf Tableau 21) :

- La majorité (68,8%) des occurrences d'emphase se trouvent en parole interprétée⁵⁷. Ce résultat est cohérent avec l'intuition courante selon laquelle la parole interprétée se caractérise par une forte expressivité. Dans l'étude précédente (cf chap. 2 § 5.2), nous avons observé que la parole interprétée contient une fréquence d'occurrence de focalisation plus élevée que les deux autres phonogènes, ce qui va également dans le sens

⁵⁷ Nous avons employé une régression logistique afin de tester la significativité de l'effet du phonogène sur la présence d'emphase. La variable dépendante est la présence d'emphase sur la syllabe focalisée, la variable fixe est le phonogène, et la variable aléatoire est le locuteur. L'effet est significatif ($\chi^2(2) = 88,1$, $p < 0,001$). Les occurrences d'emphase sont significativement plus nombreuses en parole interprétée qu'en parole lue ($z = 8,004$, $p < 0,001$) et qu'en parole spontanée ($z = 2,885$, $p < 0,01$). La différence entre la parole lue et la parole spontanée n'est pas significative ($z = 0,340$, $p > 0,1$).

de cette intuition. De plus, le fait que la parole interprétée contienne la majorité des occurrences d'emphase suggère qu'il s'agit d'un phonogène favorable à l'étude de cet ensemble de fonctions.

- La majorité (74,5%) des occurrences de marquage de focus se trouvent en parole lue. Cependant, la différence est significative par rapport à la parole interprétée uniquement⁵⁸. Contrairement au résultat précédent, ce résultat suggère que la parole interprétée est un phonogène moins favorable à l'étude du marquage de focus que la parole lue.

La seconde question de recherche de cette analyse concerne l'influence du phonogène sur la réalisation prosodique des fonctions attribuées à la focalisation. Les différences observées concernent la hauteur (cf Figure 39) et la durée syllabique (cf Figure 41) :

- Concernant la hauteur, des différences numériques sont observées entre l'insistance et l'expressivité. En parole lue et interprétée, l'insistance présente la hauteur la plus élevée. En parole spontanée, c'est au contraire l'expressivité qui présente la hauteur la plus élevée.
- Concernant la durée syllabique, des différences significatives sont observées entre l'expressivité et les deux autres fonctions :
 - En parole lue, l'expressivité présente une durée syllabique plus élevée que l'insistance. En parole spontanée, c'est au contraire l'insistance qui présente une durée plus élevée.
 - En parole lue, la supériorité de durée syllabique de l'expressivité par rapport au marquage de focus est plus importante qu'en parole spontanée et interprétée.

⁵⁸ Nous avons employé une régression logistique similaire à la précédente (cf note 57). L'effet est significatif ($\chi^2(2) = 88,1, p < 0,001$). Les occurrences de marquage de focus sont significativement plus nombreuses en parole lue qu'en parole interprétée ($z = 8,004, p < 0,001$). Elles sont aussi significativement plus nombreuses en parole spontanée qu'en parole interprétée ($z = 2,885, p < 0,01$). En revanche, la différence entre la parole lue et la parole spontanée n'est pas significative ($z = 0,340, p > 0,1$).

Cependant, ces résultats sont difficilement interprétables. En effet, les différences de hauteur entre l'insistance et l'expressivité ne sont pas significatives. De plus, aucune des différences de durée syllabique n'est répétée dans le corpus réduit. Enfin, aussi bien pour la hauteur que pour la durée, nous avons vu (cf § 4.3.1) que le corpus comprend très peu d'occurrences pour deux des interactions concernées (insistance en parole spontanée, et expressivité en parole lue).

4.5. Conclusions

Cette première étude ne nous permet pas de répondre entièrement aux deux questions de recherche initiales de cette analyse (cf § 1.1). En effet, nous avons vu (cf § 4.2.2) que le protocole de l'étude repose sur une classification fonctionnelle des occurrences de focalisation par un groupe d'experts en prosodie. Or, le taux d'accord entre les experts concernant cette classification s'avère être globalement bas. Seules trois fonctions présentent un taux d'accord suffisamment haut pour pouvoir être analysées : le marquage de focus (en tant que fonction globale), l'insistance, et l'expressivité. Nous n'avons pas pu analyser la réalisation du marquage des différentes catégories de focus (cf § 2.1). La faiblesse du taux d'accord soulève la question de la validité de la distinction entre les fonctions attribuées à la focalisation. Pourtant, l'étude révèle certaines différences prosodiques entre les fonctions analysées :

- L'emphase (composée de l'insistance et de l'expressivité) se distingue du marquage de focus par une durée syllabique supérieure. L'emphase présente aussi (bien que de manière non-significative) une hauteur supérieure au marquage de focus.
- L'insistance se distingue des deux autres fonctions par une forte fréquence d'occurrence d'accent initial. Cette relation a souvent été rapportée dans la littérature (cf § 3.2.1).
- L'expressivité se distingue des deux autres fonctions par une plus forte fréquence d'occurrence du contour H*H%. Il s'agit d'un contour montant de fin de groupe intonatif, que l'on observe notamment sur les énoncés exclamatifs (cf chap. 1, § 4.2).

- Les fonctions ne sont pas réalisées au même niveau de la structure prosodique (cf chap. 1, § 3). L'insistance est plus souvent réalisée au niveau du groupe accentuel que les deux autres fonctions, et moins souvent au niveau du groupe intonatif. Le marquage de focus est quant à lui plus souvent réalisé au niveau du groupe intermédiaire.

En revanche, l'étude ne révèle aucune influence du phonogène (parole spontanée, lue ou interprétée) sur la réalisation des fonctions de la focalisation. Une influence est cependant observée concernant la fréquence d'occurrence de ces fonctions. En effet, une majorité des occurrences d'emphase se trouve en parole interprétée, et une majorité des occurrences de marquage de focus se trouve en parole lue.

Cette étude pourrait bénéficier des mêmes améliorations que l'étude présentée dans le chapitre précédent. En effet, elle porte sur le même corpus et se base sur la même analyse prosodique. Nous rappelons ces améliorations ci-dessous (cf chap. 2, § 7) :

- Le corpus comprend moins d'enregistrements de parole spontanée que de parole lue et interprétée. L'expérience pourrait être reconduite avec un corpus plus équilibré. De plus, les locuteurs de parole lue et de parole interprétée possèdent tous une expérience de l'art dramatique (ce qui a pu introduire dans leur production des traits prosodique relevant d'un phonostyle professionnel). L'expérience pourrait être reconduite avec des locuteurs lambda.
- L'annotation du corpus par les experts en prosodie pourrait être améliorée sur plusieurs points. Il existe tout d'abord une disparité dans la compétence des experts. De plus, pour des raisons pratiques, les enregistrements n'ont pas tous été annotés par les mêmes experts. Il serait enfin possible d'effectuer l'annotation en deux temps, afin de sélectionner les meilleurs experts, ou bien de vérifier leur compréhension des consignes.
- Certains traits n'ont pas été pris en compte dans l'analyse prosodique. Au niveau phonétique, il s'agit des paramètres d'intensité et de qualité vocalique, et des mesures de configuration tonale et de constituance syllabique des accents de focalisation (cf Astésano 2001). Au niveau phonologique, nous n'avons pas pris en compte le trait de désaccentuation post-focale ou pré-focale (souvent attribué à la focalisation).

5. Etude de perception

5.1. Introduction

Cette seconde étude a pour but de répondre à la première question de recherche de cette analyse (cf § 1.1). Nous rappelons cette question ci-dessous :

Question 1 : Les fonctions attribuées à la focalisation présentent-elles des réalisations différentes ?

Contrairement à l'étude précédente, cette étude ne prend pas en compte l'influence du phonogène sur la réalisation des fonctions (cf § 4.4.3). Le but de l'étude est de vérifier, du point de vue de la perception, certains des résultats qui ont été observés dans l'étude précédente. Pour des raisons méthodologiques, l'étude porte uniquement sur la réalisation phonétique des fonctions⁵⁹ (hauteur et durée syllabique, cf § 4.2.3). Le protocole consiste à demander à des auditeurs de porter des jugements d'acceptabilité sur des énoncés contenant une occurrence de focalisation. Ces énoncés sont extraits du corpus de l'expérience de production (cf § 4.2.1). Nous avons manipulé leur prosodie sur le logiciel Praat. La manipulation a consisté à obtenir quatre réalisations phonétiques différentes pour chaque énoncé. Ces réalisations varient selon la hauteur et la durée de la syllabe initiale du constituant focalisé⁶⁰ :

- **réalisation 1** : augmentation de hauteur sur la syllabe initiale
- **réalisation 2** : augmentation de durée sur la syllabe initiale
- **réalisation 3** : augmentation de hauteur et de durée sur la syllabe initiale
- **réalisation 4** : pas de manipulation prosodique sur la syllabe initiale

Les énoncés ont été présentés visuellement aux participants. Nous leur avons également présenté, de manière à la fois visuelle et auditive, le contexte d'origine de chaque énoncé

⁵⁹ La prise en compte du niveau phonologique (contour intonatif, étendue du contour et accentuation initiale, cf § 4.2.3) aurait nécessité de conduire des expériences supplémentaires. En effet, le protocole employé dans cette expérience nécessite de limiter la variation prosodique entre les stimuli présentés aux participants.

⁶⁰ Cette restriction de la manipulation à la syllabe initiale s'explique elle aussi par la nécessité de limiter la variation prosodique entre les stimuli de l'expérience (cf note 59).

(c'est-à-dire le ou les énoncés qui précèdent cet énoncé dans le corpus). Les participants ont ensuite entendu les quatre réalisations phonétiques de l'énoncé. La tâche des participants a consisté à évaluer l'acceptabilité de chaque réalisation sur une échelle de zéro à cinq. Nous avons demandé aux participants d'évaluer cette acceptabilité en fonction du contexte de l'énoncé. Nous considérons que le contexte contribue à indiquer aux participants la fonction de l'occurrence de focalisation (parmi d'autres facteurs tels que le contenu de l'énoncé et la situation de communication). En demandant aux participants d'évaluer la relation entre la réalisation et le contexte, on obtient indirectement un jugement sur la relation entre cette réalisation et la fonction de l'occurrence de focalisation. Cette fonction est déterminée au préalable sur la base de la classification qui a été effectuée par un groupe d'experts dans l'étude précédente (cf § 4.2.2). Comme dans l'étude précédente, seules les fonctions d'insistance, d'expressivité et de marquage de focus (en tant que fonction globale) sont analysées. Pour rappel (cf § 4.3.1), ceci est dû au fait qu'il s'agit des seules fonctions pour lesquelles le taux d'accord entre les experts n'est pas quasi-nul.

Dans l'étude précédente, plusieurs différences phonétiques ont été observées entre les trois fonctions (cf § 4.3.2 et § 4.3.3). L'insistance et l'expressivité présentent une durée syllabique significativement plus élevée que celle du marquage de focus. De manière non-significative, l'expressivité présente également une durée supérieure à celle de l'insistance. Enfin, de manière non-significative, l'insistance présente la hauteur la plus élevée, suivie de l'expressivité et du marquage de focus. Il est donc possible de formuler des prédictions concernant la question de recherche de cette étude de perception. Nous nous attendons à observer les distributions suivantes pour chaque fonction :

- **insistance** : réalisations 1 et 3 plus fréquentes
- **expressivité** : réalisations 2 et 3 plus fréquentes
- **marquage de focus** : réalisation 4 plus fréquente

5.2. Méthodologie

5.2.1. Participants

L'expérience a été menée avec trente participants (vingt-deux femmes et huit hommes). Les participants possèdent tous le français pour langue maternelle. Ils sont âgés de dix-huit à

cinquante-huit ans (la moyenne d'âge est de vingt-neuf ans). Ils ont été recrutés sur le site internet du RISC (www.risc.cnrs.fr). Ils ont reçu une rémunération de dix euros pour leur participation à cette expérience, ainsi qu'à une autre expérience non-reliée à cette étude (et ne portant pas sur la prosodie).

5.2.2. Stimuli

Nous avons sélectionné, dans le corpus de l'expérience de production (cf § 4.2.1), trente énoncés comportant une occurrence de focalisation prosodique. Nous avons choisi des énoncés relativement courts, afin que les participants de l'expérience n'aient pas une quantité trop importante d'informations prosodiques à traiter. Nous avons choisi dix énoncés pour chacune des trois fonctions analysées (insistance, expressivité et marquage de focus). Comme nous l'avons expliqué ci-dessus (cf § 5.1), nous avons également présenté aux participants le contexte d'origine de chaque énoncé. Nous présentons ci-dessous, pour chaque fonction, un exemple d'énoncé précédé de son contexte (l'énoncé est souligné et le constituant focalisé est indiqué en majuscules) :

- **insistance** : « Bah, en plus on lui avait prévu pour son séminaire, on lui avait prévu une – séminaire de M1 – une toute petite salle, ce qui était étonnant parce qu'y avait... énormément d'étudiants, à Paris 4 en philo, et la salle était toute petite ! Et elle était BLINDEE à craquer. »
- **expressivité** : « Et le premier cours, il commente la bibliographie, et y a une étudiante qui prenait des notes ou qui soulignait, et il lui a hurlé dessus, il lui a arraché la bibliographie, il l'a déchirée en mille morceaux en disant : ECOUTEZ ce que je dis ! »
- **marquage de focus** : « A propos de mon nom, du coup j'ai une anecdote. Quand j'étais... Je devais être... Ouais, j'étais au LYCEE. »

Nous avons réenregistré chaque énoncé afin d'en obtenir une version à la prosodie relativement peu marquée⁶¹. Cette étape a pour but d'éviter que les participants soient influencés par la prosodie d'origine des stimuli. Une telle influence aurait en effet pu avoir lieu, étant donné que la manipulation prosodique ne porte que sur la syllabe initiale du constituant focalisé. Nous avons prononcé nous-même les énoncés, en employant un débit relativement lent et un registre tonal relativement étroit. Nous n'avons pas produit d'accents autres que rythmiques (ayant pour fonction de délimiter les groupes prosodiques de l'énoncé, cf chap. 1, § 2.1). Nous avons contrôlé la qualité des énoncés réenregistrés en visualisant leurs informations acoustiques sur le logiciel Praat (cf Boersma 2002). Nous avons également réenregistré les contextes d'origine des énoncés, afin de ne pas créer de disparité avec les énoncés lors de l'expérience. Nous présentons ci-dessous (cf Figure 46), à titre comparatif, la courbe de fréquence fondamentale de la version originale et de la version réenregistrée d'un énoncé.

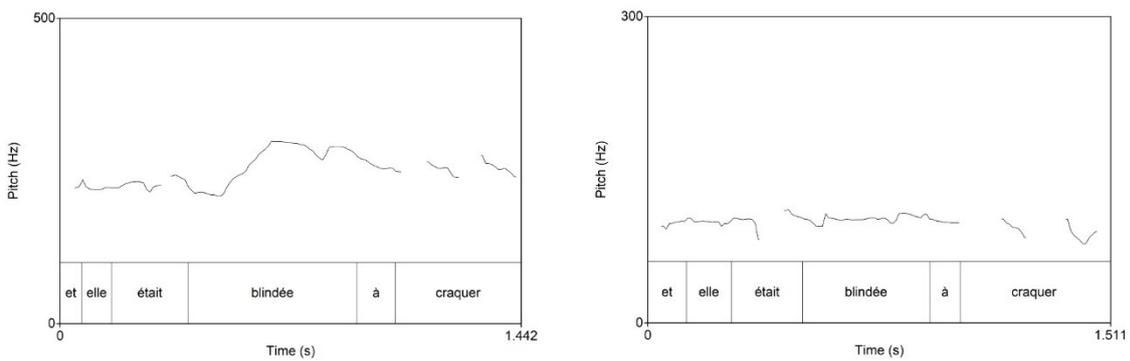


Figure 46. Courbes de fréquence fondamentale de la version originale (à gauche) et de la version réenregistrée (à droite) d'un des énoncés utilisés dans l'expérience de perception. Dans la version originale, une focalisation a lieu sur le mot blindée.

Nous avons ensuite manipulé prosodiquement les énoncés, afin d'obtenir les quatre réalisations phonétiques souhaitées (cf § 5.1). Nous détaillons ci-dessous les étapes de la manipulation :

⁶¹ L'enregistrement a été réalisé dans une chambre sourde à l'Université Paris Diderot, au moyen d'un micro de studio Rode NT1-A, d'une interface audio Roland Quad-Capture et du logiciel Audacity.

- **détermination des seuils d’augmentation de hauteur et de durée** : Il est nécessaire, pour manipuler la hauteur et la durée de la syllabe initiale du constituant focalisé, de disposer de seuils acoustiques à respecter. Pour déterminer les seuils, nous nous sommes basé sur deux sources. La première source est l’étude d’Astésano (2001), qui comprend une analyse de la réalisation des différents types d’accents en français. Concernant la hauteur, l’étude ne mentionne pas explicitement de seuil. Concernant la durée, Astésano rapporte que l’accent initial est réalisé par une augmentation moyenne de 42,5% par rapport aux syllabes non-accentuées⁶². La seconde source est le corpus de notre expérience de production (cf § 4.2.1). Nous avons comparé la hauteur et la durée des syllabes initiales accentuées et des syllabes non-accentuées du corpus⁶³. Concernant la hauteur, on observe que l’accent initial est réalisé par une augmentation moyenne de 3,89 demi-tons. Nous avons adopté ce seuil pour la manipulation prosodique. Concernant la durée, on observe une augmentation moyenne de 34,0%. Nous avons adopté comme seuil la moyenne entre cette augmentation et celle observée par Astésano (2001), soit un seuil de 38,3%.
- **manipulation de la hauteur** : Nous avons effectué la manipulation sur le logiciel Praat, à l’aide de l’extension Momel-Intsint (cf Hirst 2007). Momel-Intsint permet d’obtenir une stylisation de la courbe de hauteur de l’énoncé⁶⁴. Il est ensuite possible de manipuler la courbe de hauteur stylisée. Nous illustrons ci-dessous (cf Figure 47) le processus de manipulation de la hauteur. On voit que la courbe stylisée est composée de points successifs. En déplaçant un point vers le haut ou vers le bas, on augmente ou on diminue

⁶² Cette moyenne concerne deux types d’accent initial : l’accent initial rythmique et l’accent initial emphatique (cf chap. 1, § 2.1). La moyenne a été réalisée sur neuf locuteurs, dans trois phonogenres différents (interview, parole lue, journal radiophonique).

⁶³ Pour rappel (cf § 4.2.3), la hauteur et la durée des syllabes du corpus ont été normalisées par rapport à plusieurs paramètres. La hauteur a été normalisée par rapport aux locuteurs. La durée a été normalisée par rapport à la structure syllabique, ainsi que par rapport au débit de parole des locuteurs.

⁶⁴ La stylisation emploie la fonction mathématique de « spline quadratique » : « *The quadratic spline function used to model the macro-melodic component is defined by a sequence of target points, (couples <s, Hz>) each pair of which is linked by two monotonic parabolic curves with the spline knot occurring (by default) at the midway point between the two targets. [...] This, in fact, defines the most simple mathematical function for which the curves are both continuous and smooth.* » (Hirst 2007)

la hauteur de ce point. Ce déplacement modifie la portion de courbe qui se trouve entre le point précédent et le point suivant. Pour obtenir les réalisations 1 et 3 (cf § 5.1), nous avons déplacé vers le haut le point le plus proche de la syllabe initiale du constituant focalisé. Afin de déterminer la valeur de fréquence fondamentale à attribuer à ce point, nous avons additionné le seuil préalablement déterminé (3,89 demi-tons) à la valeur initiale du point en Hertz⁶⁵. Une fois la manipulation effectuée, nous avons rajouté des points le long de la courbe manipulée, entre les points obtenus par stylisation⁶⁶. Cette opération est nécessaire afin que l'intonation de l'énoncé paraisse plus naturelle auditivement. Notons que la réalisation 4 (sans manipulation sur la syllabe initiale) a également été obtenue en employant Momel-Intsint⁶⁷. Pour cette réalisation, nous avons simplement stylisé la courbe, sans la manipuler.

- **manipulation de la durée** : Nous avons employé le menu de manipulation de Praat, qui permet de modifier la durée d'une portion du signal. Ce processus est également illustré ci-dessous (cf Figure 47). On voit qu'il est possible d'insérer des points le long d'une ligne horizontale. La manipulation fonctionne de manière similaire à celle de la hauteur. En déplaçant un point vers le haut ou vers le bas, on augmente ou on diminue la durée de la portion qui se trouve entre le point précédent et le point suivant. Pour obtenir les réalisations 2 et 3 (cf § 5.1), nous avons inséré un point au début et à la fin de la syllabe initiale du constituant focalisé. Nous avons ensuite employé une fonction de Praat qui permet d'augmenter la durée de la portion ainsi délimitée selon le pourcentage souhaité. Nous avons indiqué comme pourcentage le seuil préalablement déterminé (38,3%) :

⁶⁵ La conversion des demi-tons en Hertz a été effectuée à l'aide d'un logiciel disponible sur Internet à l'adresse suivante : <http://users.utu.fi/jyrtuoma/speech/semitone.html>.

⁶⁶ Nous avons sélectionné pour cela l'option « Interpolate quadratically » dans le menu de manipulation de Praat.

⁶⁷ Ce choix est dû au fait que la stylisation de la courbe de hauteur donne parfois aux énoncés une qualité vocalique relativement artificielle. Afin de ne pas influencer les participants de l'expérience, il était donc nécessaire que tous les stimuli aient une courbe de hauteur stylisée.

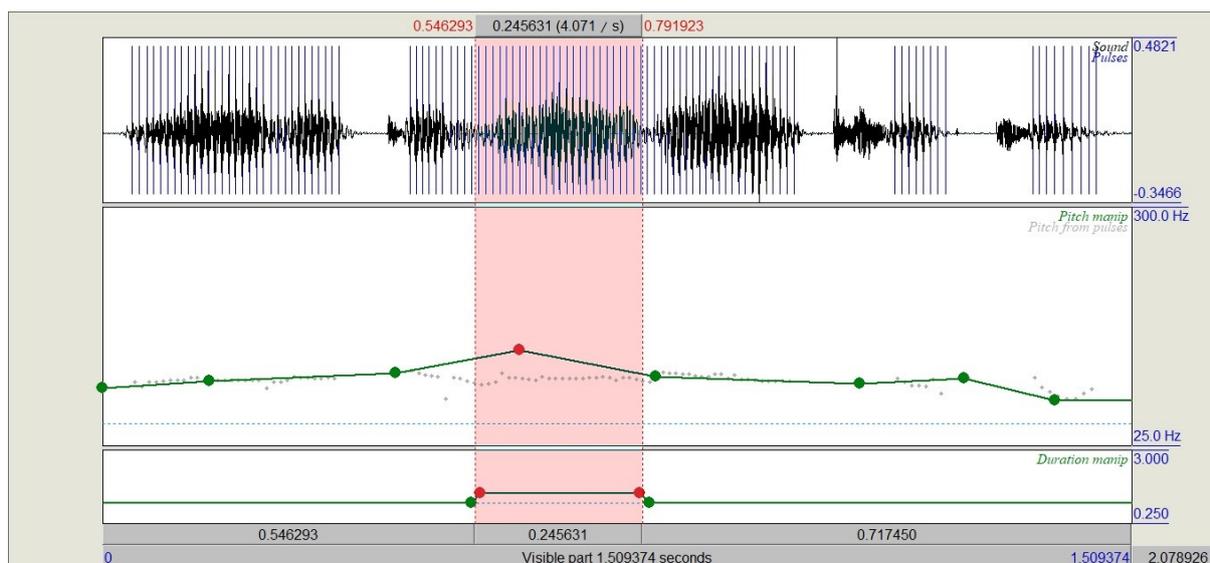


Figure 47. Capture d'écran du menu de manipulation prosodique du logiciel Praat. La fenêtre du haut contient l'oscillogramme de l'énoncé (ainsi que les pulsions laryngées détectées dans le signal). La fenêtre du milieu contient la courbe de hauteur stylisée par Momet-Intsint (en vert), ainsi que la courbe de hauteur originale (en gris). On voit qu'un des points de la courbe stylisée (en rouge) a été déplacé vers le haut, afin d'augmenter la hauteur de la syllabe. La fenêtre du bas contient la ligne de manipulation de la durée. On voit que deux points de la ligne (en rouge) ont été déplacés vers le haut, afin d'augmenter la durée de la syllabe.

Nous avons également préparé cinq items d'entraînement et cinq items de contrôle :

- Les items d'entraînement ont pour but de préparer les participants à la tâche qui leur est demandée. Il s'agit, comme pour les stimuli de l'expérience, d'énoncés sélectionnés dans le corpus de l'expérience de production et contenant une occurrence de focalisation prosodique. Les énoncés sont composés de deux occurrences d'insistance, deux occurrences d'expressivité, et une occurrence de marquage de focus. Nous avons réenregistré les énoncés ainsi que leurs contextes d'origine, puis nous avons manipulé la prosodie des énoncés afin d'obtenir les quatre réalisations phonétiques souhaitées.
- Les items de contrôle ont pour but de vérifier la capacité générale des participants à interpréter les phénomènes de nature prosodique. Il s'agit, comme pour les stimuli et les items d'entraînement, d'énoncés sélectionnés dans le corpus de l'expérience de production et contenant une occurrence de focalisation prosodique. Nous avons réenregistré les énoncés ainsi que leurs contextes d'origine, puis nous avons manipulé la prosodie des énoncés. Nous avons obtenu, pour chaque énoncé, deux réalisations a priori

acceptables⁶⁸, une réalisation neutre⁶⁹, et une réalisation a priori non-acceptable⁷⁰. Selon nos prédictions, le premier type de réalisation doit obtenir une évaluation positive de la part des participants, le second type de réalisation doit obtenir une évaluation positive ou moyenne, et le troisième type de réalisation doit obtenir une évaluation négative. Il est possible, de cette façon, d'exclure de l'analyse les participants qui n'évaluent pas les réalisations de la manière attendue.

5.2.3. Protocole

L'expérience a été présentée aux participants sur un ordinateur à l'aide de l'application E-Studio du logiciel E-Prime (version 2.0). La durée moyenne de l'expérience est de trente minutes. Nous détaillons ci-dessous le déroulement de l'expérience :

- Les consignes ont été présentées aux participants. Nous avons expliqué aux participants qu'ils allaient entendre des énoncés prononcés sur plusieurs « tons » différents. Nous leur avons demandé d'évaluer chaque prononciation en fonction de sa « correspondance » avec le contexte de l'énoncé. Nous avons par ailleurs fait croire aux participants que les réalisations phonétiques avaient été produites par un système de synthèse de la parole,

⁶⁸ Pour trois des cinq items de contrôle, nous avons synthétisé un contour terminal montant ou descendant sur le constituant focalisé, ainsi qu'une désaccentuation post-focale sur la séquence qui suit. Comme nous l'avons vu plus haut (cf § 2.2.1), cette réalisation est souvent associée dans la littérature à la fonction de marquage de focus informationnel étroit. Nous avons sélectionné des énoncés dont le contexte d'origine est propice à la réalisation de cette fonction. Dans l'exemple (i) ci-dessous, le contour terminal porte sur *matches*, et la désaccentuation porte sur *en général* :

- (i) En général, j'avais ma place, tu vois. Je jouais TOUS LES MATCHS, en général.

Pour les deux autres items de contrôle, nous avons synthétisé une forte augmentation de hauteur ou de durée sur le constituant focalisé. Comme nous l'avons vu (cf § 3.2.2), cette réalisation est associée par Rossi (1999) et Di Cristo (1999b) à la fonction d'expressivité. Nous avons là aussi sélectionné des énoncés dont le contexte d'origine est propice à la réalisation de cette fonction. Dans l'exemple (ii) ci-dessous, l'augmentation de hauteur ou de durée porte sur *vraiment* :

- (ii) Il l'a déchirée en mille morceaux en disant : « Écoutez ce que je dis, plutôt que de prendre des notes ! »
Il s'est VRAIMENT énervé.

⁶⁹ Pour obtenir cette réalisation, nous avons simplement stylisé la courbe de hauteur à l'aide de Momel-Intsint, sans la manipuler.

⁷⁰ Pour obtenir cette réalisation, nous avons notamment augmenté la hauteur de certaines syllabes au-delà de la limite supérieure du registre tonal du locuteur.

et que l'expérience avait pour but d'améliorer la qualité de ce logiciel⁷¹. L'inclusion de cette information dans les consignes est due aux résultats d'un test préliminaire que nous avons conduit avec cinq participants⁷². Ce test avait pour but de vérifier la faisabilité générale de l'expérience. Les résultats ont révélé que les participants ont attribué des notes basses à toutes les réalisations phonétiques. Après leur avoir fait passer le test, nous avons demandé aux participants quelles améliorations pourraient être apportées à l'expérience. Plusieurs participants ont mentionné le fait que la plupart des réalisations leur ont paru peu acceptables. Ce résultat est probablement dû au fait que nous avons réenregistré les stimuli avec une prosodie relativement peu marquée (cf § 5.2.2). L'information dans les consignes concernant le logiciel de synthèse de la parole a pour but d'atténuer ce biais.

- Les participants ont tout d'abord évalué les cinq items d'entraînement. Nous avons informé les participants que ces évaluations constituaient une phase d'entraînement.
- Les participants ont ensuite évalué les trente énoncés de l'expérience. Les énoncés ont été présentés dans un ordre aléatoire aux participants. Nous illustrons ci-dessous (cf Figure 48) le déroulement de l'évaluation pour chaque énoncé. Les participants voient d'abord apparaître le contexte d'origine de l'énoncé. Ils entendent simultanément sa version enregistrée. Ils voient ensuite apparaître l'énoncé. Ils entendent alors successivement les quatre réalisations phonétiques de l'énoncé. Après chaque réalisation, une consigne apparaît pour demander aux participants d'évaluer la réalisation sur une échelle de zéro à cinq^{73 74}. Les participants n'ont pas eu la possibilité d'écouter plusieurs fois le contexte

⁷¹ Une fois l'expérience terminée, pour des raisons éthiques, nous avons informé les participants que ces informations étaient en réalité fausses.

⁷² Les participants du test préliminaire sont composés de trois hommes et deux femmes, possédant tous le français pour langue maternelle. Ils ont participé au test de manière bénévole.

⁷³ Pour attribuer leur note, les participants ont utilisé les touches numériques du clavier de l'ordinateur.

⁷⁴ Contrairement à d'autres échelles (par exemple de 1 à 5), cette échelle présente l'avantage de contenir un nombre pair de notes possibles (ici six). De cette façon, les participants ne peuvent pas choisir une note médiane, et sont forcés d'évaluer la réalisation plus ou moins positivement.

d'origine de l'énoncé, ni les réalisations phonétiques de l'énoncé (ceci en raison de limitations du logiciel).

- Les participants ont enfin évalué les cinq items de contrôle. Nous n'avons pas informé les participants que ces évaluations constituaient une phase de contrôle.

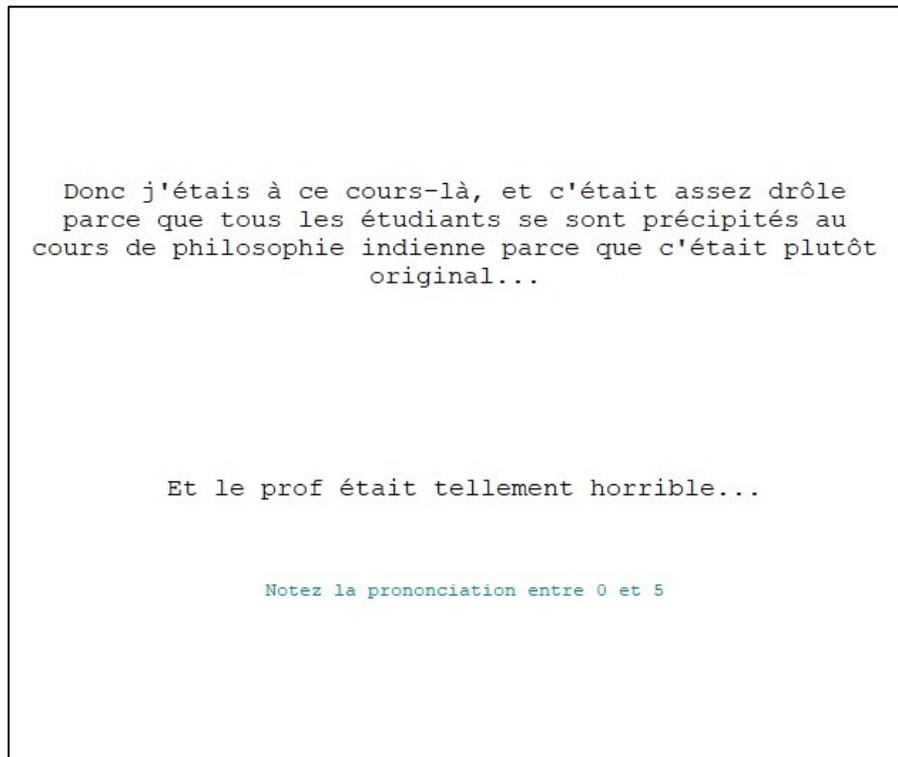


Figure 48. Capture d'écran de l'application E-Studio du logiciel E-Prime. Le contexte d'origine de l'énoncé est affiché en haut. Il est suivi de l'énoncé (« Et le prof était tellement horrible... »), puis de la consigne d'évaluation donnée aux participants (« Notez la prononciation entre 0 et 5 »).

5.3. Résultats

Nous avons standardisé en scores-z les notes attribuées aux réalisations phonétiques par les participants. Cette standardisation a pour but de tenir compte des différences entre les participants concernant la manière de noter les réalisations. Pour rappel, nous avons proposé aux participants une échelle de zéro à cinq. Certains participants ont utilisé une échelle plus réduite que cette échelle, en notant par exemple les réalisations de un à cinq, ou bien de zéro à quatre. La standardisation a été effectuée pour chaque participant en employant la formule

suivante : $z = \frac{x-m}{\sigma}$. Dans cette formule, x est la note attribuée par le participant, m est la moyenne des notes attribuées par le participant, et σ est l'écart-type des notes attribuées par le participant.

5.3.1. Evaluation des items de contrôle

Comme on peut le voir ci-dessous (cf Figure 49), les réalisations des items de contrôle ont été globalement évaluées de la manière attendue (cf § 5.2.2). On voit que, numériquement, la réalisation « non-acceptable » obtient des notes très inférieures aux autres réalisations. On voit également que les deux réalisations « acceptables » sont mieux notées que la réalisation « neutre ». Nous avons employé un modèle linéaire mixte afin de tester l'effet de la réalisation sur les notes des participants. Le modèle a été effectué sous le logiciel R, à l'aide de la fonction *lmer* de la bibliothèque « lme4 » (Bates et Maechler 2009). La variable dépendante est la note, et la variable fixe est la réalisation. Nous avons inclus le participant et l'item de contrôle en tant que variables aléatoires. Les valeurs-p ont été obtenues en effectuant un test du rapport des vraisemblances entre le modèle ainsi qu'un modèle ne contenant pas la variable de la réalisation. L'effet de la réalisation est significatif ($\chi^2(3) = 231,2$, $p < 0,001$). Nous présentons ci-dessous (cf Tableau 33) les résultats du modèle. On voit, d'après les valeurs-t et les valeurs-p⁷⁵, que les différences entre la réalisation « non-acceptable » et les trois autres réalisations sont significatives. On observe également une différence significative entre une des réalisations « acceptables » et la réalisation « neutre ». Les autres différences sont, en revanche, non-significatives.

⁷⁵ Pour rappel, pour que la différence de notation entre deux réalisations soit significative à 95%, la valeur-t doit être supérieure à 1,96 en valeur absolue, ce qui correspond à une valeur-p d'un maximum de 0,05.

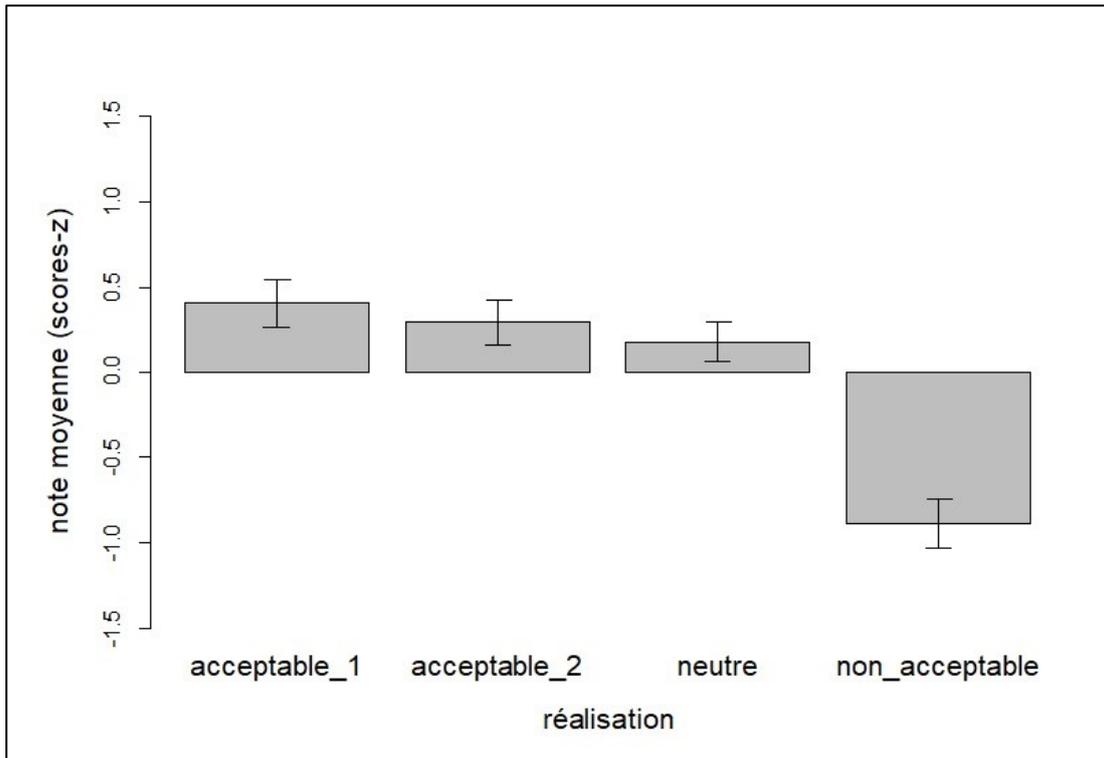


Figure 49. Notes standardisées moyennes attribuées aux quatre réalisations prosodiques des items de contrôle. Les barres d'erreur indiquent les intervalles de confiance à 95%.

	estimation (et erreur standard)	valeur-t	valeur-p
réal. acceptable 1/ réal. non-acceptable	1,29 (0,09)	14,788	< 0,001
réal. acceptable 2/ réal. non-acceptable	1,18 (0,09)	13,581	< 0,001
réal. neutre/ réal. non-acceptable	1,01 (0,09)	12,251	< 0,001
réal. acceptable 1/ réal. neutre	0,22 (0,09)	2,563	0,0106
réal. acceptable 2/ réal. neutre	0,12 (0,09)	1,333	0,1831
réal. acceptable 1/ réal. acceptable 2	0,11 (0,09)	1,232	0,218

Tableau 33. Résultats du modèle linéaire mixte concernant l'effet de la réalisation sur les notes attribuées aux items de contrôle par les participants de l'expérience de perception.

Nous avons exclu des résultats présentés ci-dessous (cf § 5.3.2, § 5.3.3 et § 5.3.4) les cinq participants ayant attribué des notes trop hautes (supérieures à 2) aux réalisations « non-acceptables » des items de contrôle. Le traitement des données porte donc sur les évaluations de vingt-cinq participants sur trente.

5.3.2. Evaluation des réalisations

On peut voir ci-dessous (cf Figure 50) les notes moyennes attribuées à chaque réalisation, indépendamment de la fonction. On voit que la réalisation 4 (sans manipulation) est numériquement la réalisation la mieux notée, suivie de la réalisation 1 (augmentation de hauteur). La réalisation 2 (augmentation de durée) et la réalisation 3 (augmentation de hauteur et de durée) obtiennent des notes inférieures. Afin de tester la significativité de l'effet de la réalisation, nous avons employé un modèle linéaire mixte similaire à celui employé pour les items de contrôle (cf § 5.3.1). L'effet de la réalisation est significatif ($\chi^2(3) = 111,9$, $p < 0,001$). Nous présentons ci-dessous (cf Tableau 34) les résultats du modèle. On voit que les différences entre la réalisation 4 et les trois autres réalisations sont significatives. Les

différences entre la réalisation 1 et les réalisations 2 et 3 sont également significatives. La différence entre les réalisations 2 et 3 n'est en revanche pas significative.

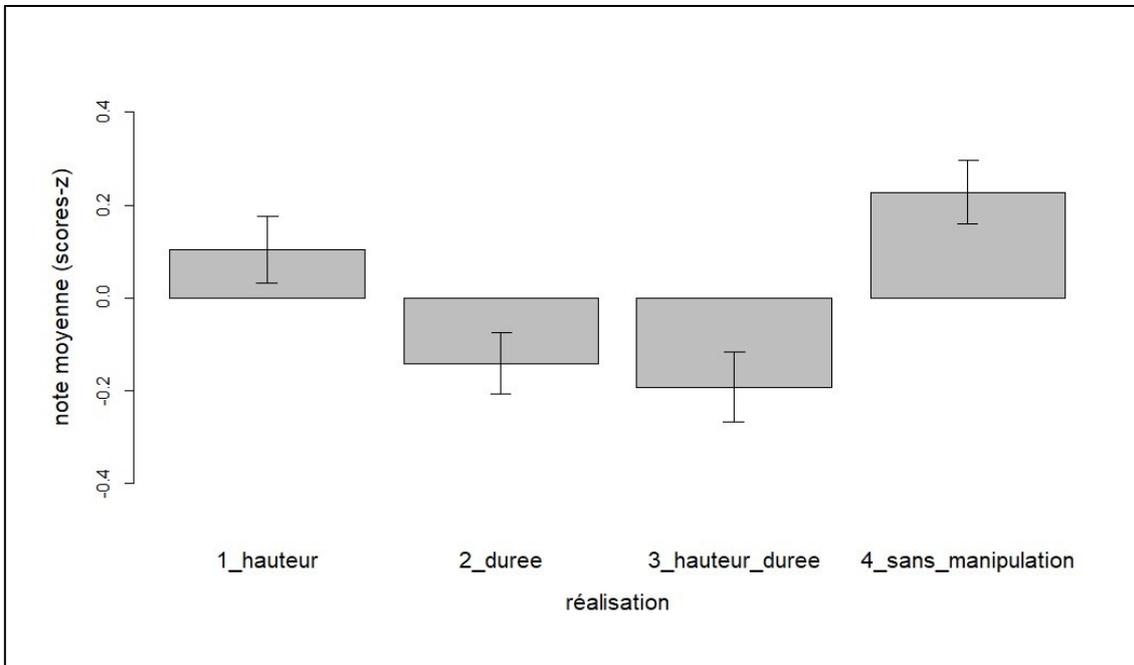


Figure 50. Notes standardisées moyennes attribuées aux quatre réalisations phonétiques (toutes fonctions confondues). Les barres d'erreur indiquent les intervalles de confiance à 95%.

	estimation (et erreur standard)	valeur-t	valeur-p
hauteur/ sans manipulation	-0,12 (0,05)	-2,664	0,00777
durée/ sans manipulation	-0,37 (0,05)	-8,063	< 0,001
hauteur et durée/ sans manipulation	-0,42 (0,05)	-9,137	< 0,001
hauteur/durée	0,25 (0,05)	5,394	< 0,001
hauteur/ hauteur et durée	0,30 (0,05)	6,470	< 0,001
durée/ hauteur et durée	0,05 (0,05)	1,082	0,2795

Tableau 34. Résultats du modèle linéaire mixte concernant l'effet de la réalisation sur les notes attribuées aux énoncés par les participants de l'expérience de perception.

5.3.3. Evaluation des fonctions

On peut voir ci-dessous (cf Figure 51) les notes moyennes attribuées à chaque fonction, indépendamment de la réalisation. On voit que l'insistance est numériquement la fonction la mieux notée, suivie du marquage de focus et de l'expressivité. Afin de tester la significativité de l'effet de la fonction, nous avons employé un modèle linéaire mixte avec la note comme variable dépendante, la fonction comme variable fixe, et le participant et l'item comme variables aléatoires. L'effet de la fonction est significatif ($\chi^2(3) = 18,3$, $p < 0,001$). Nous présentons ci-dessous (cf Tableau 35) les résultats du modèle. On voit que les différences entre l'insistance et les deux autres fonctions sont significatives. La différence entre l'expressivité et le marquage de focus n'est en revanche pas significative.

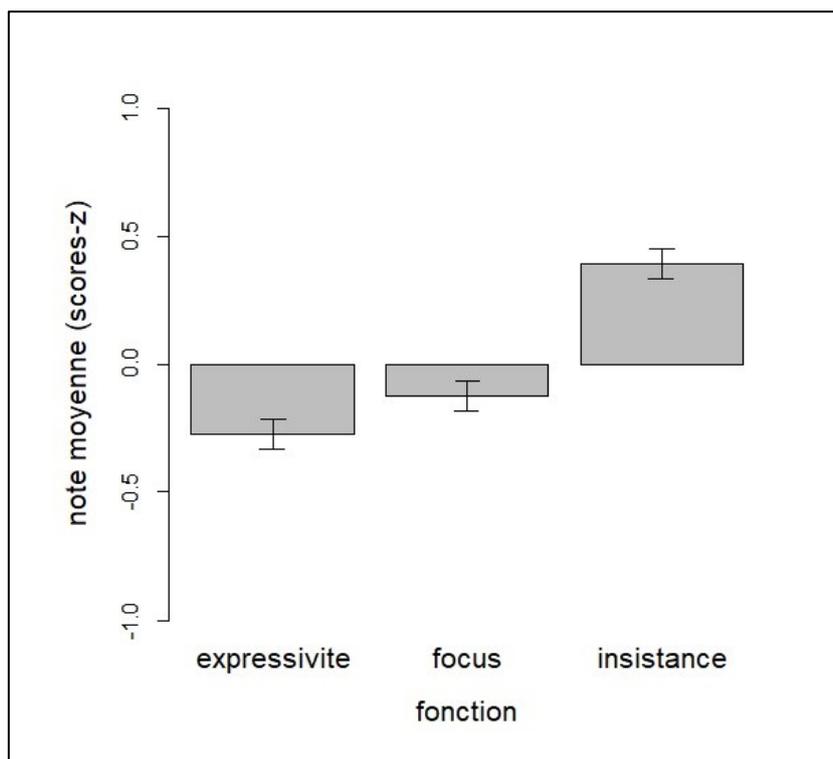


Figure 51. Notes standardisées moyennes attribuées aux trois fonctions analysées (toutes réalisations confondues). Les barres d'erreur indiquent les intervalles de confiance à 95%.

	estimation (et erreur standard)	valeur-t	valeur-p
expressivité/ insistance	-0,67 (0,15)	-4,545	0,000104
expressivité/ marquage de focus	-0,15 (0,15)	-1,030	0,31221
insistance/ marquage de focus	0,52 (0,15)	3,515	0,00157

Tableau 35. Résultats du modèle linéaire mixte concernant l'effet de la fonction sur les notes attribuées aux énoncés par les participants de l'expérience de perception.

5.3.4. Evaluation des réalisations par fonction

On peut voir ci-dessous (cf Figure 52), pour chaque fonction, les notes moyennes attribuées à chaque réalisation. On observe plusieurs différences numériques :

- **expressivité** : La réalisation 4 (sans manipulation) est mieux notée que les trois autres réalisations. Parmi ces trois réalisations, la réalisation 1 (augmentation de hauteur) est la mieux notée, suivie de la réalisation 2 (augmentation de durée) et de la réalisation 3 (augmentation de hauteur et de durée).
- **marquage de focus** : On observe une configuration similaire à celle de l'expressivité. Cependant, la réalisation 1 obtient une note supérieure.
- **insistance** : La réalisation 2 (augmentation de durée) est moins bien notée que les trois autres réalisations. Parmi ces trois réalisations, la réalisation 1 est la mieux notée, suivie de la réalisation 3 et de la réalisation 4.

Afin de tester la significativité de l'effet de la réalisation, nous avons employé un modèle linéaire mixte pour chaque fonction (en excluant des données les évaluations portant sur les énoncés des autres fonctions). Nous présentons ci-dessous (cf Tableau 36, Tableau 37 et Tableau 38) les résultats des modèles pour chaque fonction. On observe les résultats suivants :

- **expressivité** : L'effet est significatif. Les différences suivantes sont significatives :
 - La réalisation 4 (sans manipulation) est mieux évaluée que les trois autres réalisations.
 - La réalisation 1 (augmentation de hauteur) et la réalisation 2 (augmentation de durée) sont mieux évaluées que la réalisation 3 (augmentation de hauteur et de durée).
- **marquage de focus** : L'effet est significatif. Les différences suivantes sont significatives :
 - La réalisation 1 et la réalisation 4 sont mieux évaluées que les réalisations 2 et 3.
 - La réalisation 2 est mieux évaluée que la réalisation 3.

- **insistance** : L'effet est significatif. On observe une seule différence significative : les réalisations 1, 3 et 4 sont mieux évaluées que la réalisation 2 (augmentation de durée).

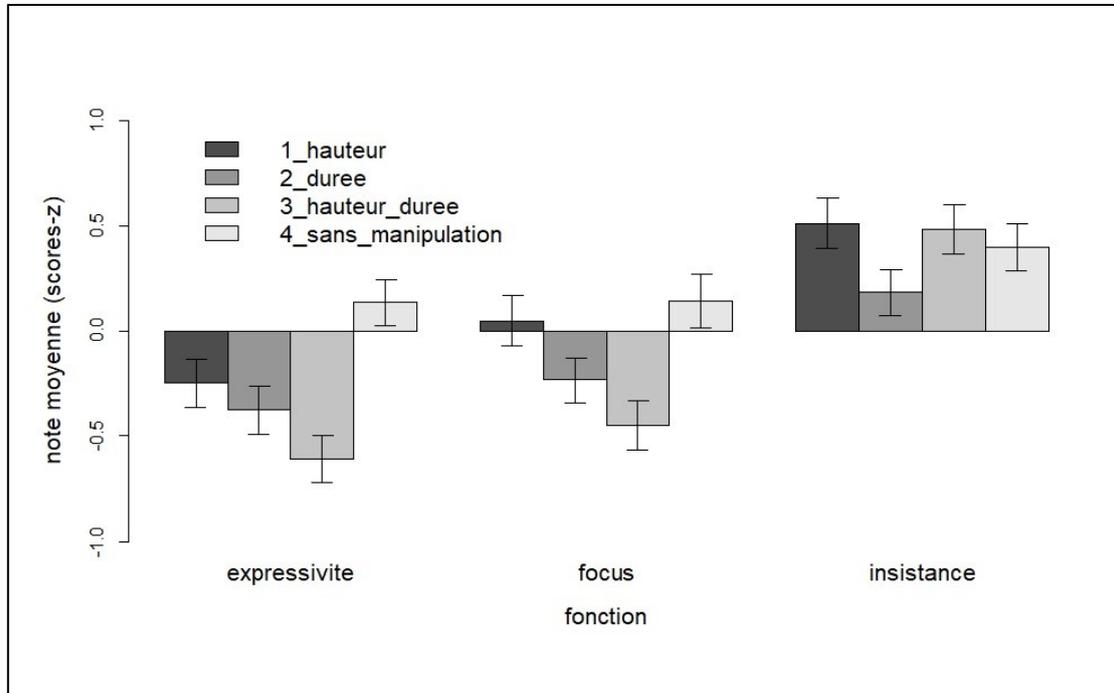


Figure 52. Notes standardisées moyennes attribuées aux quatre réalisations phonétiques pour chaque fonction. Les barres d'erreur indiquent les intervalles de confiance à 95%.

expressivité	$\chi^2(3) = 86,0, p < 0,001$		
	estimation (et erreur standard)	valeur-t	valeur-p
hauteur/ sans manipulation	-0,39 (0,08)	-4,774	< 0,001
durée/ sans manipulation	-0,52 (0,08)	-6,369	< 0,001
hauteur et durée/ sans manipulation	-0,75 (0,08)	-9,230	< 0,001
hauteur/durée	0,13 (0,08)	1,588	0,11254
hauteur/ hauteur et durée	0,36 (0,08)	4,451	< 0,001
durée/hauteur et durée	0,23 (0,09)	2,867	0,00423

Tableau 36. Résultats du modèle linéaire mixte concernant l'effet de la réalisation sur les notes attribuées aux énoncés contenant une occurrence d'expressivité par les participants de l'expérience de perception.

marquage de focus	$\chi^2(3) = 71,0, p < 0,001$		
	estimation (et erreur standard)	valeur-t	valeur-p
hauteur/ sans manipulation	-0,08 (0,08)	-1,026	0,3052
durée/ sans manipulation	-0,38 (0,08)	-4,846	< 0,001
hauteur et durée/ sans manipulation	-0,60 (0,08)	-7,576	< 0,001
hauteur/durée	0,30 (0,08)	3,816	0,000144
hauteur/ hauteur et durée	0,52 (0,08)	6,543	< 0,001
durée/ hauteur et durée	0,22 (0,08)	2,728	0,00649

Tableau 37. Résultats du modèle linéaire mixte concernant l'effet de la réalisation sur les notes attribuées aux énoncés contenant une occurrence de marquage de focus par les participants de l'expérience de perception.

insistance	$\chi^2(3) = 22,7, p < 0,001$		
	estimation (et erreur standard)	valeur-t	valeur-p
hauteur/ sans manipulation	0,11 (0,08)	1,414	0,15764
durée/ sans manipulation	-0,24 (0,08)	-2,970	0,003046
hauteur et durée/ sans manipulation	0,07 (0,08)	0,849	0,39586
hauteur/durée	0,35 (0,08)	4,386	< 0,001
hauteur/ hauteur et durée	0,05 (0,08)	0,563	0,573371
durée/ hauteur et durée	-0,31 (0,08)	-3,818	0,000143

Tableau 38. Résultats du modèle linéaire mixte concernant l'effet de la réalisation sur les notes attribuées aux énoncés contenant une occurrence d'insistance par les participants de l'expérience de perception.

5.4. Discussion

5.4.1. Evaluation des réalisations et des fonctions

Avant de répondre à la question de recherche initiale (cf § 5.1), nous discutons tout d'abord les différences globales d'évaluation qui sont observées entre les réalisations (cf Figure 50) ainsi qu'entre les fonctions (cf Figure 51) :

- La réalisation 4 (sans manipulation prosodique sur la syllabe initiale) est significativement mieux évaluée que les autres réalisations (avec augmentation de hauteur ou de durée sur la syllabe initiale). Cette différence d'évaluation est probablement due à une différence de naturalité entre les réalisations. Pour rappel (cf § 5.2.2), la réalisation 4 a simplement été obtenue en stylisant la courbe de hauteur de l'énoncé, sans manipuler celle-ci. L'absence d'augmentation de hauteur ou de durée a probablement rendu cette réalisation plus naturelle pour les participants, indépendamment de la fonction de l'énoncé. Le résultat est donc peu pertinent pour notre analyse, et ne sera pas pris en compte dans la discussion qui suit.

- La réalisation 1 (augmentation de hauteur) est significativement mieux notée que la réalisation 2 (augmentation de durée) et que la réalisation 3 (augmentation de hauteur et de durée). Ce résultat est inattendu, en particulier concernant la réalisation 3. En effet, d'après la littérature (cf chap. 1, § 2.4), l'accent initial est réalisé en français par une augmentation de hauteur, mais aussi de durée (et ce quelle que soit sa fonction). Dans l'étude présentée dans le chapitre précédent (cf chap. 2, § 5.3.2), nous avons également observé que la focalisation prosodique est réalisée par une augmentation significative de durée. Il est possible que l'augmentation de durée ait rendu les réalisations 2 et 3 moins naturelles que la réalisation 1 (malgré le seuil d'augmentation qui a été déterminé, cf § 5.2.2).
- La fonction d'insistance est significativement mieux notée que les fonctions d'expressivité et de marquage de focus. Ce résultat est probablement dû au fait que les manipulations prosodiques effectuées sur les stimuli portent uniquement sur la syllabe initiale du constituant focalisé (cf § 5.2.2). En effet, l'étude précédente a montré (cf § 4.3.6) que l'insistance présente un taux d'accentuation initiale significativement plus élevé que les deux autres fonctions. La relation entre insistance et accent initial est par ailleurs souvent décrite dans la littérature (cf § 3.2.1). Bien que cette relation ne fasse pas directement l'objet de cette expérience, elle est confirmée par notre étude.

5.4.2. Réalisation phonétique des fonctions

La question de recherche concerne les différences de réalisation phonétique entre les fonctions attribuées à la focalisation. Nous discutons ci-dessous les différences d'évaluation qui sont observées entre les réalisations concernant chaque fonction (cf Figure 52) :

- Concernant l'insistance, on observe que la réalisation 2 (augmentation de durée) est significativement moins bien évaluée que les autres réalisations. Ce résultat correspond à nos prédictions (cf § 5.1), selon lesquelles la réalisation 1 (augmentation de hauteur) et la réalisation 3 (augmentation de hauteur et de durée) seraient plus fréquentes pour cette fonction. Pour rappel, ces prédictions sont basées sur la tendance non-significative qui a été observée dans l'étude de production (cf § 4.3.2). Dans cette étude, l'insistance présente une hauteur supérieure à celle des deux autres fonctions. Cette tendance est

donc confirmée par l'expérience de perception. Cependant, le résultat est contradictoire avec certaines études précédentes (cf § 3.2.1), qui attribuent également une augmentation de durée à l'insistance.

- Concernant l'expressivité, on observe que la réalisation 1 (augmentation de hauteur) et la réalisation 2 (augmentation de durée) sont significativement mieux évaluées que la réalisation 3 (augmentation de hauteur et de durée). Ce résultat ne correspond pas entièrement à nos prédictions, selon lesquelles les réalisations 2 et 3 seraient les plus fréquentes chez cette fonction. Pour rappel, ces prédictions sont basées sur les différences qui ont été observées dans l'étude de production (cf § 4.3.3). Dans cette étude, l'expressivité présente, de manière non-significative, une durée syllabique supérieure à celle de l'insistance. De plus, ces deux fonctions présentent une durée significativement supérieure à celle du marquage de focus. Le résultat de l'expérience de perception est également contradictoire avec la littérature, qui attribue à l'expressivité une forte augmentation de hauteur, mais aussi de durée (cf § 3.2.2). La faible note moyenne attribuée à la réalisation 3 est peut-être due au fait qu'il s'agit de la réalisation qui implique le plus de manipulation prosodique. Cela a pu rendre cette réalisation moins naturelle que les autres pour les participants. Cependant, notons qu'il est possible de faire une autre interprétation des différences d'évaluation entre les réalisations pour l'expressivité. Le fait que la réalisation 2 soit significativement mieux évaluée que la réalisation 3 pourrait en effet signifier qu'il existe bien une relation entre l'augmentation de durée et cette fonction. La faible note moyenne de la réalisation 3 serait alors due à la présence d'une augmentation de hauteur dans cette réalisation. En effet, on voit que la réalisation 1 (augmentation de hauteur) obtient elle aussi une note moyenne assez faible. Cela signifierait donc que l'expressivité est plus reliée à l'augmentation de durée qu'à l'augmentation de hauteur. Cela confirmerait la tendance non-significative qui a été observée dans l'étude de production.
- Concernant le marquage de focus, on observe que la réalisation 1 (augmentation de hauteur) et la réalisation 4 (sans manipulation) sont significativement mieux évaluées que la réalisation 2 (augmentation de durée) et que la réalisation 3 (augmentation de hauteur et de durée). La réalisation 2 est également significativement mieux évaluée que la

réalisation 3. Ces résultats correspondent en partie à nos prédictions, selon lesquelles la réalisation 4 serait la plus fréquente chez cette fonction. Pour rappel, ces prédictions sont basées, là aussi, sur les résultats observés dans l'étude de production. Dans cette étude, le marquage de focus présente une hauteur inférieure à celle des autres fonctions, bien que la différence soit non-significative (cf § 4.3.2). De manière significative, le marquage de focus présente également une durée inférieure à celle des autres fonctions. Cependant, comme nous l'avons mentionné au début de cette discussion, le fait que la réalisation 4 soit bien évaluée est probablement simplement dû à une différence de naturalité par rapport aux autres réalisations, et ne devrait pas être pris en compte. Concernant les autres réalisations, la supériorité de la note de la réalisation 1 suggère que le marquage de focus est plus relié à l'augmentation de hauteur qu'à l'augmentation de durée. Cependant, on s'attendrait alors à ce que la réalisation 3 soit mieux notée que la réalisation 2, ce qui n'est pas le cas. Les résultats sont par ailleurs contradictoires avec la littérature, qui attribue généralement une augmentation de hauteur, mais aussi de durée, au marquage des différentes catégories de focus (cf § 2.2).

5.5. Conclusions

Cette étude permet de répondre en partie à la première question de recherche de cette analyse (cf § 1.1). En effet, on observe certaines relations entre les réalisations et les fonctions analysées :

- L'insistance présente une préférence pour l'augmentation de hauteur. On observe également une forte relation entre l'insistance et l'accentuation initiale (bien que cette relation ne fasse pas directement l'objet de l'expérience).
- L'expressivité présente, de manière potentielle, une préférence pour l'augmentation de durée. Cette préférence se base toutefois sur une différence d'évaluation entre deux réalisations ayant globalement obtenu des évaluations assez basses. Il s'agit donc d'une relation qui reste à explorer.
- Le marquage de focus présente une légère préférence pour l'augmentation de hauteur. Cette relation reste également à explorer.

Notons que le protocole de l'expérience possède plusieurs limitations :

- Les stimuli présentent un certain manque de naturalité. En effet, nous avons vu (cf § 5.2.2) qu'il s'agit d'énoncés issus du corpus de l'expérience de production, qui ont été réenregistrés avec une prosodie relativement neutre, puis manipulés prosodiquement. Afin d'atténuer ce biais, nous avons fait croire aux participants (cf § 5.2.3) que les stimuli ont été produits par un logiciel de synthèse de la parole. Néanmoins, il est probable que ce manque de naturalité ait influencé les participants dans leur évaluation des réalisations phonétiques. On voit en effet que la réalisation 4 (pour laquelle l'énoncé est réenregistré mais pas manipulé prosodiquement) est globalement la réalisation la mieux évaluée (cf Figure 50).
- Afin de tester l'existence de relations entre réalisations phonétiques et fonctions de la focalisation, nous avons employé une méthode indirecte. Nous avons demandé aux participants (cf § 5.1) d'évaluer l'acceptabilité de chaque réalisation en fonction du contexte d'origine de l'énoncé dans le corpus de l'expérience de production. On peut en effet considérer que le contexte contribue à indiquer la fonction de l'occurrence de focalisation présente dans l'énoncé. Cependant, il est possible que le caractère indirect de cette méthode ait eu une influence sur les résultats.
- L'expérience porte uniquement sur la réalisation phonétique des fonctions de la focalisation. Seuls les paramètres de la hauteur et de la durée sont étudiés. De plus, les variations de hauteur et de durée portent uniquement sur la syllabe initiale du constituant focalisé. De nouvelles expériences de perception peuvent être menées, afin de prendre en compte les catégories phonologiques que nous avons analysées dans les deux études de production (contour intonatif, étendue du contour, et accentuation initiale). Comme nous l'avons mentionné plus haut (cf § 4.5), des traits prosodiques supplémentaires peuvent également être pris en compte : les variations d'intensité et de qualité vocalique, la configuration tonale et la constituance syllabique des accents de focalisation (cf Astésano 2001), ainsi que la désaccentuation post-focale ou pré-focale.

6. Discussion générale

Dans ce chapitre, nous avons posé deux questions de recherche (cf § 1.1). Nous rappelons une nouvelle fois ces questions ci-dessous :

Question 1 : Les fonctions attribuées à la focalisation présentent-elles des réalisations différentes ?

Question 2 : La réalisation des fonctions de la focalisation est-elle influencée par le phonogène ?

Afin de répondre à ces questions, nous avons mené une étude de production (cf § 4) et une étude de perception (cf § 5). Ces études nous ont principalement permis de répondre à la première question. En effet, les résultats de l'analyse de l'influence du phonogène dans l'étude de production sont difficilement interprétables, notamment en raison du manque de données (cf § 4.4.3). Concernant la question de la réalisation des fonctions de la focalisation, nous avons vu (cf § 4.4.1) que la portée des résultats de l'étude de production est diminuée par la faiblesse du taux d'accord entre les experts qui ont annoté le corpus. En raison de cette faiblesse du taux d'accord, nous n'avons par ailleurs pu analyser que trois fonctions : l'insistance, l'expressivité, et le marquage de focus (en tant que fonction globale). Nous rappelons ci-dessous les résultats de l'expérience de production (cf § 4.4.2) :

- On observe des différences de hauteur et de durée syllabique entre les fonctions. L'emphase (qui comprend l'insistance et l'expressivité) présente une durée supérieure à celle du marquage de focus. L'expressivité présente, bien que de manière non-significative, une durée supérieure à celle de l'insistance. Concernant la hauteur, des différences non-significatives sont également observées : l'emphase présente une hauteur supérieure à celle du marquage de focus, et l'insistance présente une hauteur supérieure à celle de l'expressivité. Ces résultats suggèrent néanmoins qu'il existe bien une distinction phonétique entre les fonctions emphatiques et les fonctions sémantico-pragmatiques de la focalisation.
- On observe également des différences concernant le pourcentage d'occurrence de certaines catégories phonologiques :

- L'expressivité présente plus d'occurrences du contour H*H% que les autres fonctions. Il s'agit d'un contour montant de fin de groupe intonatif, que l'on observe notamment dans les énoncés exclamatifs. La modalité exclamative est cohérente avec la fonction d'expressivité.
- Les fonctions ne sont pas réalisées au même niveau de la structure prosodique. L'insistance présente plus de frontières de groupe accentuel que les autres fonctions, et moins de frontières de groupe intonatif. Ce résultat est cohérent avec la littérature. En effet, l'insistance est souvent décrite comme le soulignement d'un élément dans l'énoncé, tandis que le marquage de focus et l'expressivité peuvent affecter l'énoncé entier. On observe également que le marquage de focus présente plus de frontières de groupe intermédiaire que les autres fonctions. Cette différence est principalement due au plus grand nombre d'occurrences du contour H*H- pour le marquage de focus. Dans notre corpus, ce contour montant sert généralement à exprimer une continuation.
- L'insistance présente plus d'occurrences d'accent initial que les deux autres fonctions. Ce résultat confirme la relation, souvent rapportée dans la littérature, entre insistance et accentuation initiale.

Certains de ces résultats sont confirmés par l'expérience de perception (cf § 5.4.2) :

- L'insistance obtient globalement de meilleures évaluations que les autres fonctions. Etant donné que trois des quatre réalisations phonétiques ont été obtenues en augmentant la hauteur ou la durée sur la syllabe initiale du constituant focalisé, l'expérience confirme indirectement la relation entre insistance et accentuation initiale. Ce résultat, qui est significatif dans l'étude de production, constitue le principal résultat de cette analyse.
- On observe également, pour chaque fonction, plusieurs différences d'évaluation entre les réalisations :
 - Concernant l'insistance, les deux réalisations comprenant une augmentation de hauteur sont mieux évaluées que l'augmentation de durée. On a donc confirmation de

la relation entre insistance et augmentation de hauteur (qui est uniquement numérique dans l'étude de production).

- Concernant l'expressivité, une préférence potentielle est observée pour l'augmentation de durée. En effet, cette réalisation est mieux évaluée que l'augmentation de hauteur et de durée. La faible note moyenne de cette dernière réalisation pourrait être due à la présence de l'augmentation de hauteur (une réalisation qui obtient également une faible note moyenne). Bien que cette relation entre expressivité et augmentation de durée reste à explorer, elle est là aussi cohérente avec la différence numérique observée dans l'étude de production.

Le fait que certains résultats soient significatifs en perception et numériques seulement en production peut avoir plusieurs raisons :

- L'expérience de perception présente nécessairement un certain caractère artificiel (cf § 5.5). En effet, afin de pouvoir contrôler la variation prosodique entre les stimuli, l'expérience porte uniquement sur l'augmentation de hauteur ou de durée sur la syllabe initiale du constituant focalisé. Les stimuli présentent par ailleurs un certain manque de naturalité, étant donné qu'ils ont été réenregistrés et manipulés prosodiquement.
- Inversement, l'étude de production est moins contrôlée et contient donc plusieurs facteurs qui ont pu influencer les résultats (cf § 4.4.1). Le corpus analysé (composé de conversations spontanées et de leurs reproductions en parole lue et interprétée) ne se prête pas nécessairement à l'étude de toutes les fonctions attribuées à la focalisation. En effet, les études précédentes sur ces fonctions portent généralement sur de la parole de laboratoire. Nous avons également mentionné d'autres facteurs tels que la disparité dans la compétence des experts, ou bien l'influence du contenu sémantique des constituants focalisés.
- Notons enfin que plusieurs études ont démontré la présence de variations interindividuelles dans la production et la perception de la prosodie. Un certain nombre de ces études portent sur les différences prosodiques dans le marquage des différentes catégories de focus (e.g. Krahmer et Swerts 2001, Cangemi et al 2015, Ouyang et Kaiser

2015), ce qui est l'un des objets de notre analyse. Il est possible que de telles variations aient eu une influence sur les résultats de nos études.

Cette analyse présente plusieurs perspectives. En plus des traits prosodiques supplémentaires à prendre en compte (cf § 4.5 et § 5.5), il serait notamment intéressant de répéter l'étude de production en améliorant son protocole. Ce protocole est relativement novateur, et cette étude nous a permis d'identifier certaines de ses limitations. Le taux d'accord entre les experts concernant la classification fonctionnelle demande principalement à être amélioré. Cela permettrait d'analyser davantage de fonctions, et de vérifier les résultats observés en menant de nouvelles expériences de perception. Enfin, il est nécessaire de collecter davantage de données, afin de pouvoir effectuer une meilleure analyse de l'influence du phonogène sur la réalisation des fonctions.

7. Résumé

L'analyse présentée dans ce chapitre concerne la réalisation prosodique des fonctions attribuées à la focalisation. Ces fonctions se divisent en deux ensembles : le marquage de focus et l'emphase. L'analyse a également pour but d'évaluer l'influence du phonogène (parole spontanée, lue ou interprétée) sur la réalisation des fonctions. Les questions de recherche suivantes sont posées : 1) Les fonctions attribuées à la focalisation présentent-elles des réalisations différentes ? 2) La réalisation des fonctions de la focalisation est-elle influencée par le phonogène ? Afin de répondre à ces questions, l'analyse comprend une étude de production et une étude de perception.

Un état de l'art est consacré à la définition théorique et la réalisation prosodique des deux ensembles de fonctions de la focalisation, à commencer par le marquage de focus. Cet ensemble désigne le marquage des différentes catégories de focus (informationnel, contrastif, associatif et *verum*). La notion de focus désigne globalement la distinction d'une « figure » sur un « fond » dans un énoncé. Une définition est adoptée pour chacune des fonctions. Le marquage de focus informationnel désigne la distinction par le locuteur de ce qu'il affirme spécifiquement (la totalité du contenu dans le cas d'un focus informationnel large, et une partie seulement du contenu dans le cas d'un focus informationnel étroit). Le marquage de focus contrastif désigne la distinction d'un élément appartenant à un ensemble d'alternatives

présent dans le contexte de discours immédiat. Le marquage de focus associatif désigne la distinction de l'argument d'un opérateur sémantique tel que *seulement* ou *aussi*. Le marquage de verum focus consiste à insister sur la vérité de la proposition. Nous présentons ensuite les résultats des études sur la réalisation prosodique des différentes fonctions en français. Enfin, nous évaluons le poids des méthodes expérimentales de ces études dans les résultats observés.

L'ensemble des fonctions emphatiques est ensuite abordé. Cet ensemble comprend les fonctions d'insistance et d'expressivité. Ces fonctions ont notamment été étudiées par le domaine de la phonostylistique. L'insistance désigne le soulignement, auprès de l'interlocuteur, d'une partie ou de la totalité du contenu de l'énoncé. L'expressivité désigne l'expression d'une émotion ou une attitude par rapport à une partie ou la totalité du contenu de l'énoncé. L'état de l'art concernant la réalisation prosodique de ces deux fonctions en français est ensuite présenté, puis le poids des méthodes expérimentales est évalué.

Les deux études sont ensuite présentées, à commencer par l'étude de production. Cette étude porte sur le même corpus multi-genres que l'étude présentée dans le chapitre précédent. Elle a tout d'abord consisté à établir une classification fonctionnelle des occurrences de focalisation relevées dans le corpus. La classification a été effectuée par le même groupe d'experts qui a effectué le relevé des occurrences dans l'étude précédente. Les experts se sont tous référés aux mêmes définitions théoriques des fonctions de la focalisation. Seules les occurrences auxquelles trois experts sur quatre ont attribué la même fonction ont été conservées pour l'analyse prosodique. Cette dernière est identique à celle de l'étude précédente. Les résultats révèlent tout d'abord que le taux d'accord entre les experts est globalement bas pour toutes les fonctions. Seules trois fonctions présentent un taux d'accord suffisamment haut pour pouvoir être analysées : l'insistance, l'expressivité, et le marquage de focus (en tant que fonction globale). L'insistance et l'expressivité présentent un taux d'accord supérieur à celui du marquage de focus. Nous présentons ensuite les différences de fréquence d'occurrence entre ces fonctions. L'insistance est la fonction la plus fréquente, suivie de l'expressivité. La distribution de ces fonctions parmi les phonogenres est hétérogène : l'insistance et l'expressivité sont majoritaires en parole interprétée, tandis que le marquage de focus est majoritaire en parole lue. Nous présentons ensuite les résultats de l'analyse prosodique. Aucune différence significative de hauteur n'est observée. En revanche,

l'insistance et l'expressivité présentent une durée syllabique plus élevée que celle du marquage de focus. On observe une influence du phonogène sur les différences de durée entre l'expressivité et les autres fonctions. Concernant les catégories phonologiques des constituants focalisés, plusieurs différences sont observées. L'insistance présente plus d'accents initiaux et de contours Hi (ton haut initial) que les autres fonctions. L'expressivité présente plus de contours H*H% (contour montant de fin de groupe intonatif) que l'insistance. Des différences sont observées concernant le type de frontière prosodique : l'insistance présente plus de frontières de groupe accentuel et moins de frontières de groupe intonatif que les autres fonctions, et le marquage de focus présente plus de frontières de groupe intermédiaire. Enfin, comme dans l'étude précédente, nous présentons les résultats d'une analyse menée sur un corpus réduit constitué des enregistrements ayant obtenu les meilleurs taux d'accord entre les experts concernant le relevé des occurrences de focalisation. Les résultats sont globalement similaires aux résultats de l'analyse menée sur le corpus entier. Les exceptions concernent tout d'abord le taux d'accord entre experts, qui augmente pour les fonctions d'insistance et d'expressivité. Les différences de durée syllabiques deviennent par ailleurs toutes non-significatives, de même que les différences concernant les contours intonatifs et les frontières de groupe intermédiaire.

Une interprétation est effectuée des résultats de l'étude de production. La faiblesse du taux d'accord entre les experts réduit la portée des résultats de l'analyse prosodique. Elle soulève également plusieurs questions méthodologiques et théoriques. Il est possible que les définitions théoriques des fonctions n'aient pas été bien assimilées par les experts, ou que ceux-ci se soient référés à d'autres définitions présentes dans la littérature. Une disparité dans la compétence des experts pourrait également avoir joué un rôle. Le corpus d'étude (qui comprend des conversations spontanées et leur reproduction en parole lue et interprétée) ne permet peut-être pas d'obtenir suffisamment d'occurrences de chaque fonction. Il est enfin possible que la distinction théorique entre les fonctions ne soit pas reflétée au niveau prosodique. La focalisation remplirait donc une seule fonction globale, de nature à la fois sémantico-pragmatique et expressive. La différence de taux d'accord entre l'emphase et le marquage de focus pourrait quant à elle s'expliquer par les différences de hauteur et de durée syllabique observées entre ces fonctions, ou bien par la présence fréquente de contenu intensif ou expressif dans les occurrences d'emphase. Nous discutons ensuite les différences de fréquence d'occurrence observées entre les fonctions. Le fait que l'emphase soit plus

fréquente que le marquage de focus est cohérent avec la différence de taux d'accord entre ces deux fonctions. Parmi les occurrences d'emphase, le fait que l'insistance soit plus fréquente que l'expressivité pourrait s'expliquer par la forte relation qui est observée dans le corpus entre l'insistance et l'accentuation initiale (cette relation étant par ailleurs souvent décrite dans la littérature). La question de recherche 1 est ensuite abordée. Les différences de hauteur et de durée syllabique entre les fonctions suggèrent qu'il existe bien une distinction phonétique entre l'emphase et le marquage de focus. Ces différences sont par ailleurs globalement cohérentes avec la littérature. La relation entre l'insistance et l'accentuation initiale confirme également la littérature. La relation entre l'expressivité et le contour H*H% est cohérente, car une des fonctions possibles de ce contour est l'indication de la modalité exclamative de l'énoncé. Enfin, les différences concernant le type de frontière prosodique sont cohérentes : l'insistance est en effet souvent décrite comme le soulignement d'un élément de l'énoncé, tandis que l'expressivité et le marquage de focus peuvent affecter l'énoncé entier. La question de recherche 2 est ensuite abordée. Les résultats concernant l'influence du phonogène sur la réalisation phonétique des fonctions sont difficilement interprétables, notamment en raison du manque de données. Cependant, l'influence du phonogène sur la fréquence d'occurrence des fonctions est un résultat intéressant : il suggère notamment que la parole interprétée est un phonogène favorable à l'étude de l'emphase, tandis que la parole lue est favorable à l'étude du marquage de focus. Pour conclure, nous rappelons les moyens possibles de renforcer l'étude de production (amélioration du protocole expérimental, et prise en compte de plus de traits dans l'analyse prosodique).

L'expérience de perception est ensuite présentée. Cette expérience concerne uniquement la question de recherche 1. Elle a pour but de vérifier les résultats observés dans l'étude de production concernant les différences de hauteur et de durée syllabique entre les fonctions. Elle a consisté à obtenir des jugements d'acceptabilité sur des énoncés contenant une occurrence de focalisation. Ces énoncés sont extraits du corpus de l'expérience de production. Leur prosodie a été manipulée, afin d'obtenir quatre réalisations possibles de la syllabe initiale du constituant focalisé : 1) augmentation de hauteur, 2) augmentation de durée, 3) augmentation de hauteur et de durée, 4) pas de manipulation. Les énoncés sont présentés aux participants précédés de leur contexte d'origine dans le corpus. Nous considérons que ce contexte contribue à indiquer aux participants la fonction de l'occurrence de focalisation. Cette fonction est déterminée sur la base de la classification effectuée par le groupe d'experts

dans l'étude de production. L'expérience compte trente participants. Les stimuli comprennent trente paires d'énoncé et de contexte (dix pour chaque fonction), cinq items d'entraînement et cinq items de contrôle. Les stimuli ont été réenregistrés avec une prosodie relativement neutre, puis manipulés prosodiquement en employant des seuils d'augmentation de hauteur et de durée (tirés de la littérature et du corpus de production). Les résultats montrent tout d'abord que la réalisation 4 est globalement la réalisation la mieux évaluée, suivie de la réalisation 1. On observe également que l'insistance est globalement mieux évaluée que les autres fonctions. Nous analysons ensuite les différences d'évaluation entre réalisations pour chaque fonction. Concernant l'insistance, la réalisation 2 est moins bien évaluée que les autres réalisations. Concernant l'expressivité, la réalisation 4 est la mieux évaluée, et les réalisations 1 et 2 sont mieux évaluées que la réalisation 3. Concernant le marquage de focus, les réalisations 1 et 4 sont mieux évaluées que les réalisations 2 et 3, et la réalisation 2 est mieux évaluée que la réalisation 3.

Une interprétation est effectuée des résultats de l'expérience de perception. Le fait que la réalisation 4 (sans manipulation) soit globalement la mieux évaluée est probablement dû à une différence de naturalité par rapport aux autres réalisations. Le résultat n'est donc pas pris en compte dans l'interprétation. Le fait que la réalisation 1 (augmentation de hauteur) soit globalement mieux évaluée que la réalisations 2 (augmentation de durée) et que la réalisation 3 (augmentation de hauteur et de durée) est inattendu. En effet, la littérature attribue à l'accent initial une augmentation de hauteur, mais aussi de durée. Le fait que l'insistance soit globalement mieux évaluée que les autres fonctions est probablement dû au fait que la manipulation prosodique porte sur la syllabe initiale du constituant focalisé. Le résultat confirme donc indirectement la relation entre l'insistance et l'accentuation initiale (déjà observée en production). La question de recherche 1 est ensuite abordée. Concernant l'insistance, la supériorité des réalisations 1 et 3 confirme une tendance significative observée en perception, mais est contradictoire avec la littérature. Concernant l'expressivité, la supériorité des réalisations 1 et 2 ne correspond pas entièrement aux différences observées en production, et est également contradictoire avec la littérature. Cependant, la supériorité de la réalisation 2 par rapport à la réalisation 3 pourrait également être interprétée comme une préférence de l'expressivité pour l'augmentation de durée (ce qui confirmerait une tendance non-significative observée en production). Concernant le marquage de focus, la supériorité de la réalisation 1 suggère que le marquage de focus présente une préférence pour

l'augmentation de hauteur. Ce résultat est en partie contradictoire avec la littérature, qui attribue également une augmentation de durée à cette fonction. Pour conclure, nous présentons les limitations du protocole de l'expérience. Les stimuli présentent un certain manque de naturalité. Une méthode indirecte a été employée pour tester l'existence de relations entre réalisations phonétiques et fonctions de la focalisation. Enfin, l'expérience porte uniquement sur l'augmentation de hauteur et de durée sur la syllabe initiale du constituant focalisé. Des expériences supplémentaires doivent être menées afin de prendre en compte d'autres traits prosodiques.

La discussion générale reprend les résultats de l'analyse. Le résultat principal est la relation entre l'insistance et l'accentuation initiale, qui est significative à la fois en production et en perception. Certains résultats sont significatifs en perception et numériques seulement en production : la relation entre l'insistance et l'augmentation de hauteur, et la relation entre l'expressivité et l'augmentation de durée. Cela pourrait être dû au caractère artificiel de l'expérience de perception (limitation de la variation prosodique entre les stimuli, et manque de naturalité des stimuli), ou bien aux nombreux facteurs aléatoires de l'étude de production (type de corpus, disparité dans la compétence des experts, et influence du contenu sémantique). De plus, plusieurs études ont démontré la présence de variations interindividuelles dans la production et la perception de la prosodie (notamment concernant le marquage de focus). Enfin, les perspectives de cette analyse sont présentées.

Conclusion générale

Au terme de cette recherche, nous pouvons résumer de la façon suivante les principaux apports de nos études :

- Nous analysons l'influence du phonogène (parole spontanée, lue ou interprétée) sur la fréquence d'occurrence et la réalisation prosodique de la focalisation. Bien que plusieurs études se soient déjà intéressées aux différences prosodiques entre ces phonogènes, aucune de ces études ne porte explicitement sur la focalisation. Nous observons que la fréquence d'occurrence de la focalisation est la plus haute en parole interprétée, suivie de la parole lue et de la parole spontanée. La réalisation prosodique n'est globalement pas influencée par le phonogène, à l'exception de la durée syllabique, qui est la plus élevée en parole lue.
- Nous proposons des protocoles expérimentaux relativement novateurs pour étudier la focalisation et ses fonctions. Nos études sont donc également instructives d'un point de vue méthodologique. Dans l'expérience de production, le relevé des occurrences de focalisation et l'assignation d'une fonction à ces occurrences ont été effectuées par un groupe d'experts en prosodie. Dans l'expérience de perception, les relations entre fonctions et traits prosodiques sont testées en sélectionnant des énoncés issus du corpus de production et en les présentant aux participants dans leur contexte d'origine. Il ressort principalement que, dans l'expérience de production, le taux d'accord entre les experts ayant effectué le relevé et la classification fonctionnelle des occurrences de focalisation est globalement faible, ce qui soulève plusieurs questions.
- Nous comparons la réalisation prosodique de trois fonctions attribuées à la focalisation : l'insistance, l'expressivité, et le marquage de focus (en tant que fonction globale). Les études précédentes portent généralement sur la réalisation d'une seule fonction à la fois. De plus, notre corpus comprend des conversations spontanées, alors que les études précédentes portent généralement sur des énoncés isolés lus à haute voix. Nous observons notamment une relation entre l'insistance et l'accentuation initiale, ce qui

confirme plusieurs études précédentes. L'insistance et l'expressivité présentent également une durée syllabique supérieure à celle du marquage de focus, et l'insistance présente la hauteur la plus importante. En revanche, nous n'observons aucune influence du phonogène sur la réalisation de ces fonctions.

- Nous confirmons certains des traits prosodiques attribués à la focalisation (augmentation de hauteur et de durée, et accentuation initiale) et nous faisons quelques observations supplémentaires sur sa réalisation. Nous évaluons également la fréquence d'occurrence de la focalisation dans notre corpus (11,3 %, soit environ un tiers du nombre total de proéminences du corpus).

Nous détaillons ci-dessous les résultats des analyses que nous avons menées, ainsi que leur interprétation. Nous présentons ensuite les perspectives de notre recherche.

Relevé des occurrences de focalisation

Le taux d'accord entre les experts ayant effectué le relevé des occurrences de focalisation est relativement bas, ce qui limite la portée des résultats de l'analyse de la focalisation. Cette faiblesse du taux d'accord pourrait être due à des défauts dans la méthodologie de l'expérience de production. Il est en effet possible que les experts aient eu des difficultés à relever les occurrences de focalisation de manière purement auditive, sans se baser sur un logiciel d'analyse acoustique. De plus, on note une disparité dans la compétence des experts (selon leur niveau académique, ainsi que leurs connaissances en phonétique-phonologie).

On observe également des différences de taux d'accord selon le phonogène. La parole spontanée présente le taux d'accord le plus élevé, et la parole lue le taux d'accord le plus bas. On aurait pu s'attendre à observer le résultat inverse, en raison des nombreuses disfluences de la parole spontanée et du caractère normatif de la parole lue. Cependant, plusieurs facteurs ont pu influencer ce résultat. Etant donné que les locuteurs de parole lue et de parole interprétée possèdent une expérience de l'art dramatique, leur prosodie relève peut-être d'un phonostyle du comédien, ce qui a pu rendre plus difficile l'annotation pour les experts. De plus, il est possible que les productions de parole lue et de parole interprétée soient moins naturelles que celles de parole spontanée (en raison de la double tâche cognitive qu'implique

la lecture à haute voix, et de l'aspect esthétique qu'implique l'interprétation d'un texte). Par ailleurs, en raison du plus faible nombre de productions spontanées dans le corpus, ces dernières ont toutes été annotées par les mêmes experts, ce qui n'est pas le cas des autres productions. Enfin, la supériorité du taux d'accord en parole interprétée par rapport à la parole lue pourrait être due au fait que les productions de parole interprétée ont été préparées à l'avance par les locuteurs (ce qui a pu introduire plus de régularité dans leur prosodie).

Fréquence d'occurrence de la focalisation

Le pourcentage de syllabes focalisées est de 11,3 % dans tout le corpus. A notre connaissance, c'est la première fois que la fréquence d'occurrence de la focalisation est ainsi évaluée. Il est intéressant de comparer cette fréquence au pourcentage de syllabes proéminentes détecté de manière automatisée dans le corpus, soit 35,8 %. Cela signifie qu'environ un tiers des proéminences du français sont des accents de focalisation, c'est-à-dire remplissant une fonction sémantico-pragmatique (marquage des différentes catégories de focus) ou emphatique (insistance et expressivité).

La fréquence d'occurrence de la focalisation varie selon le phonogène. Conformément à notre prédiction, elle est la plus élevée en parole interprétée (14,2 %), suivie de la parole lue (10,7 %) et de la parole spontanée (6,6 %). Ce résultat est cohérent avec l'intuition selon laquelle la parole interprétée présente une forte expressivité, ainsi qu'avec l'étude de Goldman et al (2011) qui montre que la fréquence d'accents initiaux augmente avec le degré de « préparation » du discours. De plus, cela suggère que la parole interprétée est, d'un point de vue méthodologique, un phonogène approprié pour étudier la focalisation prosodique. Cependant, notons que le résultat a pu être influencé par deux facteurs. Comme mentionné précédemment, les productions de parole lue et interprétée relèvent potentiellement d'un phonostyle du comédien. Par ailleurs, la parole interprétée présente globalement un débit plus lent que la parole lue (ce qui a pu faciliter le relevé des occurrences de focalisation pour les experts en parole interprétée).

Réalisation prosodique de la focalisation

Au niveau phonétique, la focalisation est réalisée par une forte augmentation de hauteur et de durée syllabique. Ces résultats confirment plusieurs études précédentes sur la focalisation en français. Au niveau phonologique, on observe tout d'abord une grande variété de contours intonatifs sur les constituants focalisés. De manière majoritaire, ces contours sont montants et commencent par un ton haut. Le contour le plus fréquent est H*H-, un contour de fin de groupe intermédiaire qui sert généralement à exprimer une continuation dans notre corpus. On observe également que la focalisation peut avoir lieu à tous les niveaux de la structure prosodique (groupe accentuel, intermédiaire ou intonatif), ce qui suggère que la focalisation peut aussi bien affecter un mot qu'un groupe de mots ou qu'un énoncé entier. Pour les constituants de plusieurs syllabes, l'étendue syllabique du contour intonatif varie peu, contrairement au type de contour : il s'agit majoritairement de la dernière syllabe. Une proportion importante mais minoritaire de constituants polysyllabiques comporte un accent initial, ce qui est cohérent avec la littérature, tout en montrant que ce trait n'est pas obligatoire.

L'augmentation de hauteur n'est pas influencée par le phonogène. Cependant, l'augmentation de durée syllabique est plus importante en parole lue qu'en parole interprétée. L'absence d'influence concernant la hauteur contredit certaines études sur les différences prosodiques entre parole spontanée et parole interprétée. On aurait en effet pu s'attendre à ce que la hauteur des syllabes focalisées soit plus importante en parole interprétée. Toutefois, les locuteurs de parole interprétée ont peut-être été inhibés par le contexte expérimental. La différence de durée entre parole lue et parole interprétée est d'autant plus intéressante qu'elle n'est pas due à une différence globale de débit entre les deux phonogènes. Cela suggère que la réalisation de la focalisation au moyen d'un fort allongement de durée est un trait caractéristique du phonogène de la lecture à haute voix.

On n'observe pas d'influence du phonogène sur les catégories phonologiques de la focalisation (contour intonatif, étendue du contour et accentuation initiale). Cela est surprenant concernant le type de contour intonatif (étant donné la grande variété qui a été observée), mais pas concernant l'étendue syllabique (étant donné que les contours ont majoritairement pour étendue la dernière syllabe du constituant focalisé). Concernant

l'accentuation initiale, l'absence d'effet du phonogène contredit l'étude de Goldman et al (2011), dans laquelle ce trait prosodique est relié au degré de préparation du phonogène.

Classification fonctionnelle des occurrences de focalisation

Le taux d'accord entre les experts concernant l'assignation d'une fonction aux occurrences de focalisation est globalement faible pour toutes les fonctions, ce qui réduit la portée de l'analyse de la réalisation prosodique de ces fonctions. Seules trois fonctions présentent un taux d'accord suffisamment haut pour être prises en compte dans cette analyse : l'insistance, l'expressivité, et le marquage de focus (envisagé en tant que fonction globale). La faiblesse globale du taux d'accord soulève plusieurs questions. Bien que les experts se soient tous référés aux mêmes définitions théoriques des fonctions (basées sur la littérature la plus récente), il est possible qu'ils n'aient pas bien assimilé ces définitions. De plus, comme nous l'avons évoqué, il existe une disparité dans la compétence des experts. Il est également possible que le protocole de l'expérience de production (qui consiste à enregistrer des conversations spontanées et à les faire reproduire en parole lue et interprétée) ne permette pas d'obtenir suffisamment d'occurrences de chaque fonction, contrairement à certains protocoles plus classiques consistant à enregistrer des énoncés indépendants. Enfin, d'un point de vue théorique, on peut se demander si la focalisation sert bien à remplir les différentes fonctions qui lui sont traditionnellement attribuées, ou bien une seule fonction globale de soulignement d'un constituant.

On observe également que le taux d'accord est plus élevé pour les fonctions emphatiques (insistance et expressivité) que pour le marquage des différentes catégories de focus. Ce résultat pourrait être dû au fait que, dans notre étude, l'emphase présente une durée syllabique (et, de manière non significative, une hauteur) supérieure à celle du marquage de focus. Il est possible que ce trait soit particulièrement associé à l'emphase pour les experts, ce qui aurait facilité l'assignation des fonctions emphatiques à certaines occurrences de focalisation. Par ailleurs, il est possible que le taux d'accord de l'emphase ait été influencé par la forte fréquence de mots intensifs ou expressifs parmi les constituants focalisés dans notre corpus.

Réalisation prosodique des fonctions de la focalisation

Au niveau phonétique, on observe certaines différences entre les fonctions analysées. L'insistance et l'expressivité présentent une durée syllabique (et, de manière non significative, une hauteur) supérieure à celle du marquage de focus. Cela suggère qu'il existe une distinction phonétique entre les fonctions emphatiques et les fonctions sémantico-pragmatiques de la focalisation. L'insistance se distingue par une forte augmentation de hauteur, et l'expressivité présente potentiellement une préférence pour l'augmentation de durée. Ces résultats sont uniquement numériques en production, et significatifs en perception.

Au niveau phonologique, on observe notamment une relation entre l'insistance et l'accentuation initiale. Ce résultat, significatif à la fois en production et en perception, confirme les études précédentes sur la réalisation de l'insistance. On observe également que le contour H*H% (un contour montant de fin de groupe intonatif), est plus fréquent pour l'expressivité que pour l'insistance. Ce résultat est cohérent, car une des fonctions possibles de ce contour est l'indication de la modalité exclamative de l'énoncé. Enfin, on observe que la réalisation des fonctions n'a pas lieu au même niveau de la structure prosodique. L'insistance est plus souvent réalisée au niveau du groupe accentuel que les deux autres fonctions, et moins souvent au niveau du groupe intonatif. Cela est cohérent avec la littérature sur ces fonctions : en effet, l'insistance porte généralement sur un élément de l'énoncé, tandis que l'expressivité et le marquage de focus peuvent porter sur l'énoncé entier. Le marquage de focus est, quant à lui, plus souvent réalisé au niveau du groupe intermédiaire que les deux autres fonctions. Cela est dû à la forte occurrence de contours H*H- pour le marquage de focus. Ce contour montant sert généralement à exprimer une continuation dans notre corpus. En revanche, nos résultats ne révèlent pas d'influence du phonogène sur la réalisation des fonctions. Bien que certaines différences soient observées concernant la hauteur et la durée syllabique, ces différences sont difficilement interprétables (notamment en raison du manque de données concernant certaines interactions entre fonctions et phonogènes).

Perspectives de recherche

Une première perspective consiste à mener de nouvelles expériences afin d'obtenir des résultats plus consistants. Du point de vue de la production, il est nécessaire d'obtenir plus de données, afin notamment de pouvoir analyser davantage de fonctions de la focalisation, ainsi

que l'influence du phonogène sur la réalisation de ces fonctions. Comme nous l'avons vu, le protocole de constitution et d'annotation du corpus présente plusieurs défauts. Cependant, il est probable que le taux d'accord entre les experts en prosodie ne puisse pas être amélioré de manière importante. En effet, bien que les consignes que nous avons données aux experts soient relativement claires, le fait d'avoir recours à des linguistes présente le risque qu'ils se basent sur leurs propres définitions des phénomènes étudiés. Or, en particulier concernant les fonctions de la focalisation, la littérature propose de nombreuses définitions différentes et parfois contradictoires. Une méthodologie plus sûre consisterait à avoir recours à des participants lambda pour relever les occurrences de focalisation et leur assigner une fonction. Ces participants porteraient des jugements naïfs sans se baser sur une terminologie linguistique (comme c'est le cas dans notre expérience). De nombreuses études récentes emploient une méthodologie similaire à travers la technologie du *crowdsourcing* (cf Gibson et Fedorenko 2010, Munro et Tily 2011). Du point de vue de la perception, le processus de manipulation prosodique entraîne nécessairement un certain manque de naturalité des stimuli. Une nouvelle expérience pourrait être menée avec des énoncés naturels dans lesquels le locuteur ferait varier les paramètres désirés (ici la hauteur et la durée de la première syllabe du constituant focalisé). La validité de ces stimuli serait contrôlée dans un second temps à l'aide d'un logiciel d'analyse acoustique.

Par ailleurs, plusieurs traits prosodiques n'ont pas été pris en compte dans nos analyses. Au niveau phonétique, il s'agit des paramètres d'intensité et de qualité vocalique, ainsi que la configuration tonale et la constituance syllabique des accents de focalisation (cf Astésano 2001). Au niveau phonologique, il s'agit en particulier du trait de désaccentuation post-focale ou pré-focale, qui est fréquemment attribué à la focalisation. Concernant ce trait, nous prévoyons également de mener une étude sur les contextes dans lesquels il est employé en parole interprétée. Comme nous l'avons expliqué dans le premier chapitre, une des fonctions attribuées à la désaccentuation est le marquage des informations au statut informationnel « ancien ». Nous souhaitons montrer qu'il arrive aux locuteurs de parole interprétée (et éventuellement de parole lue) de désaccentuer des constituants dont le statut informationnel est pourtant « nouveau » d'après le contexte. Ce phénomène est, selon nous, une erreur de langage due au fait que ces locuteurs savent à l'avance ce qu'ils vont dire, contrairement aux locuteurs de parole spontanée. Cependant, selon certains auteurs (e.g. Bolinger 1986 : 110-127, Beyssade et al 2009) la fonction de la désaccentuation n'est pas de marquer l'ancienneté

d'une information, mais une fonction plus large qui englobe les fonctions emphatiques que nous avons étudiées dans cette thèse. L'étude que nous projetons de faire sur la désaccentuation en parole interprétée est donc pertinente par rapport à la question très actuelle (e.g. Marandin 2005, Trotzke 2017) des relations entre la structure informationnelle de l'énoncé et la dimension expressive du langage.

Bibliographie

- Ammann, H. (1928). *Die Menschliche Rede : Sprachphilosophische Untersuchungen*. Lahr: Schauenburg.
- Astésano, C. (2001). *Rythme et accentuation en français : invariance et variabilité stylistique*. Paris: L'Harmattan.
- Audibert, N., Aubergé, V., & Rilliard, A. (2008). How we are not equally competent for discriminating acted from spontaneous expressive speech. *Proceedings of Speech Prosody 2008*, (pp. 693-696). Campinas, Brésil.
- Audibert, N., Aubergé, V., & Rilliard, A. (2010). Prosodic correlates of acted vs. spontaneous discrimination of expressive speech: a pilot study. *Proceedings of Speech Prosody 2010*, (pp. 640-643). Chicago, Etats-Unis.
- Avanzi, M. (2012). *L'interface prosodie/syntaxe en français. Dislocations, incises et asyndètes*. Bruxelles: Peter Lang.
- Avanzi, M., Lacheret-Dujour, A., & Victorri, B. (2008). ANALOR: a tool for semi-automatic annotation of french prosodic structure. *Proceedings of Speech Prosody 2008*, (pp. 119-122). Campinas, Brésil.
- Avanzi, M., Obin, N., Lacheret-Dujour, A., & Victorri, B. (2011). Toward a continuous modeling of French prosodic structure: using acoustic features to predict prominence location and prominence degree. *Proceedings of Interspeech 2011*, (pp. 2033-2036). Florence, Italie.
- Avanzi, M., Simon, A.-C., Goldman, J.-P., & Auchlin, A. (2010). C-PROM : un corpus de français parlé annoté pour l'étude des proéminences. *Actes des XVIIIes Journées d'Etude sur la Parole (JEP)*, (pp. 73-76). Mons, Belgique.
- Bally, C. (1944). *Linguistique générale et linguistique française* (éd. 2e). Paris: Ernest Leroux.
- Bänziger, T., & Klaus, S. R. (2005). The role of intonation in emotional expressions. *Speech Communication*, 46(3-4), pp. 252-267.
- Bates, D., Maechler, M., & Bolker, B. (2011). Linear mixed-effects models using s4 classes (R package version 1.1-10). Récupéré sur <http://cran.R-project.org/package=lme4>

- Batliner, A., Kompe, R., Kießling, A., Nöth, E., & Niemann, H. (1995). Can you tell apart spontaneous and read speech if you just look at prosody? Dans A. Ayuso, & J. Soler (Éds.), *Speech Recognition and Coding. NATO ASI Series (Series F: Computer and Systems Sciences)* (Vol. 147). Berlin, Heidelberg: Springer.
- Beckman, M. E. (1997). A typology of spontaneous speech. Dans Y. Sagisaka, N. Campbell, & N. Higuchi (Éds.), *Computing Prosody*. New York: Springer.
- Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.
- Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., & Rauzy, S. (2008). Le CID – Corpus of Interactional Data – annotation et exploitation multimodale de parole conversationnelle. *Traitement Automatique des Langues*, 49(3), pp. 105-134.
- Beyssade, C., Delais-Roussarie, E., Doetjes, J., Marandin, J.-M., & Rialland, A. (2004). Prosody and information in French. Dans F. Corblin, & H. de Swart (Éds.), *Handbook of French semantics* (pp. 477-500). Stanford: CSLI Publications.
- Beyssade, C., Hemforth, B., Marandin, J.-M., & Portes, C. (2009). Prosodic marking of information focus in French. *Proceedings of LPST09 (Linguistic and Psycholinguistic approaches to Text Structuring 2009)*, (pp. 109-122). Paris, France.
- Beyssade, C., Marandin, J.-M., & Portes, C. (2008). L'association avec le focus en question : « seulement » et son associé. *Actes des XXVIIes Journées d'Etudes sur la Parole (JEP)*, (pp. 195-203). Avignon, France.
- Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, 14(4), 359-375.
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5(9-10), 341-345.
- Bolinger, D. (1985). *Intonation and its parts : melody in spoken English*. Stanford: Stanford University Press.
- Bolinger, D. (1989). *Intonation and its uses : melody in grammar and discourse*. Stanford: Stanford University Press.
- Boula de Mareüil, P., Rilliard, A., & Allauzen, A. (2011). A diachronic study of initial stress and other prosodic features in the French news announcer style : corpus-based measurements and perceptual experiments. *Language and Speech*, 55(2), 263-293.

- Boulakia, G., & Galazzi, E. (2003). Voix et émotions dans une interprétation italienne filmée du texte de « La Voix humaine » de Jean Cocteau (Anna Magnani dans « Una voce umana » de Roberto Rossellini). Dans *Lingua, cultura e testo. Miscellanea di studi francesi in onore di Sergio Cigada* (Vol. 1, pp. 133-152). Milan: Vita e Pensiero.
- Büring, D. (1998). *The meaning of topic and focus : the 59th street bridge accent*. Oxford: Routledge.
- Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific productions in intonation. Dans S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Éds.), *Individual Differences in Speech Production and Perception*. Berne: Peter Lang.
- Carton, F. (1973). L'accent d'insistance en français contemporain. *Actes du 13ème Congrès International de linguistique romane* (pp. 25-39). Québec: Presses de l'Université Laval.
- Chafe, W. (1974). Language and consciousness. *Language*(50), 111-133.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. Dans C. N. Li (Éd.), *Subject and Topic* (pp. 25-55). New York: Academic Press.
- Chafe, W. (2003). An exploration of the functions of prosody. *Proceedings of Prosody and Pragmatics*. Preston, Angleterre.
- Chomsky, N. (1970). Deep structure, surface structure, and semantic interpretation. Dans R. Jakobson, & S. Kawamoto (Éds.), *Studies in general and oriental linguistics* (pp. 183-216). Tokyo: TEC Corporation for Language Research.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Clark, H. H. (1994). Managing problems in speaking. *Speech Communication - Special issue on Spoken Dialogue*, 15(3-4), 243-250.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. Dans R. O. Freedle (Éd.), *Discourse Production and Comprehension* (pp. 1-40). Norwood: Ablex Publishing Corporation.
- Clements, G. N. (1990). The role of the sonority cycle in core syllabification. Dans J. Kingston, & M. E. Beckman (Éds.), *Papers in Laboratory Phonology* (pp. 283-333). Cambridge: Cambridge University Press.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

- Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, 25(7-9), 1141-1177.
- Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.
- Daneš, F. (1974). Functional sentence perspective and the organization of the text. Dans F. Daneš (Éd.), *Papers on functional sentence perspective* (pp. 106-128). Prague: Academia.
- De Looze, C. (2010). *Analyse et interprétation de l'empan temporel des variations prosodiques en français et en anglais (Thèse de doctorat)*. Université de Provence – Aix-Marseille.
- De Ruiter, L. E. (2015). Information status marking in spontaneous vs. read speech in story-telling tasks – evidence from intonation analysis using GToBI. *Journal of Phonetics*, 48, pp. 29-44.
- Degand, L., & Simon, A.-C. (2009). On identifying basic discourse units in speech: theoretical and empirical issues. *Discours. Revue de linguistique, psycholinguistique et informatique*, 4.
- Delais-Roussarie, E., & Di Cristo, A. (à paraître). L'accentuation. Dans A. Abeillé, D. Godard, & A. Delaveau (Éds.), *La Grande Grammaire du Français*. Arles: Actes Sud.
- Delais-Roussarie, E., & Post, B. (2008). Unités prosodiques et grammaire de l'intonation : vers une nouvelle approche. *Actes des XVIIes Journées d'Etude sur la Parole (JEP)*. Avignon, France.
- Delais-Roussarie, E., & Yoo, H. (2014). Rythme et synthèse de la parole : études comparées des patrons rythmiques de différents genres. *Nouveaux cahiers de linguistique française*, 31, 237-247.
- Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., . . . Yoo, H. (2015). Intonational phonology of French: developing a ToBI system for French. Dans S. Frota, & P. Prieto (Éds.), *Intonation in Romance* (pp. 63-100). Oxford: Oxford University Press.
- Delais-Roussarie, E., Rialland, A., Doetjes, J., & Marandin, J.-M. (2002). The prosody of post-focus sequences in French. Dans *Proceedings of Speech Prosody 2002* (pp. 239–242). Aix-en-Provence, France.

- Delais-Roussarie, E., Yoo, H., & Post, B. (2011). Quand frontières prosodiques et frontières syntaxiques se rencontrent. *Langue française*, 170(2), 29-44.
- Delattre, P. (1966). Les dix intonations de base du français. *The French Review*, 40(1), pp. 1-14.
- Dellwo, V., Leemann, A., & Kolly, M.-J. (2015). The recognition of read and spontaneous speech in local vernacular: the case of Zurich German. *Journal of Phonetics*, 48, pp. 13-28.
- Di Cristo, A. (1999a). Le cadre accentuel du français contemporain : essai de modélisation (première partie). *Langues*, 2(3), pp. 184-205.
- Di Cristo, A. (1999b). Le cadre accentuel du français contemporain : essai de modélisation (seconde partie). *Langues*, 2(4), pp. 258-267.
- Di Cristo, A. (2013). *La prosodie de la parole*. Paris: De Boeck Supérieur.
- Di Cristo, A., & Hirst, D. (1993). Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix*, 15, pp. 9-24.
- Di Cristo, A., & Jankowski, L. (1999). Prosodic organisation and phrasing after focus in French. *Proceedings of ICPhS (International Congress of Phonetic Sciences) 14*, (pp. 1565-1568). Chicago, Etats-Unis.
- Doetjes, J., Rebuschi, G., & Riolland, A. (2005). Cleft sentences. Dans F. Corblin, & H. de Swart (Éds.), *Handbook of French Semantics* (pp. 529-552). Chicago: CSLI Publications.
- Dohen, M., & Løevenbruck, H. (2004). Pre-focal rephrasing, focal enhancement and postfocal deaccentuation in French. *Proceedings of ICSLP (International Conference on Spoken Language Processing) 2004*, (pp. 1313-1316). Jeju, Corée du Sud.
- Duez, D. (1991). *La pause dans la parole de l'homme politique*. Paris: Editions du CNRS.
- Eskenazi, M. (1993). Trends in speaking styles research. *Proceedings of Eurospeech 1993*, (pp. 501-509). Berlin, Allemagne.
- Feldhausen, I., & Delais-Roussarie, E. (2012). La structuration prosodique et les relations syntaxe/prosodie dans le discours politique. *Actes des XIXes Journées d'Etude sur la Parole (JEP)*, (pp. 9-16). Grenoble, France.

- Féry, C. (2001). Focus and phrasing in French. Dans C. Féry, & W. Sternefeld (Éds.), *Audiatu Vox Sapientiae. A Festschrift for Arnim von Stechow (Studia grammatica, vol. 52)* (pp. 153-181). Berlin: Akademie Verlag.
- Firbas, J. (1992). *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), pp. 378-382.
- Fónagy, I. (1964). L'information du style verbal. *Linguistics*, 2(4), pp. 19-47.
- Fónagy, I. (1980). L'accent en français : accent probabilitaire. Dans I. Fónagy, & P. R. Léon (Éds.), *L'accent en français contemporain (Studia Phonetica, vol. 15)* (pp. 123-233). Paris: Didier.
- Fónagy, I. (1983). *La vive voix : essais de psycho-phonétique*. Paris: Payot.
- Fónagy, I., & Bérard, E. (2006). Functions of intonation. Dans Y. Kawaguchi, & T. Moriguchi (Éds.), *Prosody and Syntax: Cross-linguistic perspectives (Usage-based linguistic informatics, vol. 3)* (pp. 19-46). Amsterdam: John Benjamins.
- Frota, S., & Prieto, P. (Éds.). (2015). *Intonation in Romance*. Oxford: Oxford University Press.
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). Various coefficients of interrater reliability and agreement (R package version 0.84). Récupéré sur <https://CRAN.R-project.org/package=irr>
- Gendrot, C., & Raynal, C. (2004). Phénomènes prosodiques de constituants « appartenant au fond » autour de deux constituants distingués : le focus et la portée de 'seulement'. *Actes des Journées d'Etude en Linguistique 2004*, (pp. 205-209). Nantes, France.
- Gendrot, C., Adda-Decker, M., & Schmid, C. (2012). Comparaison de parole journalistique et de parole spontanée : analyses de séquences entre pauses. *Actes des XIXes Journées d'Etude sur la Parole (JEP)*, (pp. 649-656). Grenoble, France.
- Gibson, E., & Fedorenko, E. (2010). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1-2), pp. 1-37.
- Godement-Berline, R. (2014). L'emploi de la focalisation prosodique dans le jeu d'acteur. *Nouveaux Cahiers de Linguistique Française*, 31, pp. 129-139.

- Godement-Berline, R. (2016). Using a replication task to study prosodic highlighting. *Proceedings of Speech Prosody 2016*, (pp. 528-532). Boston, Etats-Unis.
- Goldman Eisler, F. (1968). *Psycholinguistics: experiments in spontaneous speech*. New York: Academic Press.
- Goldman, J.-P. (2011). EasyAlign: an automatic phonetic alignment tool under Praat. *Proceedings of Interspeech 2011*, (pp. 3233-3236). Florence, Italie.
- Goldman, J.-P., Auchlin, A., & Simon, A.-C. (2011). Discrimination de styles de parole par analyse prosodique semi-automatique. *Actes de IDP (Interface Discours Prosodie) 2011*, (pp. 207-221). Paris, France.
- Goldman, J.-P., Pršir, T., Christodoulides, G., Simon, A.-C., & Auchlin, A. (2014). Phonogenre identification: a perceptual experiment with 8 delexicalised speaking styles. *Nouveaux Cahiers de Linguistique Francaise*, 31, pp. 51-62.
- Grice, H. P. (1975). Logic and conversation. (P. Cole, & J. L. Morgan, Éd.) *Syntax and Semantics*, 3, pp. 41-58.
- Gutzmann, D., & Castroviejo Miró, E. (2011). The dimensions of verum. Dans O. Bonami, & P. Cabredo Hofherr (Éds.), *Empirical Issues in syntax and semantics, vol. 8* (pp. 143–165).
- Halliday, M. A. (1967). Notes on transitivity and theme in English (part 2). *Journal of Linguistics*, 3(2), pp. 199-244.
- Hirst, D. (2007). A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation. *Proceedings of ICPhS (International Congress of Phonetic Sciences) 16*, (pp. 1233-1236). Saarbrücken, Allemagne.
- Hockett, C. F. (1958). *A course in modern linguistics*. New York: Macmillan .
- Höhle, T. N. (1992). Über verum-fokus im deutschen. Dans J. Jacobs (Éd.), *Informationsstruktur und Grammatik. Linguistische Berichte Sonderhefte* (pp. 112-141). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, 10(2), pp. 163-169.
- Jackendoff, R. (1974). *Semantic interpretation in generative grammar*. Cambridge: The MIT Press.

- Jacobs, J. (1983). *Fokus und Skalen : Zur Syntax und Semantik der Gradpartikel im Deutschen (Linguistische Arbeiten, vol. 138)*. Tübingen: Niemeyer.
- James, W. (1890). *The principles of psychology*. New York: Henry Holt and Company.
- Joos, M. (1968). The isolation of styles. Dans J. A. Fishman (Éd.), *Readings in the sociology of language* (pp. 185-191). La Haye: Mouton De Gruyter.
- Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, 15(2), pp. 189-226.
- Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. Dans A. Botinis (Éd.), *Intonation: analysis, modeling and technology* (pp. 209-242). Dordrecht: Kluwer Academic Publishers.
- Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of acting experience on emotion expression and recognition in voice: non-actors provide better stimuli than expected. *Journal of Nonverbal Behavior*, 39(3), pp. 195-214.
- Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Psychology*, 2, pp. 180-191.
- Koch, P., & Oesterreicher, W. (2001). Langage parlé et langage écrit. Dans G. Holtus, M. Metzeltin, & C. Schmitt (Éds.), *Lexikon der romanistischen Linguistik, tome 1* (pp. 584-627). Tübingen: Max Niemeyer Verlag.
- Koffka, K. (1922). Perception: an introduction to the Gestalt-theorie. *Psychological Bulletin*, 19, pp. 531-585.
- Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, 34(4), pp. 391-405.
- Krahmer, E., & Swerts, M. (2008). On the role of acting skills for the collection of simulated emotional speech. *Proceedings of Interspeech 2008*, (pp. 261-264). Brisbane, Australie.
- Krifka, M. (1992). A compositional semantics for multiple focus constructions. Dans J. Jacobs (Éd.), *Informationsstruktur und Grammatik (Linguistische Berichte Sonderhefte, vol. 4)* (pp. 17-53). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Krifka, M. (2006). Association with focus phrases. Dans V. Molnár, & S. Winkler (Éds.), *The architecture of focus* (pp. 105-136). Berlin, New York: Mouton De Gruyter.

- Kuno, S. (1987). *Functional syntax : anaphora, discourse and empathy*. Chicago: University of Chicago Press.
- Kuno, S., & Kaburaki, E. (1977). Empathy and syntax. *Linguistic Inquiry*, 8(4), pp. 627-672.
- Laan, G. P. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, 22(1), pp. 43-65.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphie: University of Pennsylvania Press.
- Lacheret, A., Obin, N., Goldman, J.-P., & Avanzi, M. (2008). Comparaison de trois outils de détection automatique de proéminences en français parlé. *Actes des XVIIes Journées d'Etude sur la Parole (JEP)*. Avignon, France.
- Ladd, D. R. (1980). *The structure of intonational meaning: evidence from English*. Bloomington, Londres: Indiana University Press.
- Lambrecht, K. (1994). *Information structure and sentence form : topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), pp. 159-174.
- Laurens, F., Marandin, J.-M., Patin, C., & Yoo, H. (2011). The used and the possible: the use of elicited conversations in the study of prosody. *Actes de IDP (Interface Discours Prosodie) 2011*, (pp. 239-257). Paris, France.
- Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
- Léon, P. R. (1969). Principes et méthodes en phonostylistique. *Langue française*, 3(1), pp. 73-84.
- Léon, P. R. (1993). *Précis de phonostylistique : parole et expressivité*. Paris: Nathan Université.
- Léon, P. R. (2007). *Phonétisme et prononciations du français* (éd. 5e). Paris: Armand Colin.
- Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustical Society of America*, 34, pp. 922-927.

- Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., & Miller, M. (1985). Measures of the sentence intonation of read and spontaneous speech in American English. *Journal of the Acoustical Society of America*, 77(2), pp. 649-657.
- Lucci, V. (1980). L'accent didactique. Dans I. Fónagy, & P. R. Léon (Éds.), *L'accent en français contemporain* (pp. 107-121). Paris: Didier.
- Marandin, J.-M. (2005). Formatage de l'information : focus et contexte. Dans F. Corblin, & C. Gardent (Éds.), *Interpréter en contexte* (pp. 31-79). Paris: Hermès.
- Marouzeau, J. (1923). Accent affectif et accent intellectuel. *Bulletin de la Société de Linguistique de Paris*, 25, pp. 80-86.
- Martin, P. (1975). Analyse phonologique de la phrase française. *Linguistics*, 13(146), pp. 35-68.
- Martin, P. (1987). Prosodic and rhythmic structures in French. *Linguistics*, 25(5), pp. 925-950.
- Martin, P. (2009). *Intonation du français*. Paris: Armand Colin.
- Martin, P. (2015). *The structure of spoken language*. Cambridge: Cambridge University Press.
- Mathesius, V. (1961). *A functional analysis of present-day English on a general linguistic basis*. Prague: Československá Akademie Věd.
- Mathon, C. (2014). Perception des phonostyles et représentativité du phonogène : le cas du commentaire sportif en direct. *Nouveaux cahiers de linguistique française*, 31, pp. 93-103.
- Mertens, P. (1993). Accentuation, intonation et morphosyntaxe. *Travaux de linguistique*, 26, pp. 21-69.
- Mertens, P. (2004). The Prosogram: semi-automatic transcription of prosody based on a tonal perception model. *Proceedings of Speech Prosody 2004*, (pp. 23-26). Nara, Japon.
- Michelas, A. (2011). *Caractérisation phonétique et phonologique du syntagme intermédiaire en français : de la production à la perception (Thèse de doctorat)*. Université de Provence – Aix-Marseille.
- Milner, J.-C., & Régnault, F. (1987). *Dire le vers : court traité à l'intention des acteurs et des amateurs d'alexandrins*. Paris: Editions du Seuil.

- Mixdorff, H., & Pfitzinger, H. R. (2005). Analysing fundamental frequency contours and speech rate in map task dialogs. *Speech Communication*, 46(3-4), pp. 310-325.
- Munro, R., & Tily, H. (2011). The start of the art: an introduction to crowdsourcing technologies for language and cognition studies. *Crowdsourcing Technologies for Language and Cognition Studies*, (pp. 210-266). Boulder, Etats-Unis.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
- Ouyang, I. C., & Kaiser, E. (2015). Individual differences in the prosodic encoding of informativity. Dans S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Éds.), *Individual differences in speech production and perception* (pp. 147-188). Berlin: Peter Lang.
- Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Halle: Niemeyer.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50, pp. 157-175.
- Post, B. (2000). *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.
- Post, B. (2002). French tonal structures. *Proceedings of Speech Prosody 2002*, (pp. 583–586). Aix-en-Provence, France.
- Potts, C. (2007). The expressive dimension. *Theoretical Linguistics*, 33(2), pp. 165-198.
- Prince, E. F. (1981). Toward a taxonomy of given-new information. Dans P. Cole (Éd.), *Radical pragmatics* (pp. 223-255). New York: Academic Press.
- Pršir, T., Goldman, J.-P., & Auchlin, A. (2014). Prosodic features of situational variation across nine speaking styles in French. *Journal of Speech Sciences*, 4(1), pp. 41-60.
- R Core Team. (2018). *R: a language and environment for statistical computing*. Vienne: R Foundation for Statistical Computing. Récupéré sur <https://www.R-project.org>
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1(1), pp. 75–116.
- Rossi, M. (1980). Le français, langue sans accent ? Dans I. Fónagy, & P. R. Léon (Éds.), *L'accent en français contemporain (Studia Phonetica, vol. 15)* (pp. 13-52). Paris: Didier.

- Rossi, M. (1999). *L'intonation : le système du français*. Paris: Ophrys.
- Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, 99(2), pp. 143-165.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), pp. 695-729.
- Scherer, K. R., Ladd, D. R., & Silverman, K. E. (1984). Vocal cues to speaker affect: testing two models. *The Journal of the Acoustical Society of America*, 76(5), pp. 1346-1356.
- Séguinot, A. (1976). L'accent d'insistance en français standard. Dans F. Carton, D. Hirst, A. Marchal, & A. Seguinot (Éds.), *L'accent d'insistance (Studia Phonetica, vol. 12)* (pp. 1-91). Paris: Didier.
- Selkirk, E. (1978). On prosodic structure and its relation to syntactic structure. Dans T. E. Fretheim (Éd.), *Nordic Prosody II* (pp. 111-140). Trondheim: Tapir.
- Selkirk, E. (1984). *Phonology and syntax : the relation between sound and structure*. Cambridge: The MIT Press.
- Selting, M. (1994). Emphatic speech style : with special focus on the prosodic signalling of heightened emotive involvement in conversation. *Journal of pragmatics*, 22(3-4), pp. 375-408.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P. P., & Hirschberg, J. (1992). Tobi : a standard for labeling english prosody. *Proceedings of ICSLP (International Conference on Spoken Language Processing) 1992*, (pp. 867-870). Banff, Canada.
- Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), pp. 257-268.
- Simon, A.-C., Auchlin, A., Avanzi, M., & Goldman, J.-P. (2010). Les phonostyles: une description prosodique des styles de parole en français. Dans M. Abecassis, & G. Ledegen (Éds.), *Les voix des Français : en parlant, en écrivant* (pp. 71-88). Berne: Peter Lang.
- Stalnaker, R. C. (1978). Assertion. Dans P. Cole (Éd.), *Pragmatics (Syntax and semantics, vol. 9)* (pp. 315-332). New York: Academic Press.

- Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. *Proceedings of ICSLP (International Conference on Spoken Language Processing) 1996*, (pp. 1501-1504). Philadelphie, Etats-Unis.
- Touati, P. (2003). Registre et expansion tonals du français. L'usage rhétorique de la voix dans les discours politiques. Dans M. Demers (Éd.), *Registre et voix sociale* (pp. 59-78). Montréal: Éditions Nota bene.
- Trotzke, A. (2017). *The grammar of emphasis : from information structure to the expressive dimension*. Berlin, New York: Mouton De Gruyter.
- Troubetzkoy, N. S. (1939). *Principes de phonologie*. Paris: Klincksieck.
- Turco, G., Dimroth, C., & Braun, B. (2013). Intonational means to mark Verum focus in German and French. *Language and Speech*, 56(4), pp. 461-491.
- Uldall, E. T. (1964). Dimensions of meaning in intonation. Dans D. Abercrombie, D. Butler Fry, P. MacCarthy, N. C. Scott, & J. L. Trim (Éds.), *In honor of Daniel Jones: papers contributed on the occasion of his eightieth birthday* (pp. 271–279). London: Longman.
- Vachek, J. (1966). *The linguistic school of Prague; an introduction to its theory and practice*. Bloomington, Londres: Indiana University Press.
- Vallduví, E., & Vilkuna, M. (1998). On rheme and kontrast. Dans P. Culicover, & L. McNally (Éds.), *The limits of syntax* (pp. 79-108). New-York: Academic Press.
- Vallduví, E., & Zacharski, R. (1998). Accenting phenomena, association with focus, and the recursiveness of focus-ground. *Proceedings of the Ninth Amsterdam Colloquium*, (pp. 683-702). Amsterdam, Pays-Bas.
- von der Gabelentz, G. C. (1869). Ideen zu einer vergleichenden Syntax. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft*, 6, pp. 376-384.
- von Stechow, A. (1990). Focusing and backgrounding operators. Dans W. Abraham (Éd.), *Discourse particles* (pp. 37-84). Amsterdam: John Benjamins.
- Wagner, P., & Windmann, A. (2016). Re-enacted and spontaneous conversational prosody: how different? *Proceedings of Speech Prosody 2016*, (pp. 518-522). Boston, Etats-Unis.
- Wilting, J., Krahmer, E., & Swerts, M. (2006). Real vs. acted emotional speech. *Proceedings of ICSLP (International Conference on Spoken Language Processing) 2006*, (pp. 805–808). Pittsburgh, Etats-Unis.

Wioland, F. (1985). *Les structures syllabiques du français*. Genève: Slatkine.

Wunderli, P. (1987). *L'intonation des séquences extraposées en français*. Tübingen: Gunter Narr Verlag.

Annexe A : Matériel de l'expérience de production

Texte intégral de la production A :

– A propos de mon nom, du coup j'ai une anecdote : quand j'étais... je devais être... ouais, j'étais au lycée... Je jouais au foot, tu vois, à Bormes... Et pour savoir qui jouait... Tu vois, donc t'avais l'entraînement la semaine... Et le samedi, dans le journal, dans la Prov... enfin à l'époque c'était Bormes-Matin... à Bormes-les-Mimosas, tu avais les convocations des joueurs. Donc, « sont convoqués à telle heure pour jouer contre telle équipe... »

– Ah ça paraissait dans le journal ? Tu savais quand tu... ? Non, toi tu le savais d'avance.

– Non, tu savais pas.

– Ah ouais. C'est dans le journal que tu apprenais que tu...

– Si tu veux, tu t'entraînais la semaine, tu t'entraînais par exemple le mardi et le vendredi, tu vois, ou le mercredi et le vendredi... Enfin ça dépendait. Au départ c'était le mercredi, après quand t'étais junior, tout ça, c'était le soir. Et eux, ils faisaient le...

– Ah ouais, ils publiaient le...

– Ouais bah tu sais, dans les villages, de toute façon t'avais pas... A la limite ça les arrangeait les journaux, au moins ils avaient des trucs à mettre ! Y'avait au moins déjà, je sais pas, cinquante personnes qui achetaient le journal pour... enfin bref. Et donc tu achetais le journal samedi, tu savais si tu avais été sélectionné ou pas, tu vois. Et puis... Bon, moi, en général, si tu veux, je me démerdais pas trop mal pour dans un village, donc en général j'avais ma place, tu vois. Je jouais en général tous les matchs.

– Hm-hm.

– Et puis un samedi... j'ouvre... 'tain, j'y suis pas. Alors que j'étais allé à l'entraînement, j'avais pas fait de connerie, je m'étais pas battu... Enfin bref. J'y suis pas. Alors je tire la tronche, tu vois, je me dis : « Merde, pourquoi je suis pas sélectionné », tout ça... Tu sais, comme tu peux être, des fois, con...

– Ouais.

– Et puis mon père il dit : « Ben attends, on va... y'a peut-être une erreur ! Enfin te mets pas dans cet état-là, on va... j'appelle. » Alors il appelle, si tu veux, il appelle un des dirigeants...

– Le maire de Bormes !

– Le maire de Bormes ! C’est le voisin, tu vois, le maire de Bormes. Il appelle un dirigeant, il dit : « Ouais, on a vu le journal et... Adam il est pas convoqué... C’est normal ou pas, ou c’est une... ? » « – Mais si, il est convoqué. Pourquoi ? Il est dans le journal ! » Et puis le dirigeant, il ouvre le journal, tu vois, et puis il dit : « Ah ouais il y est pas. Bon, c’est un oubli. » Et puis il dit : « Mais... qu’est-ce que c’est que cette histoire ? » Et en fait... si tu veux, donc y’avait marqué : « sont convoqués à telle heure »... Et puis tu as donc un joueur, t’as le joueur... Et puis y’en a un : « entre parenthèses, 6h10 ». Alors il dit : « Qu’est-ce que c’est que cette connerie ? 6h10... Pourquoi il est... Il est pas convoqué à 6h10, le match il est à 10h ! » etc. Et en fait, si tu veux...

– Ouais, Ghio...

– Voilà !

– Le G ça fait un 6...

– Quand les dirigeants ils avaient écrit la liste des joueurs, ils avaient filé ça au journaliste, le journaliste il avait vu G-H-I-O, le mec il avait dû un peu fermer le G, tu vois, et ça avait fait 6-H-1-0. Et du coup, lui, il était convoqué à 6h10. Et du coup après, pendant toute mon adolescence, on m’appelait 6h10...

Texte intégral de la production B :

– Donc j’étais à ce cours-là, et c’était assez drôle parce que tous les étudiants se sont précipités au cours de philosophie indienne parce que c’était plutôt original... Et le prof était tellement horrible...

– Ah bon !

– Bah, en plus on lui avait prévu pour son séminaire, on lui avait prévu une – séminaire de M1 – une toute petite salle, ce qui était étonnant parce qu’y avait...

– Hmm, ça attire en fait... Ouais, ils se doutent que...

– Enormément d’étudiants, à Paris 4 en philo, et la salle était toute petite ! Et elle était blindée à craquer le premier jour et j’étais assez étonnée...

– Ils ont pas mesuré, en fait pour eux ils se sont dit on fait un petit truc comme ça...

– Hmm et en fait...

– Mais c’est ce qui attire les étudiants !

– Mais en fait ils avaient raison, parce que au bout de deux semaines y avait plus personne, tellement il était insupportable.

– Ah d’accord, en fait ils savaient très bien ce qu’ils faisaient !

– Bah ouais je pense. Il était immonde, ce prof. Il était vraiment horrible. Et il était passionnant, mais il était...

– Et ses cours étaient quand même biens...

– Ah oui ses cours étaient passionnants.

– Mais c’est lui, humainement, qui était...

– Ah ouais il était terrible.

– Mais pourquoi, qu’est-ce qu’il faisait par exemple ?

– Le premier jour, il nous a distribué une bibliographie avec tous les livres à lire, donc c’était un recto verso, A4, pas de marge, écrit en point 10, sans interlignes, donc déjà tu... fallait réussir à lire, et pas les titres les uns après les autres mais vraiment...

– Ah ouais !

– A l’affilée... Et le premier cours, il commente la bibliographie, et y a une étudiante qui prenait des notes ou qui soulignait, et il lui a hurlé dessus, il lui a arraché la bibliographie, il l’a déchirée

en mille morceaux en disant : « Ecoutez ce que je dis, plutôt que de prendre des notes ! » Il s'est vraiment énervé. Il était comme ça un peu tout le temps. A avoir des réactions super agressives et super violentes. Donc c'était drôle, parce que les étudiants qui venaient pour écouter la philosophie indienne, c'était plutôt des gens qui avaient un peu une image de la philosophie indienne comme quelque chose assez proche de la spiritualité, du yoga, de la méditation, enfin la vision occidentale d'un truc cool et hippie, quoi, et le prof c'était la caricature inverse. Donc c'était assez drôle.

– Mais t'as tenu bon.

– Oui, ben oui oui, nan, c'est...

– T'as pas eu envie de lui arracher la gueule dès le début...

– Moi j'ai fait partie de ceux à qui il a pas arraché un papier donc...

– Ouais.

– Nan, une fois qu'on était pas nombreux, il était beaucoup plus...

– Ah peut-être c'est une stratégie lui aussi pour être un peu plus...

– Ouais, je pense.

– Tranquille ?

– Mais c'est une stratégie que t'as dans... A chaque fois que les gens ils font des études indiennes, j'ai l'impression que ça attire beaucoup de gens un peu branchés « je mange végétarien, je fais du yoga etc. » Parce que j'ai fait du sanskrit, c'était la même chose ! Les deux débuts d'année, t'avais une quinzaine de personnes habillées avec des pantalons un peu genre « je fais du didgeridoo sur la plage, des bolas, et je fais du yoga, et j'essaye d'être végétarien », et les profs, c'est exactement l'inverse, quoi ! Les gens qui étudient le sanskrit, la grammaire du sanskrit, ou les textes védiques, c'est pas du tout des rigolos et des gens cools, quoi... Ils sont hyper carrés, hyper sérieux, et du coup c'est un peu le clash culturel en début d'année à chaque fois en sanskrit, et là en philo indienne c'était la même chose. Et du coup je pense, par réaction, les profs sont encore plus... Je pense que ça les agace, cette caricature... qui est en plus une caricature hyper...

– Occidentale ?

– Ouais. Mais du coup ils deviennent assez horribles très vite. Et les étudiants sont vite dégoûtés.

- Oui, du coup il reste que des gens qui sont intéressés par le fond et pas par toute cette mode...
- Ouais. Hm. Ou qui arrivent à le cacher...
- Comme toi...
- Voilà, discrètement...
- Qui vont pas venir... ils laissent leurs pantalons spéciaux à la maison...
- Ouais. Faudrait que je m'achète un... Oh j'aurais dû essayer. De faire de la provoc et d'y aller avec un bon didgeridoo dans mon sac à dos. Mon djembe. Et pieds nus. Mais du coup oui, je suis allée en Inde l'année après mon mémoire de philo indienne.

Consignes données aux locuteurs de parole lue et interprétée :

Premier enregistrement (parole lue)

Le texte qui vous a été donné provient d'une conversation réelle entre deux amis.

Nous vous demandons de jouer le rôle du personnage de M.

Votre tâche est de lire le texte à voix haute, comme si vous étiez le personnage et que vous participiez réellement à la conversation.

Vous pouvez lire le texte pour vous-même avant de commencer l'enregistrement.

Veuillez ne pas lire à voix haute les répliques du personnage de G et simplement les marquer par un temps de silence.

Veuillez respecter le plus possible les mots exacts du texte, y compris les hésitations, répétitions etc. Si vous faites une erreur, vous pouvez vous reprendre (veuillez répéter la phrase depuis le début).

Second enregistrement (parole interprétée)

Nous vous avons demandé d'apprendre par cœur le rôle du personnage de M.

Votre tâche est de jouer le texte de mémoire. Nous vous demandons d'enregistrer deux versions du texte :

1. une « italienne » : veuillez simplement restituer le texte, sans volonté d'interprétation
2. votre interprétation : veuillez donner votre version personnelle du texte, en y ajoutant spontanéité, expressivité etc.

Veuillez ne pas jouer les répliques du personnage de G et simplement les marquer par un temps de silence.

Veuillez respecter le plus possible les mots exacts du texte, y compris les hésitations, répétitions etc. Si vous faites une erreur, vous pouvez vous reprendre (veuillez répéter la phrase depuis le début).

Formulaire de consentement rempli par les locuteurs de parole lue et interprétée :

Je soussigné, M./Mme/Melle personne majeure,
domicilié(e) à, né(e) à..... en
..... (année),

déclare avoir été informé(e) de la finalité des données enregistrées fournies à **Rémi Godement-Berline** (Laboratoire de Linguistique Formelle, UMR 7110, CNRS et Université Paris-Diderot).

Je manifeste mon accord :

- 1) pour que ces données soient traitées dans le cadre de la thèse de Rémi Godement-Berline et de ses publications scientifiques ultérieures, conformément à la loi *Informatique et Liberté*, et à la directive européenne 95/46/CE.
- 2) pour que ces enregistrements soient transcrits, analysés et utilisés à des fins scientifiques et pédagogiques en préservant mon anonymat ;
- 3) qu'ils soient diffusés ou représentés en préservant mon anonymat.

J'autorise donc le Laboratoire de Linguistique Formelle, par le biais de ses membres chercheurs et ingénieurs, à reproduire et publier par quelque moyen les enregistrements auxquels j'ai participé, ceci à des fins purement académiques et scientifiques et sans but commercial. J'autorise également la diffusion de ces données à d'autres institutions de recherche afin de mettre en œuvre des projets de recherche dans les Sciences du Langage, à condition que l'Université ou l'Institution qui en bénéficient n'en profitent aucunement pour faire des activités commerciales.

Consignes données aux experts en prosodie :

Veillez écouter les enregistrements sans utiliser de logiciel d'analyse prosodique (e.g. Praat, WinPitch), ou alors en cachant le spectrogramme, la courbe de F0 et la courbe d'intensité.

Relevé auditif des soulignements prosodiques

Veillez relever, dans les répliques du personnage de M uniquement, les soulignements prosodiques réalisés au moyen des traits suivants, isolément ou en combinaison :

- augmentation de la hauteur sur une partie ou la totalité du constituant souligné
- augmentation de la durée syllabique sur une partie ou la totalité du constituant souligné
- augmentation de l'intensité sur une partie ou la totalité du constituant souligné
- accent initial sur le constituant souligné
- contour terminal à la fin du constituant souligné
- compression du registre intonatif sur le(s) constituant(s) à droite du constituant souligné
- compression du registre intonatif sur le(s) constituant(s) à gauche du constituant souligné

NB. Si vous jugez que le soulignement est purement rythmique (accent servant uniquement à marquer une frontière de groupe accentuel ou intonatif), veuillez ne pas le considérer.

Classification fonctionnelle des soulignements prosodiques

Veillez associer à chaque soulignement relevé une ou plusieurs des fonctions suivantes :

- insistance
- expressivité
- focus informationnel étroit ou large (préciser)
- focus contrastif
- focus associatif
- verum focus
- autre fonction (préciser)

Veillez vous référer aux définitions suivantes :

- insistance : le locuteur souligne, pour l'interlocuteur, l'importance d'une partie ou de la totalité du contenu de son énoncé ;
- expressivité : le locuteur exprime une émotion ou une attitude par rapport à une partie ou la totalité du contenu de son énoncé (ex : colère, mépris, joie, admiration...) ;
- focus informationnel : le locuteur distingue ce qu'il affirme spécifiquement, c'est-à-dire l'information qu'il présente à l'interlocuteur comme nouvelle dans la conversation en cours (la totalité du contenu dans le cas d'un focus informationnel large, une partie seulement du contenu dans le cas d'un focus informationnel étroit) ;
- focus contrastif : le locuteur distingue un élément appartenant à un ensemble d'alternatives présent dans le contexte de discours immédiat ;
- focus associatif : le locuteur distingue l'argument d'un adverbe/quantifieur tel que *seulement* ou *aussi* (ex : *Paul a seulement vu BERNADETTE à Paris* interprété comme « Paul a vu Bernadette, et personne d'autre, à Paris ») ;
- verum focus : le locuteur insiste sur la vérité de la proposition (ex : *Paul DID order soup* interprété comme « Il est bel et bien vrai que Paul a commandé de la soupe »).

Annexe B : Autres résultats des études sur corpus réduit

Nous présentons ci-dessous les résultats similaires aux résultats obtenus dans les deux études de production menées sur le corpus entier (cf chap. 2, § 5 et chap. 3, § 4.3).

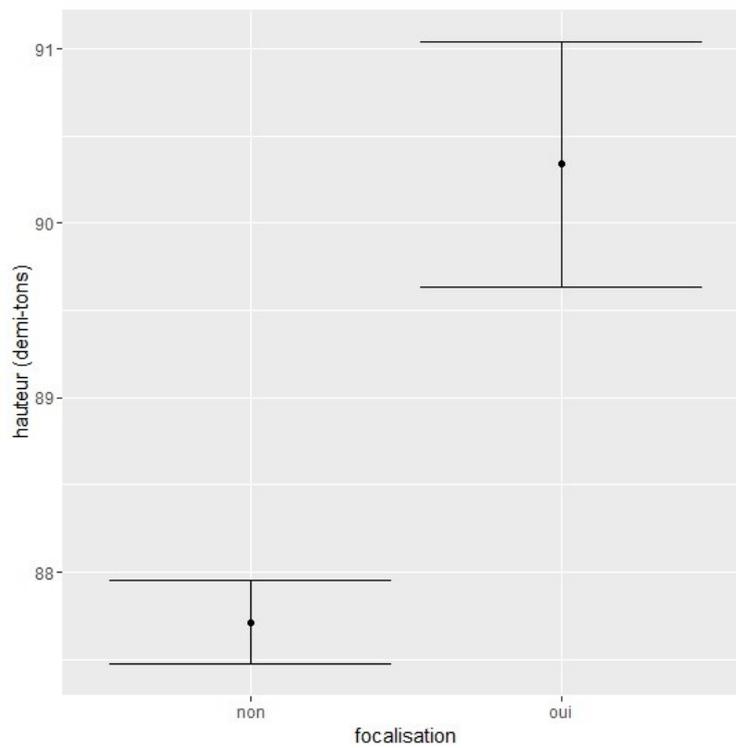
Fréquence d'occurrence de la focalisation prosodique :

	parole spontanée	parole interprétée	corpus réduit
pourcentage de syllabes focalisées	6,59%	16,39%	11,59%

Pourcentage de syllabes focalisées pour chaque phonogène et dans le corpus réduit.

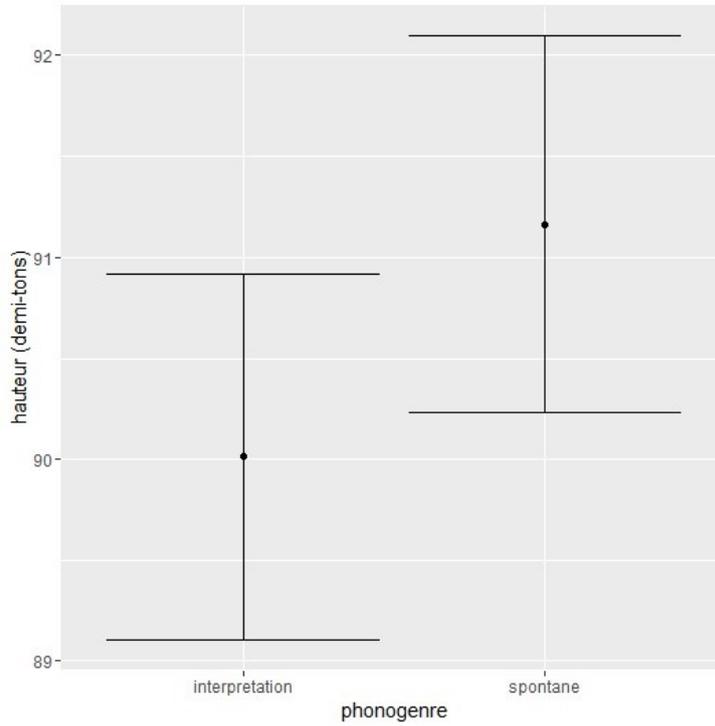
L'effet du phonogène sur la présence de focalisation est significatif ($\chi^2(1) = 7,1$, $p < 0,01$). La différence entre parole spontanée et parole lue est significative ($z = 4,663$, $p < 0,001$).

Hauteur des syllabes focalisées :



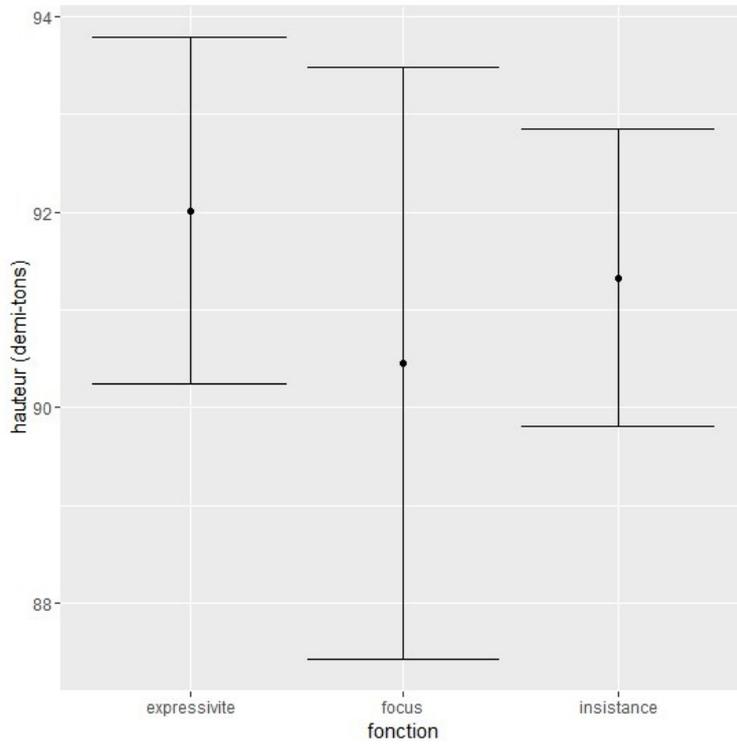
Hauteur moyenne des syllabes focalisées (« oui ») et des autres syllabes du corpus (« non ») dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

L'effet de la présence de focalisation sur la hauteur est significatif ($\chi^2(1) = 234,1, p < 0,001$). La différence de hauteur entre les syllabes focalisées et les syllabes non-focalisées est bien significative ($t = 15,62, p < 0,001$).



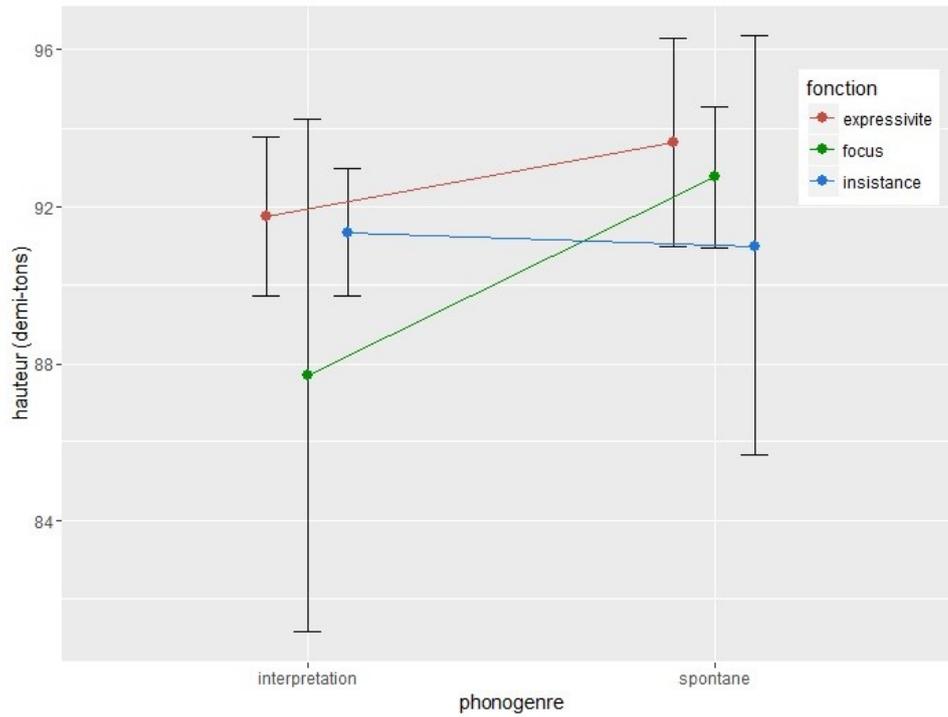
Hauteur moyenne des syllabes focalisées en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané ») dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

L'effet du phonogène sur la hauteur des syllabes focalisées n'est pas significatif ($\chi^2(1) = 0,4, p > 0,5$).



Hauteur moyenne des syllabes focalisées pour chaque fonction dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

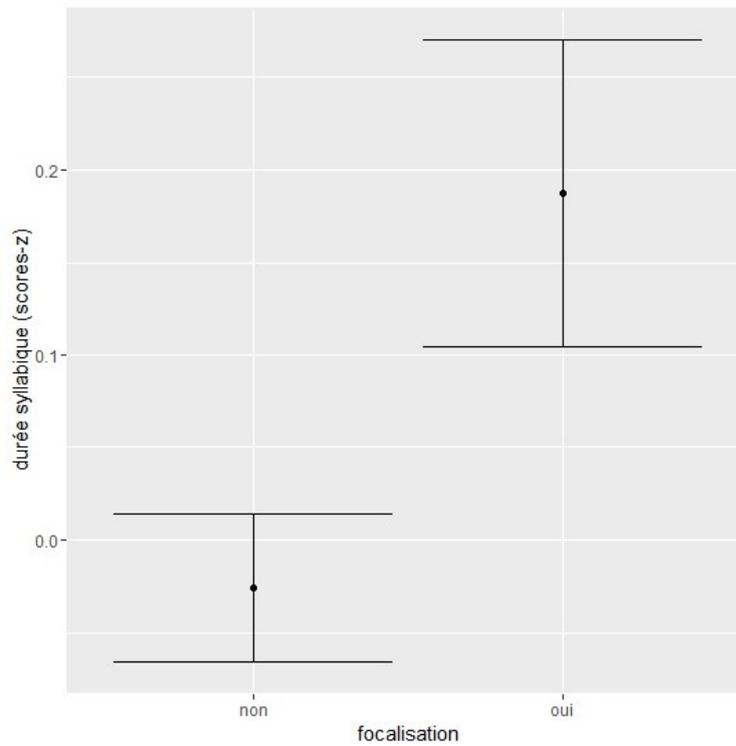
L'effet de la fonction sur la hauteur n'est pas significatif ($\chi^2(2) = 1,6$ $p > 0,1$), de même que l'effet du phonogène sur les différences de hauteur entre fonctions ($\chi^2(1) = 0,6$ $p > 0,1$).



Hauteur moyenne des syllabes focalisées selon la fonction et le phonogenre dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

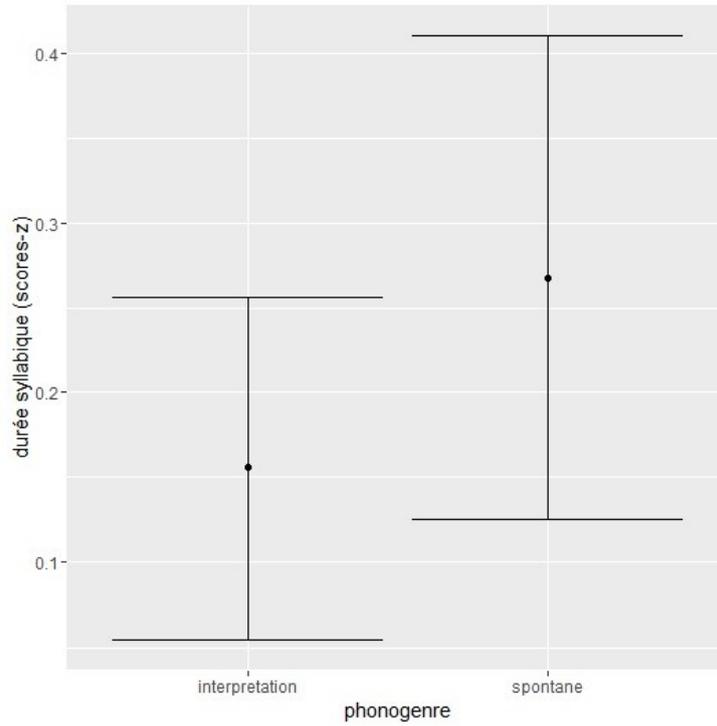
L'effet de l'interaction entre fonction et phonogenre sur la hauteur n'est pas significatif ($\chi^2(2) = 0,5$ $p > 0,1$).

Durée des syllabes focalisées :



Durée moyenne des syllabes focalisées (« oui ») et des autres syllabes du corpus (« non ») dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

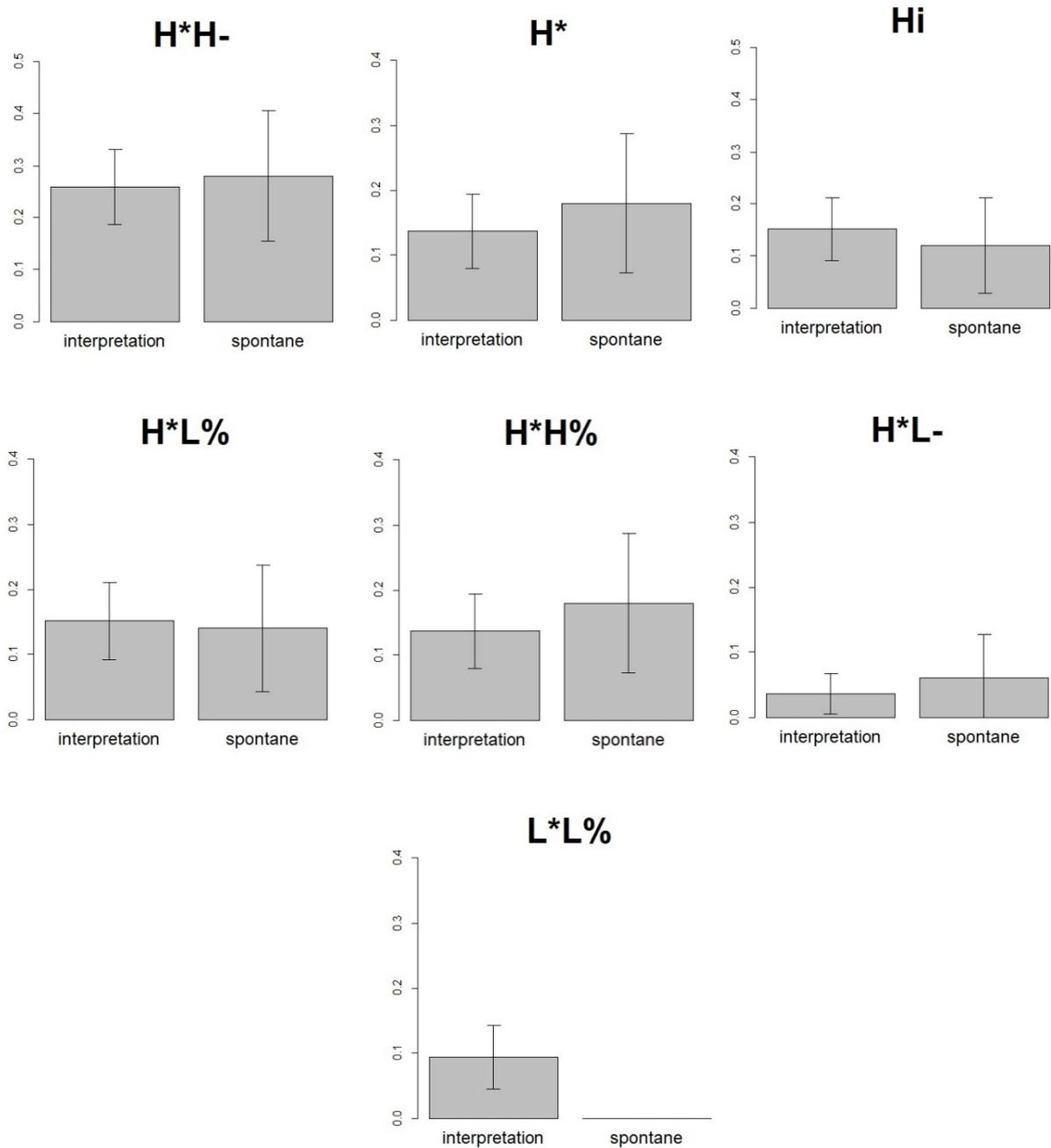
L'effet de la présence de focalisation sur la durée est significatif ($\chi^2(1) = 13,9, p < 0,001$). La différence de durée entre les syllabes focalisées et les syllabes non-focalisées est bien significative ($t = 3,727, p < 0,001$).



Durée moyenne des syllabes focalisées en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané ») dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

L'effet du phonogène sur la durée des syllabes focalisées n'est pas significatif ($\chi^2(1) = 1,5, p > 0,1$).

Contour intonatif des constituants focalisés :



Pourcentages d'occurrence des différents contours intonatifs en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané ») dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

L'effet du phonogène sur le type de contour intonatif n'est pas significatif ($\chi^2(9) = 7,8, p > 0,5$). Il n'y a pas non plus d'effet significatif du phonogène sur le type de frontière prosodique ($\chi^2(2) = 1,0, p > 0,5$).

	résultats de la régression logistique
contour H*H-	$\chi^2(1) = 0,002, p > 0,1$
contour H*	$\chi^2(1) = 0,0001, p > 0,1$
contour Hi	$\chi^2(1) = 0,6, p > 0,1$
contour H*L%	$\chi^2(1) = 0,5, p > 0,1$
contour H*H%	$\chi^2(1) = 0,2, p > 0,1$
contour H*L-	$\chi^2(1) = 0,3, p > 0,1$
contour L*L%	$\chi^2(1) = 0,4, p > 0,1$
groupe accentuel	$\chi^2(1) = 0,7, p > 0,1$
groupe intonatif	$\chi^2(1) = 0,7, p > 0,1$
groupe intermédiaire	$\chi^2(1) = 0,03, p > 0,1$

Résultats des régressions logistiques concernant l'effet de la fonction et du phonogène sur le type de contour intonatif et de frontière prosodique des constituants focalisés du corpus réduit.

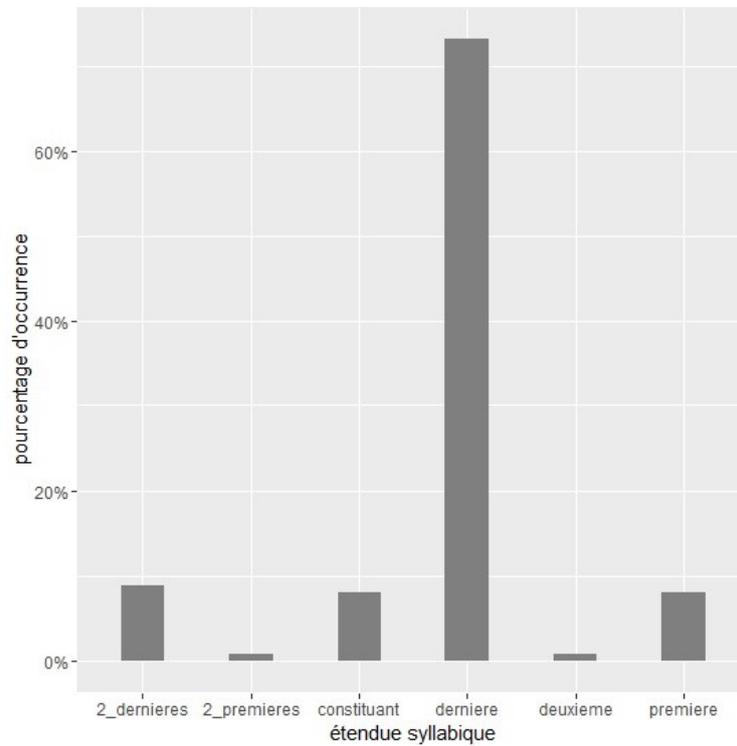
L'effet du phonogène sur les différences de type de contour intonatif et de frontière prosodique entre les fonctions n'est significatif pour aucune des catégories concernées.

	résultats de la régression logistique
contour H*H-	$\chi^2(2) = 6,2, p = 0,05$
contour H*	$\chi^2(2) = 2,3, p > 0,1$
contour Hi	$\chi^2(2) = 4,2, p > 0,1$
contour H*L%	$\chi^2(2) = 5,2, p > 0,05$
contour H*H%	-
contour H*L-	-
contour L*L%	-
groupe accentuel	$\chi^2(2) = 4,2, p > 0,1$
groupe intonatif	$\chi^2(2) = 1,9, p > 0,1$
groupe intermédiaire	-

Résultats des régressions logistiques concernant l'effet de l'interaction entre fonction et phonogène sur le type de contour intonatif et de frontière prosodique des constituants focalisés du corpus réduit.

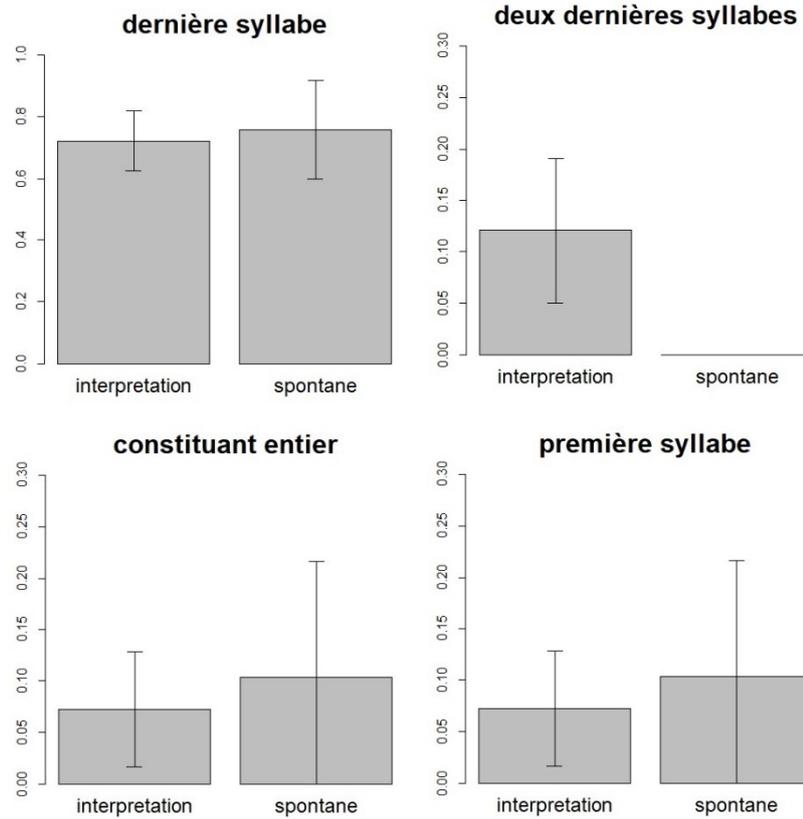
L'effet de l'interaction entre phonogène et fonction sur le type de contour intonatif et de frontière prosodique n'est significatif pour aucune des catégories concernées (pour certaines catégories, le modèle ne révèle aucun résultat).

Etendue syllabique du contour intonatif des constituants focalisés :



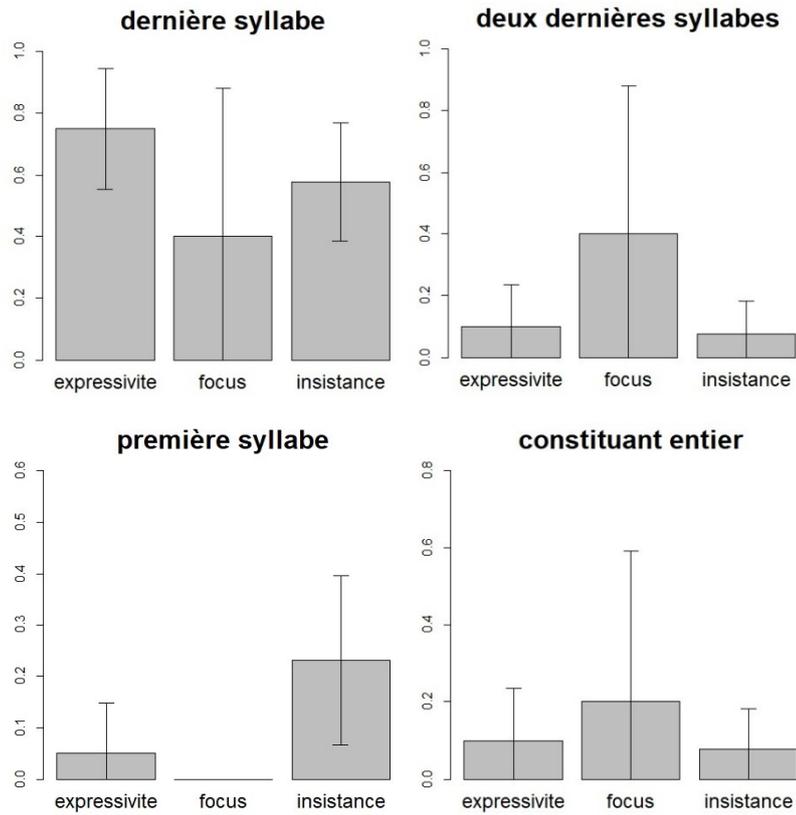
Pourcentage d'occurrence des différents types d'étendue syllabique du contour intonatif sur les constituants focalisés (« 2_dernières » : deux dernières syllabes, « 2_premières » : deux premières syllabes, « constituant » : constituant entier, « dernière » : dernière syllabe, « deuxième » : deuxième syllabe, « première » : première syllabe) dans le corpus réduit. Seuls les constituants de plusieurs syllabes sont concernés.

Comme dans le corpus entier, pour la grande majorité (73,2 %) des constituants focalisés polysyllabique, le contour intonatif est porté par la dernière syllabe.



Pourcentage d'occurrence des différentes étendues syllabiques du contour intonatif en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané ») dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

L'effet du phonogène sur le type d'étendue syllabique n'est pas significatif ($\chi^2(5) = 7,3$, $p = 0,1$).



Pourcentages d'occurrence des différents types d'étendue syllabique du contour intonatif pour chaque fonction dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

L'effet de la fonction sur le type d'étendue syllabique du contour intonatif n'est pas significatif ($\chi^2(8) = 9,9, p > 0,1$).

	résultats de la régression logistique
dernière syllabe	$\chi^2(1) = 0,1, p > 0,1$
deux dernières syllabes	-
constituant entier	$\chi^2(1) = 1,5, p > 0,1$
première syllabe	$\chi^2(1) = 0,8, p > 0,1$

Résultats des régressions logistiques concernant l'effet de la fonction et du phonogène sur le pourcentage d'occurrences des différents types d'étendue syllabique du contour intonatif dans le corpus réduit.

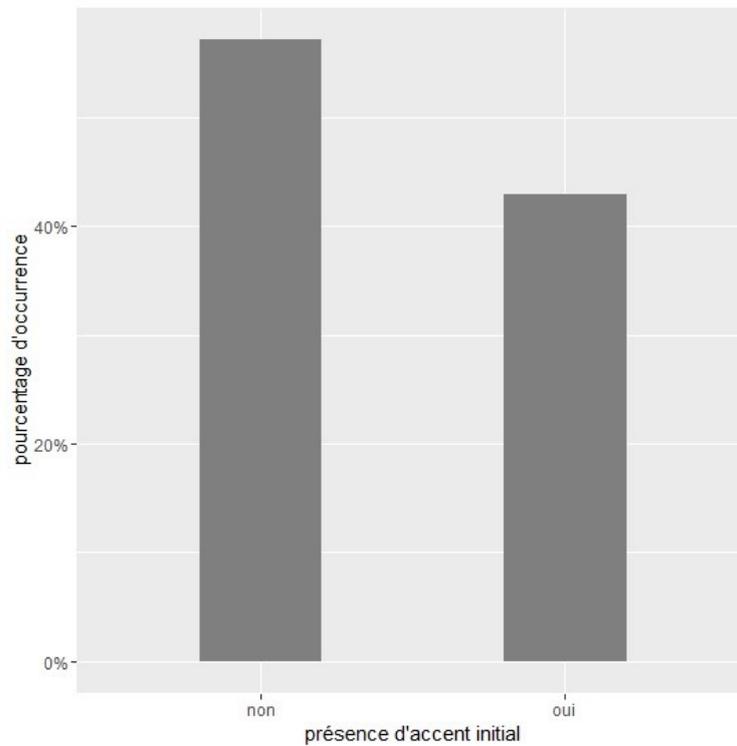
L'effet du phonogène sur les différences de type d'étendue syllabique du contour intonatif entre les fonctions n'est significatif pour aucun type d'étendue (pour un des types d'étendue, le modèle ne révèle aucun résultat).

	résultats de la régression logistique
dernière syllabe	-
deux dernières syllabes	-
constituant entier	-
première syllabe	-

Résultats des régressions logistiques concernant l'effet de l'interaction entre fonction et phonogène sur le pourcentage d'occurrences des différents types d'étendue syllabique du contour intonatif dans le corpus réduit.

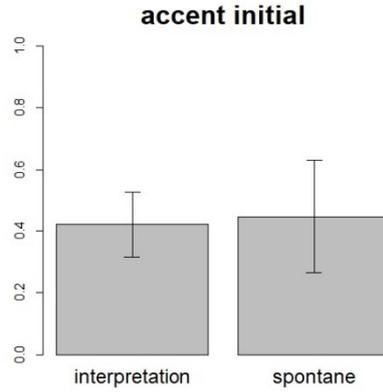
Les régressions logistiques concernant l'effet de l'interaction entre phonogène et fonction sur le type d'étendue syllabique du contour intonatif ne révèlent aucun résultat.

Accentuation initiale sur les constituants focalisés :



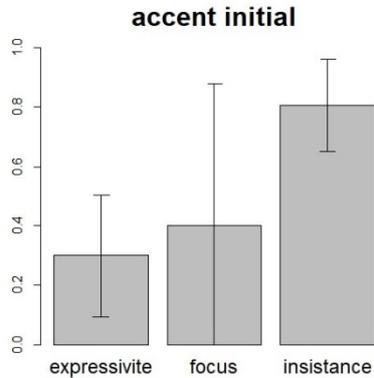
Pourcentage d'occurrence des constituants focalisés portant un accent initial (« oui ») et ne portant pas d'accent initial (« non ») dans le corpus réduit. Seuls les constituants de plusieurs syllabes sont concernés.

Comme dans le corpus entier, une proportion importante bien que minoritaire (42,9 %) de constituants focalisés polysyllabiques comporte un accent initial.



Pourcentage d'occurrence de la présence d'accent initial sur le constituant focalisé en parole interprétée (« interprétation »), en parole lue (« lecture ») et en parole spontanée (« spontané ») dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

L'effet du phonogène sur la présence d'accent initial n'est pas significatif ($\chi^2(1) = 0,1$ $p > 0,1$).



Pourcentage d'occurrence de la présence d'accent initial sur le constituant focalisé pour chaque fonction dans le corpus réduit. Les barres d'erreur indiquent les intervalles de confiance à 95%.

L'effet de la fonction sur la présence d'accent initial est significatif ($\chi^2(2) = 13,1$ $p < 0,01$). La différence entre l'insistance et l'expressivité est significative ($z = 3,275$, $p = 0,00106$). La différence entre l'insistance et le marquage de focus est marginale ($z = 1,770$, $p = 0,07668$).

La différence entre le marquage de focus et l'expressivité n'est pas significative ($z = 0,427$, $p = 0,6695$).

L'effet du phonogène sur les différences de présence d'accent initial entre les fonctions n'est pas significatif ($\chi^2(1) = 1,3$, $p > 0,1$), de même que l'effet de l'interaction entre phonogène et fonction sur la présence d'accent initial ($\chi^2(2) = 1,8$, $p > 0,1$).

Annexe C : Matériel de l'expérience de perception

Enoncés de l'expérience :

Insistance :

- (1) Bah, en plus on lui avait prévu pour son séminaire, on lui avait prévu une – séminaire de M1 – une toute petite salle, ce qui était étonnant parce qu'y avait... énormément d'étudiants, à Paris 4 en philo, et la salle était toute petite ! Et elle était BLINDEE à craquer.
- (2) Donc j'étais à ce cours-là, et c'était assez drôle parce que tous les étudiants se sont précipités au cours de philosophie indienne parce que c'était plutôt original... Et le prof était TELLEMENT horrible...
- (3) Donc c'était drôle, parce que les étudiants qui venaient pour écouter la philosophie indienne, c'était plutôt des gens qui avaient un peu une image de la philosophie indienne comme quelque chose assez proche de la spiritualité, du yoga, de la méditation, enfin la vision occidentale d'un truc cool et hippie, quoi. Et le prof c'était la CARICATURE inverse.
- (4) Les gens qui étudient le sanskrit, la grammaire du sanskrit, ou les textes védiques, c'est pas du tout des rigolos et des gens cools, quoi... Ils sont HYPER carrés.
- (5) Et le samedi, dans le journal, dans la Prov... enfin à l'époque c'était Bormes-Matin... à Bormes-les-Mimosas, tu avais les convocations des joueurs. Donc « sont CONVOQUES à telle heure »...
- (6) Et donc tu achetais le journal samedi, tu savais si tu avais été sélectionné ou pas, tu vois. Et puis... Bon, moi, en général, si tu veux, je me démerdais pas trop mal pour dans un village. Donc en GENERAL j'avais ma place.
- (7) Les deux débuts d'année, t'avais une quinzaine de personnes habillées avec des pantalons un peu genre « je fais du didgeridoo sur la plage, des bolas, et je fais du yoga, et j'essaye d'être végétarien »... Et les profs, c'est EXACTEMENT l'inverse !

(8) Et le premier cours, il commente la bibliographie, et y a une étudiante qui prenait des notes ou qui soulignait, et il lui a hurlé dessus, il lui a arraché la bibliographie, il l'a déchirée en mille morceaux en disant : « Ecoutez ce que je dis, plutôt que de prendre des notes ! » Il s'est vraiment énervé. Il était comme ça un peu tout le temps. A avoir des réactions SUPER agressives.

(9) Ils sont hyper carrés, hyper sérieux, et du coup c'est un peu le clash culturel en début d'année à chaque fois en sanskrit, et là en philo indienne c'était la même chose. Et du coup je pense, par réaction, les profs sont encore plus... Je pense que ça les AGACE.

(10) Et puis un samedi... j'ouvre... 'tain, j'y suis pas. Alors que j'étais allé à l'entraînement, j'avais pas fait de connerie, je m'étais pas battu... Enfin bref. J'y suis pas. Alors je tire la tronche, tu vois, je me dis : « Merde, pourquoi je suis pas sélectionné », tout ça... Tu sais, comme tu peux être, des fois, con... Et puis mon père il dit « Ben ATTENDS »...

Expressivité :

(1) Alors je tire la tronche, tu vois, je me dis : « Merde, pourquoi je suis pas sélectionné », tout ça... Tu sais, comme tu peux être, des fois, con... Et puis mon père il dit : « Ben attends, on va... y'a peut-être une erreur ! Enfin te mets pas DANS C'T ETAT-LA.

(2) Et le premier cours, il commente la bibliographie, et y a une étudiante qui prenait des notes ou qui soulignait, et il lui a hurlé dessus, il lui a arraché la bibliographie, il l'a déchirée en mille morceaux en disant : « ECOUTEZ ce que je dis ! »

(3) Donc c'était drôle, parce que les étudiants qui venaient pour écouter la philosophie indienne, c'était plutôt des gens qui avaient un peu une image de la philosophie indienne comme quelque chose assez proche de la spiritualité... Du YOGA...

(4) Donc c'était drôle, parce que les étudiants qui venaient pour écouter la philosophie indienne, c'était plutôt des gens qui avaient un peu une image de la philosophie indienne comme quelque chose assez proche de la spiritualité, du yoga... De la MEDITATION...

- (5) Les deux débuts d'année, t'avais une quinzaine de personnes habillées avec des pantalons un peu genre « je fais du didgeridoo sur la plage, des bolas, et je fais du yoga, et j'essaye d'être végétarien », et les profs, c'est exactement l'inverse, quoi ! C'est pas du tout des RIGOLOS.
- (6) Oh j'aurais dû essayer. De faire de la provoc et d'y aller avec un bon didgeridoo dans mon sac à dos. Mon DJEMBE.
- (7) Le premier jour, il nous a distribué une bibliographie avec tous les livres à lire, donc c'était un recto verso, A4, pas de marge... ECRIT en point 10.
- (8) Et le premier cours, il commente la bibliographie, et y a une étudiante qui prenait des notes ou qui soulignait... Et il lui a HURLE dessus.
- (9) Il appelle un dirigeant, il dit : « Ouais, on a vu le journal et... Adam il est pas convoqué... C'est normal ou pas, ou c'est une... ? » « – Mais si, il est convoqué. POURQUOI ? »
- (10) Il appelle un dirigeant, il dit : « Ouais, on a vu le journal et... Adam il est pas CONVOQUE.

Marquage de focus :

- (1) A propos de mon nom, du coup j'ai une anecdote. Quand j'étais... Je devais être... Ouais, j'étais au LYCEE.
- (2) Si tu veux, tu t'entraînais la semaine, tu t'entraînais par exemple le mardi et le vendredi, tu vois, ou le mercredi et le vendredi... Enfin ça dépendait. Au départ c'était le MERCREDI.
- (3) Ouais bah tu sais, dans les villages, de toute façon t'avais pas... A la limite ça les ARRANGEAIT les journaux.

- (4) Et puis un samedi... j'ouvre... Putain, j'y suis pas. Alors que j'étais allé à l'entraînement, j'avais pas fait de connerie... Je m'étais pas BATTU...
- (5) A propos de mon nom, du coup j'ai une anecdote. Quand j'étais... Je devais être... Ouais, j'étais au lycée... Je jouais au foot, tu vois, à Bormes... Et pour savoir qui jouait... Tu vois, t'avais l'entraînement la SEMAINE.
- (6) Si tu veux, tu t'entraînais la semaine, tu t'entraînais par exemple le mardi et le vendredi, tu vois. Ou le mercredi et le VENDREDI.
- (7) Ouais bah tu sais, dans les villages, de toute façon t'avais pas... A la limite ça les arrangeait les journaux, au moins ils avaient des trucs à mettre ! Y'avait au moins déjà, je sais pas... Cinquante PERSONNES qui achetaient le journal.
- (8) Et puis mon père il dit : « Ben attends, on va... y'a peut-être une erreur ! Enfin te mets pas dans cet état-là, on va... j'appelle. » Alors il appelle, si tu veux, il appelle un des dirigeants... Le maire de Bormes ! C'est le VOISIN, le maire de Bormes.
- (9) Ouais bah tu sais, dans les villages, de toute façon t'avais pas... A la limite ça les arrangeait les journaux, au moins ils avaient des trucs à mettre ! Y'avait au moins déjà, je sais pas, cinquante personnes qui achetaient le journal pour... enfin bref. Tu achetais le journal SAMEDI.
- (10) Il était immonde, ce prof. Il était vraiment horrible. Et il était PASSIONNANT.

Items de contrôle :

- (1) Et le premier cours, il commente la bibliographie, et y a une étudiante qui prenait des notes ou qui soulignait, et il lui a hurlé dessus, il lui a arraché la bibliographie, il l'a déchirée en mille morceaux en disant : « Ecoutez ce que je dis, plutôt que de prendre des notes ! » Il s'est VRAIMENT énervé.

- (2) Et puis... Bon, moi, en général, si tu veux, je me démerdais pas trop mal pour dans un village, donc en général j'avais ma place, tu vois. Je jouais TOUS LES MATCHS en général.
- (3) Mais en fait ils avaient raison, parce que au bout de deux semaines y avait plus personne, tellement il était insupportable. Il était IMMONDE, ce prof.
- (4) Ils sont hyper carrés, hyper sérieux, et du coup c'est un peu le clash culturel en début d'année à chaque fois en sanskrit. Et là c'était la MÊME CHOSE en philo indienne.
- (5) Et en fait si tu veux, donc y avait marqué : « sont convoqués à telle heure »... Et puis donc tu as un joueur ... Et puis y'en a un où y'a marqué « 6h10 » entre parenthèses. Alors il dit : « Qu'est-ce que c'est que cette CONNERIE ? »

Items d'entraînement :

- (1) Et puis mon père il dit : « Ben attends, on va... y'a peut-être une erreur ! Enfin te mets pas dans cet état-là, on va... j'appelle. » Alors il appelle un des DIRIGEANTS...
- (2) Donc j'étais à ce cours-là, et c'était assez drôle parce que tous les étudiants se sont précipités au cours de philosophie indienne parce que c'était plutôt original... Et le prof était tellement HORRIBLE...
- (3) Et le premier cours, il commente la bibliographie, et y a une étudiante qui prenait des notes ou qui soulignait... Et il lui a HURLE dessus.
- (4) Le premier jour, il nous a distribué une bibliographie avec tous les livres à lire, donc c'était un recto verso, A4, pas de marge, écrit en point 10, sans interlignes, donc déjà fallait réussir à lire, et pas les titres les uns après les autres... Mais vraiment à L'AFFILÉE...
- (5) Il appelle un dirigeant, il dit : « Ouais, on a vu le journal et... Adam il est pas convoqué... » « C'est NORMAL ou pas ? »

Consignes données aux participants :

Bonjour.

Dans cette expérience, vous allez entendre des phrases et vous devrez dire si elles vous semblent prononcées sur le bon « ton » en fonction du contexte dans lequel elles se trouvent. Plus précisément, vous allez d’abord voir apparaître quelques phrases qui forment le contexte qui précède la phrase que vous devez juger. Exemple :

Y a plein de trucs ! Y a des trucs un peu chers, mais y a plein de trucs pas chers où tu peux bien manger, comme dans les cantines chinoises, généralement c’est blindé, y a beaucoup beaucoup de monde, mais c’est vachement bon.

Puis vous allez voir apparaître la phrase que vous devez juger. Exemple :

Après faut faire attention, faut demander pas épicé.

Vous allez entendre cette phrase prononcée sur quatre tons différents. Votre tâche sera de noter chaque prononciation de 0 à 5 en fonction de sa « correspondance » avec le contexte qui précède.

Le but de cette expérience est d'améliorer des logiciels de synthèse de la parole (à destination de systèmes de commande vocale dans des voitures, sur des smartphones, etc.). Les phrases que vous devrez juger ont toutes été produites par un logiciel de synthèse. Par conséquent, aucune d'elle ne sera parfaitement naturelle à l'écoute. Veuillez en tenir compte dans votre évaluation. Veuillez ne pas noter les phrases selon leur degré de "naturalité" mais uniquement selon leur degré de correspondance avec le contexte.

Nous allons commencer par une courte phase d’entraînement avant de passer à l’expérience.

Formulaire de consentement rempli par les participants :

Evaluation de réalisations prosodiques

Le participant entend une série de phrases. Il lui est demandé de juger le « ton » sur lequel est prononcée chaque phrase en fonction du contexte dans lequel elle se trouve.

L'expérience dure approximativement 30 minutes.

Cette expérience n'entre donc pas dans le cadre de la loi Huriot.

Je soussigné(e) comprends la nature des recherches auxquelles je vais participer et comprends que j'ai le droit d'arrêter ma participation à tout moment sans avoir à fournir d'explication. Je comprends que le paiement sera relatif à la participation.

