HAL open science

De l'emphase à la phonologie de la proéminence, pour une approche perceptuelle de la prosodie de corpus

Sophie Herment

To cite this version:

Sophie Herment. De l'emphase à la phonologie de la proéminence, pour une approche perceptuelle de la prosodie de corpus. Linguistique. Université Paris Diderot (Paris 7) Sorbonne Paris Cité, 2013. tel-01626556

HAL Id: tel-01626556

https://hal.science/tel-01626556v1

Submitted on 30 Oct 2017

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.

Sophie Herment
Maître de Conférences
Université d'Aix-Marseille
Laboratoire Parole et Langage, UMR 7309

De l'emphase à la phonologie de la proéminence, pour une approche perceptuelle de la prosodie de corpus

Dossier présenté devant l'Université Paris 7 - Diderot
en vue de l'Habilitation à Diriger des Recherches
le 9 décembre 2013

Volume I : synthèse

Jury
Nicolas Ballier, Professeur à l'Université Paris 7 - Diderot, garant des recherches
Elisabeth Delais-Roussarie, DR CNRS à l'Université Paris 7 - Diderot
Jean-Louis Duchet, Professeur honoraire à l'Université de Poitiers
Jean-Michel Fournier, Professeur à l'Université de Tours
Sylvie Hanote, Professeur à l'Université de Poitiers
Daniel Hirst, DR CNRS honoraire à l'Université d'Aix-Marseille

Remerciements

Je remercie en tout premier lieu Nicolas Ballier, mon garant pour cette habilitation, de m'avoir incitée à m'engager dans cette voie et de m'avoir ensuite accompagnée avec rigueur et gentillesse. Je lui suis reconnaissante pour sa grande disponibilité, ses conseils avisés et la perspicacité de ses commentaires.

J'exprime ma gratitude à Elisabeth Delais-Roussarie, Jean-Louis Duchet, Jean-Michel Fournier, Sylvie Hanote et Daniel Hirst de m'avoir fait l'honneur d'accepter de participer au jury.

Ces pages doivent beaucoup à Laetitia Leonarduzzi et Gabor Turcsan, mes principaux co-auteurs. Leurs encouragements m'ont été précieux.

Je remercie mes autres co-auteurs, tout particulièrement Daniel Hirst, pour la confiance qu'il m'accorde et pour les conseils qu'il continue à promulguer avec toujours autant de bienveillance, et Anne Tortel et Brigitte Bigi pour leurs encouragements.

Ce document n'aurait pas vu le jour sans la richesse que m'apportent mes étudiants. Qu'ils soient ici tous remerciés pour les questions et remises en question, pour les discussions fructueuses et les éclairages indirects.

J'ai aussi bénéficié du soutien précieux de Claire et des encouragements de mes collègues du B106, Laetitia et Gabor, mais aussi Valérie, Cécile, Isabelle et Peter, ainsi que du soutien des collègues du Laboratoire Parole et Langage. Je les en remercie.

Enfin, ce travail n'aurait certainement pas pu être mené à son terme sans l'entourage inestimable de ma famille, à qui j'exprime ma reconnaissance : Amar, pour sa présence apaisante, ses encouragements quotidiens et le réconfort dans les moments de doute, mes parents pour leur soutien sans faille, et Gautier et Elise pour la force qu'ils me transmettent et les moments salutaires de joies partagées.

Table des matières

Table des matières **1**

Table des illustrations **5**

INTRODUCTION **7**

Contextualisation *7*

Prosodie *8*

Intonation *11*

Les écoles britannique et américaine *12*

Les constituants prosodiques *13*

Transcription : contours ou niveaux ? *14*

Positionnement, enseignement et recherche *19*

Structure du document de synthèse *21*

CHAPITRE 1. Des corpus et des outils **23**

1.1. Les corpus oraux *25*

1.1.1. Qu'est-ce qu'un corpus ? 25

1.1.2. Types de corpus 26

1.1.3. Qualité sonore 32

1.1.4. Choix des locuteurs 35

1.1.5. Documentation 37

1.1.6. SLDR, mise en forme et curation des données 40

1.1.7. Proposition de typologie complétée 42

1.2. Transcription et outils *42*

1.2.1. Transcription orthographique 43

1.2.2. SPPAS et PRAAT 46

1.3. Annotation prosodique *50*

1.3.1. Subjectivité de l'annotation et influence de la théorie sous-jacente 50

1.3.2. Système d'annotation et exemple de discordance avec les TSM 51

1.4. Le primat du perceptuel et l'expérience de l'écoute *55*

1.4.1. Annotation automatique sur grands corpus 55

1.4.2. Réglages des seuils de F0 et variations micro-mélodiques 56
1.4.3. Erreurs d'interprétation de la courbe 61
1.4.4. Conclusion : progression pédagogique pour la perception et la catégorisation des contours 64

CHAPITRE 2. La proéminence prosodique, du suprasegmental au segmental..... 67

2.1. Du segmental au suprasegmental : la force de l'inertie de la tradition..... 69
2.2. Du suprasegmental au segmental..... 71
2.2.1. Acquisition et logique de la perception 71
2.2.2. Le repérage des unités intonatives..... 72
2.2.3. Détection des noyaux 75
2.2.4. Saillance et tonique de tête 79
2.2.5. Accent de phrase et accent de mot : quels niveaux de proéminence ? .. 85
2.2.6. Les syllabes inaccentuées et la qualité de la voyelle 88
2.3. Tentative de modèle de hiérarchie des proéminences 92
2.3.1. Niveaux de proéminence 92
2.3.2. Unités rythmiques 94
2.3.3. Niveaux hiérarchiques et modèle 96
2.3.4. Corrélats acoustiques et prosodiques de la proéminence 99
2.3.4.1. F0, durée, intensité et qualité vocalique..... 99
2.3.4.2. Sémantisme et syntaxe 102
2.3.4.3. Pauses..... 103
2.3.4.4. Qualité de voix 104
2.3.4.5. Attaque du mot..... 104
2.3.4.6. Corrélats acoustiques des proéminences : tendances..... 106
2.4. Conclusion : et l'emphase ?..... 107

CHAPITRE 3. De l'emphase à l'interface prosodie/discours : d'une approche perceptuelle-acoustique à une approche discursive de la prosodie et des réalisations syntaxiques..... 109

3.1. Postulats de départ..... 110
3.1.1. La loi du moindre effort 110
3.1.2. La redondance 113
3.1.3. Structures syntaxiques emphatiques..... 115
3.1.4. Théorie de la marque 117

3.1.5. Une alternative : canonique/non canonique 121
3.2. Le discours et la méthode perceptuelle 123
3.2.1. Corpus : le parti-pris syntaxique et le primat de la structure 123
3.2.2. Analyse aurale et approche croisée de la syntaxe et de la prosodie 126
3.2.2.1. Comment la mise en contexte peut changer la perception 126
3.2.2.3. Comment la « mise en son » peut changer l'interprétation de la structure syntaxique 130
3.2.2.4. Approche croisée de la syntaxe et de la prosodie 132
3.3. Réinterprétation en discours et approche de la prosodie discursive 132
3.3.1. Remise en cause de la grammaire de l'écrit 132
3.3.2. Réinterprétation en discours 137
3.3.2.1. Mise en relief non canonique 137
3.3.2.2. Informativité 138
3.3.2.3. Récapitulation (clivées) 140
3.3.2.4. Clarification (dislocation) 141
3.3.2.5. Renforcement de la relation prédicative ou de l'adverbe (do) 141
3.3.2.6. Résumé 142
3.3.3. Vers une approche discursive de la prosodie 143
3.3.3.1. Le découpage en UI (tonality) 143
3.3.3.2. La place du noyau (tonicity) 144
3.3.3.3. Les tons (tones) 145
3.4. Conclusion : l'interface prosodie/discours 147

Chapitre 4. Perspectives 148
4.1. Proéminences prosodiques 149
4.1.1. Accents secondaires 149
4.1.1.1. Hypothèses 149
4.1.1.2. Protocole 150
4.1.2. Toniques de tête 151
4.1.2.1. Niveau de l'UI 151
4.1.2.2. Au delà de l'UI 152
4.1.3. Saillance comparée du français et de l'anglais 154
4.2. Prosodie et syntaxe 157
4.2.1. Prosodie et syntaxe en discours 157
4.2.2. Ambiguïté des structures 160
4.2.3. Questions de traduction 162

4.2.3.1. Do inséré dans un contexte affirmatif 162
4.2.3.2. The prononcé /ði/ devant consonne 163
4.3. Interphonologies 164
4.3.1. Interphonologies dialectales 164
4.3.2. Interphonologies L1-L2 168

Conclusion 171

Bibliographie 175

Annexes 200
Annexe 1 : Maptask 200
Annexe 2 : Fiche de renseignements / Information sheet 201
Annexe 3 : Fiche signalétique IARI 204
Annexe 4 : Consentement éclairé LPL 206
Annexe 5 : Consentement éclairé IARI 208
Annexe 6 : Table des matières provisoire Delais et al. à paraître 211
Annexe 7 : Exemples sonores du document de synthèse 213
Annexe 8 : Contenu du CD-ROM 216

Table des illustrations

Figures

Figure 1 : niveaux d'annotation dans Aix-Marsec fs 5 47
Figure 2 : niveaux d'annotation dans AixOx fs 6 48
Figure 3 : courbe de F0 pour un extrait de fs 7 54
Figure 4 : courbe de F0 pour un extrait de fichier fs 8 55
Figure 5 : fs 10 avec plancher à 50 Hz et plafond à 240 Hz 57
Figure 6 : fs 10 avec plancher à 50 Hz et plafond à 275 Hz 57
Figure 7 : fs 10 avec plancher à 40 Hz et plafond à 249 Hz 57
Figure 8 : spectrogramme de *progress* 58
Figure 9 : courbe brute de F0 en haut et courbe lissée en bas pour la phrase *Every Thursday I have a programme here on radio 4 called 'Employment'* 59
Figure 10 : fs 11 plancher 100Hz, plafond 400 Hz 59
Figure 11 : fs 11 plancher 100 Hz, plafond 500 Hz 60
Figure 12 : oscillogramme, spectrogramme, courbe de F0 et quatre tires dont les tons intsint pour un extrait de fs 12. 61
Figure 13 : extrait de la conversation PAC Donegal, locutrice Helen. fs 13 63
Figure 14 : extrait de conversation PAC DONEGAL, locutrice Majella fs 14 64
Figure 15 : exemple d'une transcription tonétique interlinéaire (Cruttenden 1997 : 37) 79
Figure 16 : exemple d'un do auxiliaire placé en tonique de tête (fs 17) 80
Figure 17 : oscillogramme, spectrogramme et courbe de F0 pour fs 11 82
Figure 18 : exemple de transcription tonétique interlinéaire (Wells 2006 :8) 84
Figure 19 : visualisation de la mélodie dans PRAAT de l'énoncé de la figure 18 (fs 19) 84
Figure 20 : oscillogramme, spectrogramme et courbe de F0 pour le mot *calbain* (fs 21) 90
Figure 21 : relation de proéminence au niveau du mot (Herment & Turcsan, à paraître) 92

Figure 22 : premier niveau de hiérarchie des proéminences sur les mots à l'échelle de l'UI 97
Figure 23 : deuxième niveau de hiérarchie des proéminences sur les syllabes à l'échelle du mot 98
Figure 24 : correspondance avec les degrés de proéminence au niveau de la syllabe.....99
Figure 25 : courbes d'intensité et de F0 pour fs 10 100
Figure 26 : courbes d'intensité et de F0 pour fs 22 101
Figure 27 : courbe de F0 pour fs 23 102
Figure 28 : allongement du groupe consonantique initial de *scooters* fs 27....105
Figure 29 : allongement du /k/ : tenue et explosion plus longues fs 22.....105
Figure 30 : coup de glotte devant *actual* fs 28.....105
Figure 31 : courbe de F0 pour fs 29 112
Figure 32 : fichier fs 29 manipulé sous PRAAT : fs 30 113
Figure 33 : courbe de F0 pour le mot *benefited* (fs 8) 129
Figure 34 : courbe de F0 pourfs 69 : assertion en montée (HRT).....166
Figure 35 : courbe de F0 pour fs 70 : question ascendante 166
Figure 36 : courbe de F0 pour fs 71 : assertion en montée (HRT).....166
Figure 37 : courbe de F0 pour fs 72 : question ascendante 167
Figure 38 : contenu du CD-ROM.....216

Tableaux

Tableau 1 : Tonetic Stress Marks dans Aix-Marsec.....52
Tableau 2 : échelle de degrés des proéminences dans l'unité intonative 93
Tableau 3 : corrélats acoustiques de la proéminence : tendances 107
Tableau 4 : constitution du corpus ICE-GB 125
Tableau 5 : pourcentage de clivées prononcées avec une prosodie canonique ...133
Tableau 6 : prosodies les plus fréquentes en discours pour chaque structure 134
Tableau 7 : tableau récapitulatif des patrons prosodiques canoniques et non canoniques et de leur répartition pour chaque structure. 136

INTRODUCTION

Contextualisation

Mes recherches se situent en prosodie, suivant la tradition du Laboratoire Parole et Langage à Aix-en-Provence (avec Mario Rossi, Albert Di Cristo, Daniel Hirst, Mariapaola D'Imperio entre autres), où j'ai soutenu ma thèse sous la direction de Daniel Hirst, et auquel je suis rattachée depuis près de 15 ans maintenant. Je m'intéresse principalement au discours, et ce depuis mes travaux de doctorat. En effet, après avoir étudié les corrélats acoustiques et prosodiques de l'emphase dans le discours spontané anglais pendant ma thèse (Herment-Dujardin, 2001), j'ai continué ensuite à m'intéresser à l'emphase et à la proéminence, sous des angles différents. Ces perspectives nouvelles m'ont été dictées à la fois par des choix méthodologiques et théoriques que je vais développer dans ce document de réflexion, et par mes étudiants, à travers les cours que je leur dispense. Je voudrais insister ici sur le fait que les cours que je donne et les étudiants que j'encadre nourrissent véritablement ma recherche. J'en donnerai des exemples précis, tant dans les commentaires de certains exemples tirés de mes cours ou des travaux de mes étudiants que dans la référence à certains mémoires de M2 que j'ai dirigés. J'enseigne la phonétique et la phonologie anglaise, et m'inscris à nouveau dans la tradition aixoise, pour l'enseignement de l'intonation cette fois, sur les traces et dans les pas de Georges

Faure, André Raphaël et Michel Ginésy. Les théories qui sous-tendent mes choix pédagogiques sont discutées ici et ce document de synthèse propose donc un va-et-vient récurrent entre recherche et expérience d'enseignement et d'encadrement, deux volets de notre métier d'enseignant-chercheur qui sont complémentaires et intimement liés dans mes travaux. L'étude sur le son [i] (Herment 2010c) par exemple m'a été entièrement inspirée par les questions récurrentes de mes étudiants qui ne savaient jamais où placer ce son dans leurs transcriptions. L'étude sur *do* me permet de mieux expliciter les différences de sens que l'intonation permet de véhiculer. Le troisième volet, administratif, n'est pas complètement dénué de lien, mais ils sont moins évidents et n'apparaissent pas aussi clairement. Ils sont néanmoins présents et importants dans l'évolution de ma réflexion, puisqu'ayant dirigé les laboratoires de langues de l'université pendant 6 ans et le service audio-visuel et multimédia pendant 2 ans, j'ai eu l'occasion de réfléchir aux applications pédagogiques de la recherche pour les mettre à la disposition des étudiants et des enseignants en tant que directrice de services communs. Ainsi par exemple, la création d'un Espace Oralité aux laboratoires de langues de mon université a été possible, et ce aussi avec le concours des lecteurs du département d'anglais dont j'ai été la responsable pendant 5 ans. La gestion du matériel et en particulier du studio d'enregistrement a joué un rôle dans mes recherches et dans mes encadrements, tout particulièrement dans la constitution de corpus, aussi bien pour mes travaux que pour ceux de mes étudiants de MASTER.

Cette introduction va repréciser les différentes notions qui sont au cœur de mon document de synthèse et resituer brièvement la tradition britannique prosodiste à laquelle j'appartiens.

Prosodie

Mes travaux s'inscrivent donc dans une tradition prosodiste, avec un penchant pour l'intonation. Je distingue en effet les deux termes prosodie et intonation. Ils peuvent définir les mêmes concepts dans la littérature (chez Wells

2006 : 2 par exemple), et il est parfois difficile de faire le départ entre les deux, la définition pouvant varier d'un auteur à l'autre. J'entends prosodie au sens large et intonation dans un sens plus étroit. Di Cristo (2013 : 21) donne une définition très précise et complète de la prosodie :

> La prosodie est une branche de la linguistique consacrée à l'analyse des propriétés formelles (phonologie prosodique), de la matérialité (phonétique prosodique) et de la fonctionnalité des éléments non verbaux de l'expression orale, non coextensifs aux phonèmes, tels que l'accent, les tons, l'intonation, la quantité, le tempo et les pauses, que l'on qualifie collectivement de prosodèmes. La matérialité physique des prosodèmes se manifeste par les variations de la fréquence fondamentale (F0), de la durée et de l'intensité (paramètres prosodiques acoustiques), ces variations étant perçues par l'auditeur comme des changements respectifs de hauteur, longueur et de volume sonore (paramètres prosodiques auditifs). Les informations prosodiques ont la particularité d'être polysémiques et de véhiculer à la fois des informations d'ordre linguistique, paralinguistique et extralinguistique, qui se révèlent déterminantes pour l'interprétation des énoncés du discours et de la conversation. (Di Cristo 2013 : 21).

Cette définition a l'avantage de faire la distinction entre production (les paramètres prosodiques acoustiques) et perception (les paramètres perceptifs auditifs). On fait souvent abstraction de ces deux niveaux et les termes s'y référant sont en général employés de façon interchangeable alors que cela ne devrait pas être le cas, ainsi que Ladd (2008) le fait remarquer. En anglais les termes *F0*, *intensity* et *duration* devraient être employés pour l'aspect physique, acoustique (donc du point de vue de la production), alors que l'on devrait utiliser les termes *pitch*, *loudness* et *quantity* respectivement pour l'aspect psychophysique de la perception.

La définition d'Albert Di Cristo inscrit aussi clairement l'étude de la prosodie en phonétique et en phonologie. C'est également l'avis de Martin (2009) :

> La prosodie est le domaine de la phonétique qui étudie l'intonation, l'accentuation, le rythme et la durée des réalisations des phonèmes.
>
> En passant au cours des vingt dernières années du domaine de la phonétique à celui de la phonologie, le terme de prosodie en linguistique a aussi annexé l'acceptation musicale du terme : outre l'intonation, séquence de variations

> musicales attachées à chaque syllabe, le rythme, succession de durées longues ou brèves, et l'intensité, qui ont depuis longtemps constitué les paramètres traditionnels de la prosodie, la phonologie prosodique a aussi élaboré (ou tenté d'élaborer) des systèmes de règles associant les syllabes du texte avec par exemple des degrés d'accent ou des mouvements ou des hauteurs mélodiques. (Martin 2009 : 13).

Ainsi phonétique et phonologie se complètent-elles et sont indissociables dans l'étude de la prosodie. C'est à partir des études phonétiques des divers paramètres acoustiques qui constituent la prosodie que l'on peut tenter de déduire des systèmes de règles et des généralisations qui font partie du domaine de la phonologie. On peut aussi adopter la démarche inverse en partant d'hypothèses phonologiques que des analyses phonétiques vont nous permettre de confirmer ou d'infirmer. Les études que j'ai réalisées avec Gabor Turcsan (Herment & Turcsan 2103 & à paraître et Turcsan & Herment 2013 & à paraître) sont particulièrement représentatives de cette complémentarité entre phonétique et phonologie pour l'étude de la prosodie.

On subdivise traditionnellement la prosodie en trois grands volets :

- l'organisation accentuelle et rythmique. Par accentuation, on entend à la fois l'accentuation au niveau du mot (*word stress*) et au niveau de la phrase (*sentence stress*). Il existe différents degrés de proéminence qui contribuent à la formation de groupements, qui eux-mêmes constituent le rythme. J'ai travaillé sur l'accent de mot (Turcsan & Herment, 2013) et la notion de constituants prosodiques est présente dans la plupart de mes travaux, comme dans Herment (à paraître (b)) ou bien les travaux en prosodie et syntaxe puisque mes analyses se fondent pour la plupart sur l'unité intonative.
- La dimension temporelle de la prosodie. La durée des unités linguistiques, les pauses et le tempo (débit ou vitesse d'élocution) font partie de ce volet. Les deux premiers aspects sont traités dans certains de mes travaux, en particulier la durée des phonèmes dans Herment & Turcsan (à paraître) et les pauses dans Herment & Hirst (2002, 2003) et Herment (2006).
- L'intonation, c'est-à-dire le système des patrons mélodiques distinctifs de l'énoncé et de ses constituants. C'est la partie de la prosodie sur laquelle j'ai le

plus travaillé (voir l'axe 2 du volume des travaux et Turcsan & Herment à paraître).

On pourrait rajouter aux paramètres qui constituent la prosodie la qualité de voix (*cf.* Auran 2004 et 2012 pour une discussion plus approfondie sur ce sujet). Ce paramètre est malheureusement encore très peu connu et très peu étudié car très difficile à mesurer (*cf.* Coadou 2007). Je n'ai pas travaillé sur la qualité de voix, je la mentionne simplement dans quelques travaux.

Intonation

Ainsi, l'intonation fait partie de la prosodie. Dans l'acception dans laquelle je m'inscris, il s'agit des caractéristiques suprasegmentales, post-lexicales ou encore non-lexicales, regroupant des phénomènes comme la forme globale des contours mélodiques, les phénomènes de frontières de groupes, la déclinaison, les mouvements mélodiques selon Hirst & Di Cristo (1998 : 4). Cruttenden (1997 : 7) définit l'intonation en termes de contours mélodiques utilisés pour véhiculer des significations particulières. C'est la définition de Ladd (2008 : 4) que je citerai et à laquelle je souscris : « Intonation […] refers to the use of *suprasegmental* phonetic features to convey 'post-lexical' or *sentence-level* pragmatic meanings in a *linguistically structured* way ». Là encore phonétique et phonologie se croisent et se complètent : la définition de Ladd inclut à la fois les mots « phonetic features » et « meaning ». Les trois termes donnés en italiques par l'auteur lui-même sont les points-clés de cette définition. Les critères phonétiques suprasegmentaux sont la fréquence fondamentale (F0), l'intensité et la durée. C'est au niveau post-lexical, ou niveau de la phrase, que l'intonation transmet des informations, comme le type de phrase, le focus ou encore la structure informationnelle. Enfin, l'intonation est structurée linguistiquement en ce qu'elle est organisée en entités catégoriques distinctes, telles que ton bas ou montée terminale. La plupart des travaux que j'ai menés portent sur l'intonation et font appel à ces notions. J'y reviendrai donc tout au long de ce document. Certains de mes travaux portent sur les deux autres domaines de la prosodie que sont l'organisation métrique (l'accentuation avec le

travail sur les dissyllabes) et l'organisation temporelle (plus particulièrement la durée des réalisations des phonèmes et les considérations sur la différence entre forme pleine et forme réduite avec les travaux sur les voyelles hautes). Il était donc important de bien distinguer les différentes composantes de chaque notion.

Les écoles britannique et américaine

Plusieurs systèmes de description de l'intonation ont été développés mais ont pris des chemins différents en Europe et aux Etats-Unis.

Les descriptions britanniques de l'intonation anglaise remontent au 18ème siècle avec Walker (1781, 1787), qui fut le premier à utiliser des marques pour représenter les tons *rising*, *falling*, *monotone*, *fall-rise* et *rise-fall*. Le début du 20ème siècle fut marqué par Jones (1909a, 1909b, 1918), qui utilisa une courbe continue entre deux lignes horizontales pour représenter la mélodie et qui inventa les dénominations ton 1 et 2 pour décrire les deux contours basiques montants et descendants de l'intonation. Jones fut suivi par d'autres phonéticiens, dont Palmer (1922), qui introduisit les termes *nucleus* (la syllabe qui porte ou sur laquelle commence le mouvement mélodique principal du groupe), *head* (tout ce qui précède le noyau) et *tail* (tout ce qui suit le noyau). C'est sur ces fondations que les travaux de l'école britannique se sont appuyés pour une approche configurationnelle, avec par exemple Kingdon (1958), O'Connor & Arnold (1961) et bien-sûr Halliday (1967). A la suite de ce dernier, on distingue traditionnellement trois niveaux ou plus exactement trois types de décisions que le locuteur doit prendre :

- la division en unités intonatives, ou encore *tonality* ;
- le placement de la syllabe nucléaire, *tonicity* ;
- et le choix du type de contour nucléaire, *tone*.

Ces trois paramètres sont communément appelés les 3T (*the three Ts*).

La tradition britannique perdure avec notamment les ouvrages de Cruttenden (1997, 2008), Wells (2006) ou Roach (2009).

Aux Etats-Unis, la description de l'intonation de l'anglais américain s'est développée à partir des années 1940 avec Pike (1945) puis Trager & Smith (1951), qui ont défini quatre niveaux mélodiques et trois tons terminaux. Les niveaux mélodiques et les tons terminaux sont les phonèmes de la prosodie et les séquences niveaux + tons terminaux en sont les morphèmes. Les auteurs en donnent des significations précises. Une approche plus récente, l'approche métrique autosegmentale (*Autosegmental Metrical theory*, AM), inspirée de Trager & Smith, a vu le jour avec les travaux de Liberman (1975), Goldsmith (1976), Leben (1976), Bruce (1977) et surtout Pierrehumbert (1980). Seuls deux tons haut et bas sont retenus, ainsi que deux niveaux de frontière. Les tons sont alignés aux proéminences accentuelles ou associés aux frontières. L'approche autosegmentale combine phonétique et phonologie en donnant une description phonologique de séquences d'éléments catégoriquement distincts, tout en associant les éléments phonologiques à des paramètres acoustiques continus (Ladd, 2008 : 43).

Les constituants prosodiques

Les constituants prosodiques ont fait l'objet de nombreuses études se réclamant d'approches différentes (voir Shattuck-Hufnagel & Turk 1996 pour une synthèse). Dans le cadre autosegmental et selon les travaux de Pierrehumbert (1980), Beckman & Pierrehumbert (1986) et Pierrehumbert & Hirschberg (1990), le syntagme intonatif (IP pour *intonation phrase*) est défini comme un contour complet selon des critères intonatifs, comportant au moins un accent mélodique (*pitch accent*) et se terminant par un ton terminal (*boundary tone*). Dans la hiérarchie prosodique, la phrase phonologique (*Phonological Utterance*) domine l'IP et en dessous, l'IP est construite de constituants prosodiques inférieurs : le syntagme intermédiaire (ip pour *intermediate phrase*), lui-même constitué d'un ou plusieurs mots prosodiques (PWd pour *Prosodic Word*).

Dans la tradition britannique, l'unité intonative (TU pour *Tone Unit*) est analysée comme un contour tonal complet, tout comme l'IP. Elle contient, comme

indiqué plus haut et suivant la terminologie de Palmer (1922) légèrement modifiée, obligatoirement un noyau (*nucleus*) et un seul, et de façon optionnelle une tête (*head*, qui part de la première syllabe accentuée jusqu'au noyau), une pré-tête (*prehead*, toutes les syllabes inaccentuées précédant le *head*) et une queue (*tail*, toutes les syllabes inaccentuées qui suivent le noyau) (voir Crystal 1969 ou Gussenhoven 2004 pour une description plus récente). Les mouvements mélodiques traditionnellement reconnus (avec des variantes selon les auteurs) sont le *fall*, *rise*, *rise-fall* et *fall-rise*. Le mouvement mélodique commence sur la syllabe nucléaire et s'étend jusqu'à la fin de l'unité intonative.

C'est en cela que les deux systèmes américain et britannique diffèrent le plus. Tandis que le système britannique fait référence à la forme du contour global, le système de Pierrehumbert définit des cibles tonales pour décrire les accents mélodiques. Cette différence à la fois syntagmatique et paradigmatique dans le ciblage des entités pertinentes a des conséquences fortes sur les conventions transcriptionnelles de l'intonation et sur les algorithmes d'annotation automatique.

Transcription : contours ou niveaux ?

Les méthodes de transcription utilisées sont logiquement inspirées des théories sous-jacentes et sont donc complètement différentes selon que le modèle est configurationnel comme chez les linguistes de l'école britannique, ou tonal chez les adeptes de la théorie autosegmentale.

Pour les premiers, des transcriptions intégrées ou interlinéaires sont largement utilisées. L'avantage de la transcription intégrée est que la lecture en est aisée puisque les marques tonétiques (*tonetic stress marks,* désormais TSM), développées par Kingdon (1958) sont dans le texte. En revanche, seules les syllabes accentuées sont porteuses d'un signe tonétique. Différentes conventions

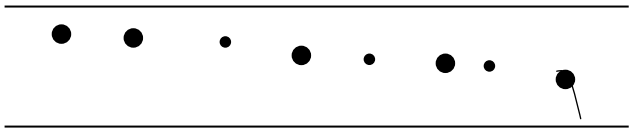
de TSM ont été utilisées et on trouve ce type de transcription dans Aix-Marsec[1] par exemple :

*<Good `morning || >`more *news about the ~Reverend _Sun *Myung `Moon |*

La transcription scalaire ainsi que la transcription tonétique interlinéaire notent les syllabes entre deux barres horizontales qui symbolisent le grave du locuteur pour celle du bas et l'aigu pour celle du haut. En transcription scalaire, les syllabes ne sont notées que sur trois niveaux possibles : l'aigu, le médium et le grave. Les syllabes accentuées sont représentées par des tirets, qui sont horizontaux s'ils sont porteurs de tons statiques et obliques s'il y a un mouvement mélodique. Les points sont les syllabes inaccentuées. Cette transcription est très schématique. Les déclinaisons (*downstepping)* ne sont pas représentées par exemple :

Why hadn't I thought of this before?

Dans cet énoncé, les syllabes qui se trouvent entre *why* et *–fore* peuvent tout à fait être prononcées en légère descente mais cela ne figure pas sur la transcription, contrairement à la transcription tonétique interlinéaire, où l'intervalle entre les deux lignes qui représentent la tessiture du locuteur n'est pas codé en niveaux. Les syllabes sont représentées par des points, les points plus gros sont les syllabes accentuées alors que les points moins larges sont les syllabes inaccentuées. Dans le même exemple ci-dessous, les points forment une ligne qui descend doucement :



Why hadn't I thought of this before?

1. Voir Knowles, Wichmann & Alderson (1996) pour les TSM de Aix-Marsec (je les détaille dans Herment 2010b et à paraître a).

Ce dernier type de transcription est vu comme une transcription étroite de l'intonation, contrairement aux systèmes de transcription qui utilisent des niveaux plutôt que des contours pour représenter l'intonation et qui sont conçus comme des transcriptions larges, phonologiques, comme c'est le cas de ToBI (*Tone and Break Indices*). ToBI est un système de transcription de l'anglais américain fondé sur les travaux de la théorie autosegmentale (Pierrehumbert 1980, Pierrehumbert & Beckman 1988, Wightman *et al.* 1991 et d'autres). Il a été proposé par Silverman *et al.* (1992) puis développé, et il est largement utilisé aujourd'hui par la communauté scientifique (voir Beckman *et al.* 2005). Il n'est pas le lieu ici de décrire en détail le système ToBI (voir notamment Beckman & Ayers 1994 pour cela) mais une brève description s'impose tout de même. L'idée principale est que les étiquettes sont conçues pour représenter les distinctions phonologiques plutôt que les détails des réalisations phonétiques. Cela implique une connaissance préalable des patrons intonatifs distinctifs de la langue étudiée. Le système ToBI est fondé sur une description de l'anglais américain donnée dans la thèse de Pierrehumbert (1980). En anglais seuls deux tons mélodiques (*pitch* accents) existent, qui sont des points-cible relatifs : haut (H pour *high*) et bas (L pour *low*). Ces tons sont associés à des syllabes et étiquetés de diacritiques :

- le symbole % est associé à un ton de frontière terminale ;
- le symbole – correspond à une frontière de constituant intermédiaire (un accent de phrase) ;
- l'étoile * est associée à une syllabe perçue proéminente ;
- le point d'exclamation ! indique une déclinaison (*downstepping*)
- le symbole HiF0 est utilisé pour localiser le pic de F0 dans chaque constituant majeur.

Ainsi H*L correspond à une chute et L*H à une montée ; un *fall-rise* dans la tradition britannique serait annoté H*L+H%. L'annotation s'effectue manuellement par des annotateurs entrainés à partir de l'écoute et de la visualisation de la courbe de F0. Du fait que ToBI requiert une description préalable des contours intonatifs distinctifs de la langue, il ne peut s'appliquer qu'à l'anglais. Cependant, des versions modifiées de ToBI ont été adaptées à

d'autres langues, notamment l'allemand (G-ToBI, Grice & Baumann 2002), le japonais (J-ToBI, Venditti, 1995, Campbell & Venditti, 1995), le coréen (K-ToBI, Beckman & Jun 1996), le grec (GR-ToBI, Arvaniti & Baltazani 2000) ou le français (F-ToBI, Delais *et al.* à paraître). L'adaptation peut se révéler très complexe, d'une part parce que les descriptions de l'intonation des langues ne sont pas toujours consensuelles, et d'autre part parce que les unités et les niveaux tonals peuvent différer grandement de l'anglais.

En parallèle du développement de ce système, Hirst & Di Cristo (1998) avaient proposé un système de transcription de l'intonation, Intsint (International Transcription System for INTonation), également inspiré par la théorie autosegmentale. Ce système fournit une représentation phonologique de surface des contours intonatifs. Il est envisagé comme une interface potentielle entre une représentation symbolique purement phonologique de l'intonation et une représentation phonétique continue (voir De Looze *et al.* 2012 pour une synthèse). Le contour intonatif est décrit comme une séquence de segments tonals. Huit symboles sont utilisés et sont de trois types :

- les tons absolus (*absolute tones*) : *Top, Mid, Bottom*, calculés selon la tessiture du locuteur ;

- les tons relatifs (*relative tones*) **:** *Higher, Same, Lower*, définis par rapport aux tons précédents ;

- les tons itératifs relatifs (*iterative relative tones*) : *Upstepped, Downstepped*, également définis relativement aux tons précédents, mais qui impliquent en général des changements mélodiques moins amples et qui apparaissent souvent dans des séquences de paliers soit montants soit descendants.

Contrairement au système ToBI, Intsint reste au niveau phonologique de surface et ne prend pas en compte les frontières par exemple. En effet, Intsint a été conçu pour décrire l'intonation de plusieurs langues, et en particulier au départ celle du français et de l'anglais. Or les deux langues étant extrêmement différentes au niveau accentuel, et au niveau du découpage en constituants prosodiques, cela rendait impossible le marquage des frontières ou le positionnement des proéminences par exemple. Intsint peut donc être considéré

comme l'équivalent d'une transcription phonétique étroite en IPA au niveau suprasegmental (Hirst & Di Cristo 1998 : 14) et se distingue ainsi fondamentalement du système ToBI. L'annotation est entièrement automatique et peut être optimalisée grâce à des corrections manuelles et un recours à la resynthèse permettant de vérifier que l'énoncé a une prosodie perceptivement équivalente à celle de l'original. Le système a été appliqué sur de nombreuses langues, comme nous l'expliquons dans De Looze *et al.* (2012).

Un autre système est IViE (Grabe *et al.* 1998, Grabe & Post 2004), modèle qui suit la théorie autosegmentale et est inspiré de ToBI. IViE a été développé pour pallier trois problèmes de ToBI : le manque d'accord inter-annotateur concernant les tons dans ToBI, la difficulté à s'accoutumer au système de transcription et le fait que ToBI n'est pas performant pour décrire des variétés d'anglais non standard (voir Nolan & Grabe 1997). IViE offre une alternative à ToBI fondée sur un inventaire de tons différent car dérivé de descriptions de l'anglais britannique et qui ne représente pas un système intonatif particulier. Le système est au contraire conçu comme un ensemble d'étiquettes qui permettent de comparer différentes variétés d'anglais. Enfin, IViE prend en compte les frontières d'unités intonatives mais repose sur un seul niveau de constituance prosodique : l'IP (*intonation phrase*).

Le dernier système qu'il me semble important de mentionner est le Prosogram de Mertens (2004a, 2004b, 2011). On peut le voir comme un système configurationnel : les transcriptions symboliques proposées par le système sont générées automatiquement à partir de « prosogrammes », qui sont des représentations stylisées des contours intonatifs. Ces prosogrammes apparaissent sur les syllabes (le système procède au préalable à un découpage automatique en syllabes) et sont indépendants de la langue. Il en résulte une transcription symbolique, qui, comme dans le système Intsint, n'est pas dépendante d'une connaissance phonologique des contours de la langue étudiée et qui n'indique donc pas non plus ce qui est distinctif. Le résultat est très intéressant car il ressemble aux transcriptions interlinéaires mentionnées plus haut (voir http://bach.arts.kuleuven.be/pmertens/prosogram/).

Positionnement, enseignement et recherche

La question se pose à ce stade de savoir quelle approche privilégier et quel système choisir. Sur le plan théorique, les deux cadres britannique et américain ne sont pas aussi éloignés qu'il y paraît et la question s'est donc posée en termes de choix pédagogiques. Traditionnellement en France, l'intonation de l'anglais est enseignée selon l'école britannique. C'est ainsi que je l'ai apprise et que j'ai commencé à l'enseigner lors de mes premiers cours d'intonation, avec les ouvrages pédagogiques de Ginésy (1995), de Roach ([1983]2009), puis avec celui de Wells (2006). Il semble en effet plus simple et pédagogiquement plus pertinent d'enseigner une approche configurationnelle avec des contours globaux plutôt qu'un système compositionnel avec des tons, même si l'étude de Toivanen (2005) montre que les deux systèmes donnent des résultats très similaires pour l'apprentissage de l'anglais L2 chez des locuteurs finnois. Car n'oublions pas que nous dispensons nos cours à des étudiants francophones pour la plupart, donc des apprenants, et que dans les années de licence, nous enseignons la prononciation et l'intonation de l'anglais L2 bien davantage que les théories phonologiques.

En ce qui concerne l'annotation prosodique, l'école britannique offre un système de notation avec des symboles diacritiques très simples, tels que (̗) pour la petite montée (*low-rise*), (̖) pour la petite chute (*low fall*), (ˇ) pour le ton creusé (*fall-rise*), etc. Ce système de transcription est par exemple celui qui est requis à l'épreuve écrite de phonologie de l'agrégation d'anglais. Les transcriptions interlinéaires représentent aussi un outil pédagogique précieux pour nos étudiants. Ces représentations de l'intonation sont visuellement bien plus parlantes qu'une suite de tons et de diacritiques pour un apprenant. Les systèmes de notation ToBI ou Intsint me semblent beaucoup plus complexes pour les étudiants et difficiles à mettre en place, même au niveau MASTER. Le système britannique traditionnel est donc resté mon domaine de prédilection pour l'enseignement.

Pour la recherche, j'ai souhaité coller au modèle britannique également, afin de pouvoir relier recherche et enseignement. Comme je l'ai dit plus haut, la mise

en relation de l'enseignement et de la recherche constitue une priorité pour moi, l'un alimentant l'autre et vice versa. Je souhaite pouvoir appliquer mon travail de recherche dans mes cours. Ce n'est pas chose aisée, mais c'est un idéal vers lequel je tends et que je commence à voir s'esquisser.

En ce qui concerne les systèmes d'annotation prosodique pour la recherche, ils vont varier en fonction de ce que l'on cherche à analyser (voir chapitre 1). Les transcriptions interlinéaires sont difficilement applicables en recherche, et les TSM sont privilégiées pour certains travaux. Il est parfois utile d'avoir recours aussi à des modes de transcription différents comme ToBI, Intsint ou IViE. Collant à la tradition britannique, je n'utilise pas le système ToBI. IViE est d'un grand intérêt et je m'en inspire d'une certaine façon. Mais c'est plutôt Intsint que j'utilise : j'y ai été familiarisée dès mon DEA (Herment 1997), au cours duquel j'en ai réalisé une validation expérimentale. J'ai donc ensuite logiquement continué à utiliser et favoriser ce système. Intsint est de plus implémenté dans le logiciel SPPAS (Bigi 2012) que j'utilise à d'autres fins (voir chapitre 1). Cependant, Intsint est un système de transcription automatique de l'intonation qui n'est donc applicable que sur des énoncés de bonne qualité acoustique. Au cours de mes travaux j'ai utilisé des corpus de qualité sonore médiocre et ai donc été amenée à faire des analyses auditives des énoncés (voir chapitre 3). La méthode que j'utilise dans ces cas-là est de tradition britannique comme je l'ai dit, c'est-à-dire qu'elle est configurationnelle. Je code donc en contours et non en niveaux tonals et m'inspire des transcriptions interlinéaires mentionnées plus haut, en considérant comme dans IViE qu'il n'y a qu'un niveau de constituance prosodique, l'unité intonative, et non deux. Je considère en effet à l'instar des linguistes de l'école britannique qu'une unité intonative ne peut contenir qu'un seul accent mélodique au sens de *pitch accent* (Bolinger 1958). Les autres accents peuvent être mélodiques mais seront moins marqués (voir chapitre 2).

Lorsque la qualité sonore le permet, j'utilise le système Intsint, bien souvent pour vérifier mes impressions auditives (voir le chapitre 1), les compléter et les affiner : « both auditory and instrumental analysis have something to offer and should be combined » (Cruttenden 1997 : 6). Ce travail d'écoute et de

reconnaissance auditive des contours mélodiques de l'anglais me semble essentiel et je pense important que les étudiants y soient formés. J'explique dans le chapitre 1 qu'il faut se méfier des logiciels et des transcriptions automatiques si l'on n'a pas une oreille aguerrie.

Structure du document de synthèse

C'est donc dans la perspective de l'intime relation entre enseignement et recherche que je voudrais construire ce document de synthèse, en montrant les ponts que l'on peut tendre entre eux et en prenant du recul sur mes travaux et sur mes cours, afin de mieux montrer comment ils sont liés, comment l'on peut appliquer la recherche à l'enseignement et comment l'on peut se nourrir de son expérience d'enseignant pour sa recherche.

Deux axes principaux se dégagent de mes travaux, les proéminences prosodiques et l'interface prosodie/syntaxe en discours, avec en filigrane les corpus et les outils, que je considère davantage comme représentant un axe transversal indissociable de mes recherches. J'ai choisi de consacrer le premier chapitre à cet axe afin d'établir les bases méthodologiques et théoriques sur lesquelles mes recherches s'appuient. Dans les deux chapitres suivants, j'essaierai de retracer les grandes lignes de mes travaux : j'en présenterai les résultats en tentant de théoriser mes points de vue, tout en passant par la méthode de l'exemple et en émaillant mon document d'études de cas concrets. Consciente que les résultats actuels méritent des approfondissements, je développerai dans le quatrième chapitre des perspectives de recherche future, montrant par la même occasion dans quelles directions je pourrais diriger des recherches.

Parce qu'il me semble difficile de faire passer mes convictions et d'expliciter mes points de vue uniquement par écrit, et parce que la méthode perceptuelle occupe une place toute particulière dans mes travaux, un CD-ROM accompagne ce dossier d'habilitation à diriger des recherches : le lecteur y trouvera tous les exemples sonores donnés dans le présent document de synthèse, numérotés dans

leur ordre d'apparition. L'annexe 7 en donne la liste avec la référence aux corpus utilisés. L'annexe 8 décrit le contenu du CD-ROM.

CHAPITRE 1

Des corpus et des outils

Mon travail de recherche repose quasiment uniquement sur des corpus. Ma thèse s'appuyait déjà sur trois corpus de parole spontanée anglaise : une émission de télévision, une émission de radio et une conversation, ce dernier corpus, le COEC (Herment 2010a), que j'avais collecté moi-même et que j'ai depuis déposé sur le SLDR (*Speech Language Data Repository*), une plateforme (sur laquelle je reviendrai dans ce chapitre) qui peut être considérée comme un outil. Le présent chapitre s'intitule « des corpus et des outils » : il est évident que l'on ne peut pas constituer un corpus sans outils, quel que soit le type de corpus, et que l'on ne peut plus analyser un corpus oral sans avoir recours à des outils (*cf.* Ballier & Martin à paraître (a) et Nguyen & Adda-Decker 2013). De nombreux corpus sont disponibles pour la communauté scientifique et j'en ai utilisé pour mes travaux, en particulier le corpus ICE-GB (Greenbaum 1996, Aarts *et al.* 2002 et plus précisément Aarts & Nelson 1999 pour la composante britannique), ou encore le corpus Aix-Marsec[2] (Auran *et al.* 2004). Les films, les émissions de radio (dont le corpus Aix-Marsec est composé) et de télévision, aussi diverses qu'elles soient, représentent une ressource très riche, largement utilisée par nos étudiants par

2. Aix-Marsec est la version de Marsec (*Machine-readable Spoken English Corpus*, Roach *et al.* 1993) enrichie au Laboratoire Parole et Langage à Aix-en-Provence par Cyril Auran, Caroline Bouzon et Daniel Hirst.

exemple. J'ai dirigé plusieurs mémoires de M1 ou de M2 pour lesquels les étudiants avaient pris comme corpus des séries télévisées ou des films. Certaines études se basent sur des corpus récupérés sur la toile, comme les vlogs de Goudet (2013a et 2013b par exemple). Malgré cette grande diversité, et pour des raisons que je détaille plus loin, il s'avère nécessaire pour certaines études de collecter un corpus soi-même. C'est ce que j'ai été amenée à faire à plusieurs reprises. Le volume des travaux, dont la troisième partie traite des outils et des corpus, inclut cinq corpus que j'ai montés moi-même, avec l'aide de collègues pour certains, entre 1999 et 2013. J'ai acquis une certaine expérience de la constitution de corpus, qui représente une part importante de mes travaux car il s'agit d'une activité chronophage. Cette expérience acquise sur plusieurs années me permettra de proposer quelques remarques d'ordre diachronique tout au long de ce chapitre et de pointer les failles des premiers corpus que j'ai constitués.

L'axe « corpus et outils » du volume des travaux de ce dossier d'habilitation à diriger des recherches constitue le troisième axe car je le conçois comme un axe transversal : tous mes articles sont fondés sur des corpus et font usage d'outils. J'ai choisi de traiter cette partie dans le premier chapitre de la synthèse afin de poser les bases théoriques et méthodologiques qui sous-tendent les travaux présentés dans le reste du document.

Le présent chapitre s'intéresse tout d'abord à la constitution des corpus oraux. M'appuyant sur une réflexion sur la méthodologie du montage de corpus, je remets en cause la typologie traditionnelle en arguant que des contraintes quasiment purement logistiques entrent en jeu dans le type de corpus que l'on crée. Une deuxième sous-section est consacrée aux outils et à l'annotation. J'aborde la question de la subjectivité de l'annotation, qui est complètement dépendante d'une théorie sous-jacente, et j'explique donc quel fondement théorique est le mien. La dernière partie de ce chapitre est consacrée à une mise en garde sur l'utilisation parfois dangereuse des outils dont nous disposons.

1.1. Les corpus oraux

1.1.1. Qu'est-ce qu'un corpus ?

En tout premier lieu on peut se demander ce qu'est un corpus en linguistique : « Le corpus est un ensemble homogène et significatif de données linguistiques observées et à partir desquelles pourra s'élaborer la théorie. » (encyclopédie universalis en ligne[3]).

Si l'on applique cette définition aux corpus oraux, il ne s'agit pas en effet uniquement d'enregistrer des locuteurs à tout va, et de stocker les données dans un ordinateur. Un corpus suppose une réflexion préalable, à plusieurs niveaux. Pourquoi monter un corpus, dans quel but et pour quoi faire ? De quelle façon ? Avec le développement technologique et la possibilité de collecter et de stocker un nombre beaucoup plus important de données, la notion de corpus a évolué et la linguistique de corpus a vu le jour, avec entre autres Sinclair, l'un des pionniers, pour qui un corpus est « a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language » (Sinclair 1996 : 4). Ainsi un ensemble de données rassemblées sans réflexion préalable sur ce qui motive la collecte n'est pas un corpus.

Pour les données orales, la collecte est soumise à une contrainte de qualité sonore impérative : là encore, une réflexion en amont est nécessaire. Va-t-on avoir recours à des analyses acoustiques ? Si oui, la qualité doit être excellente. Je reviendrai sur ce point plus bas.

Le choix des locuteurs est un autre problème crucial sur lequel je m'attarderai également plus loin.

Une fois toutes ces questions résolues et les décisions et choix scientifiques pris, la collecte peut être effectuée, avec le matériel technique adéquat (une pièce dédiée par exemple, ou bien du matériel portable).

3. http://www.universalis.fr/encyclopedie

Mais le travail est loin d'être achevé une fois la collecte effectuée : l'exploitation des données suppose un travail de mise en forme qui va varier en fonction des corpus et des buts recherchés. Ainsi Gibbon *et al.* (1997 : 79), représentant les travaux du Spoken Language Working Group de EAGLES[4], proposent la définition suivante du corpus oral : « A corpus is any collection of speech recordings which is accessible in computer readable form and which comes with annotation and documentation sufficient to allow re-use of the data in-house, or by scientists in other organisations ». Ainsi l'annotation, la documentation et la possibilité de réutiliser le corpus, c'est-à-dire sa mise à disposition, font partie intégrante d'un corpus. Je milite pour une mise à disposition gratuite, mais réservée à la communauté scientifique. Le prix exorbitant d'un corpus comme ICE-GB par exemple me paraît démesuré.

Les points que j'ai mentionnés dans cette sous-section sont essentiels lorsque l'on monte un corpus et bien souvent nos étudiants, novices en la matière, ignorent toutes ces choses. Leurs corpus sont souvent de très bonne qualité, intéressants et novateurs, mais malheureusement impossibles à réutiliser par manque de documentation.

1.1.2. Types de corpus

Une typologie des corpus a été proposée par plusieurs auteurs et est très bien résumée dans Delais-Roussarie (2003). Pour l'oral, on distingue les « corpus oraux » des « corpus de parole » :

- *les corpus oraux* qui regroupent des documents correspondant à des transcriptions, alignées ou non sur le signal, de productions orales (entretiens libres, entretiens guidés, exposés, etc.). Ces données peuvent servir pour mener des recherches sur l'oral (syntaxe de l'oral, prosodie, etc.) et ont souvent été collectées dans des situations diversifiées ;
- *les corpus de parole* qui regroupent des enregistrements de parole effectués dans des conditions expérimentales. Les données sont de nature très variée

4. EAGLES est l'acronyme pour *Expert Advisory Group on Language Engineering Standards* (coordonné par Antonio Zampolli, Consorzio Pisa Ricerche, Pisa, Italie). Ce projet a vu le jour au début des années 1990.

> (lecture de logatomes, lecture de mots, lecture de phrases isolées, lecture de textes, etc.). En outre, la taille globale du corpus peut être assez limitée. (Delais-Roussarie 2003 : 93-4)

Lors de mes recherches, j'ai travaillé sur ces deux types de corpus. ICE-GB, Aix-Marsec ou encore COEC font partie de la première catégorie, alors que le Learner's corpus of reading texts (Herment *et al.* 2010) ou le Aix-Di-Nonce (Turcsan & Herment prochainement en ligne) font partie de la deuxième. Cette division typologique est fondée sur la dichotomie entre ce que l'on appelle parole de laboratoire et parole naturelle.

Les « corpus parallèles » sont aussi une catégorie qui me concerne puisqu'ils sont constitués de documents, traduits dans une ou plusieurs langues. C'est le cas dans AixOx (Herment *et al.* 2012), qui regroupe des lectures des mêmes textes en anglais et en français.

La distinction entre « corpus de référence » et « corpus spécialisé » est également faite par les experts :

> A reference corpus is one that is designed to provide comprehensive information about a language. It aims to be large enough to represent all the relevant varieties of the language, and the characteristic vocabulary, so that it can be used as a basis for reliable grammars, dictionaries, thesauri and other language reference materials […] Special corpora are made up of texts that do not overlap as much with the large central pool. To be clearly `in a language' they must show quite a number of the grammatical and lexical features of that language, but the markedness of the patterns unique to them serves to differentiate them clearly from the general varieties of the language. In due course, and with the growing influence of internal criteria, reference corpora will be used in order to measure the deviance of special corpora. (Sinclair 1996)

Cette distinction ne me semble pas très pertinente, ni celle faite entre « corpus oraux » et « corpus de parole ». En effet, les « corpus de référence » sont divisés en sous-groupes (*sub-corpora*) : ces sous-groupes ne constituent-ils pas des « corpus de spécialité » ? Si l'on considère que le corpus ICE-GB est un « corpus de référence », on trouve différents types de données : dialogues privés, en face à face ou au téléphone, dialogues publics, en classe, dans des transactions d'affaires,

au Parlement, etc. ; monologues avec scripts ou non, à la radio ou non, etc. Prenons le cas du Parlement : il s'agit d'un type de discours très particulier, où les locuteurs ont une façon singulière de s'exprimer. Ils savent qu'ils ne seront pas coupés, ils réalisent donc beaucoup moins de montées qu'en situation de conversation spontanée : nous avons montré dans Leonarduzzi & Herment (2013b) que les extrapositions ont le plus souvent un schéma R/F (*rise* + *fall*) quand elles sont coupées en deux unités intonatives, sauf dans le discours parlementaire ou journalistique, où l'on trouve F/F (*fall*+*fall*). En ce qui concerne le discours journalistique précisément, la façon de s'exprimer est très spéciale aussi (voir les travaux de Stuart-Smith, à paraître). Comme pour le discours parlementaire, les interventions sont souvent préparées. Il s'agit d'écrit oralisé davantage que d'oral spontané. Au sein même du style journalistique, d'autres styles peuvent encore être distingués. Nous avons montré par exemple (Herment & Leonarduzzi 2012) que les clivées en *wh*- dans les commentaires sportifs ont un schéma particulier avec une montée sur la première partie de la clivée, probablement parce que le journaliste commentant un match ou une course ne sait pas ce qu'il va se passer et attend la suite en mettant la phrase en suspens grâce à la montée. C'est la raison pour laquelle certains auteurs (voir par exemple Simon *et al.*) prennent en compte les phonostyles de Léon 1993 (1993 : 3), qui les définit comme des styles sonores « tels qu'ils sont perçus en tant que caractéristiques d'un individu (jeune, vieux, homme, femme), d'un groupe social (prolétaire, bourgeois), ou d'une circonstance particulière (discours politique, sermon, etc.) ».

Quant à la distinction entre « corpus oraux » et « corpus de parole », il apparaît évident dans mes travaux et en particulier dans les corpus que j'ai constitués ou sur lesquels j'ai travaillés, que la distinction n'est pas pertinente en ces termes. Le corpus PAC[5] (voir Carr *et al.* 2004), dont la méthodologie est calquée sur celle du projet PFC[6] (Durand *et* al. 2002) regroupe par exemple à la

5. Phonologie de l'Anglais Contemporain
6. Phonologie du Français Contemporain

fois des lectures de listes de mots et d'un texte, en faisant un corpus de parole, et également des conversations formelles et informelles, typiques des corpus oraux. Comment décrire le corpus PAC alors ? Le corpus PAC a été conçu comme un corpus phonologique, c'est-à-dire qu'il comprend des productions orales spécifiquement calibrées à l'étude des systèmes phonologiques (segmentaux ou suprasegmentaux). Les différents points d'enquête, enregistrés dans des endroits particuliers pour travailler sur la variété d'anglais parlée dans un endroit précis, constituent des « corpus spécialisés ». Mais le corpus PAC dans son entier, lorsque l'on regroupe les différentes enquêtes, ne constitue-t-il pas un superbe « corpus de référence » pour l'anglais ? Les mêmes remarques s'appliquent au corpus collecté dans le cadre du projet IARI[7] (Prieto *et al.* 2010-2014) : pour cet Atlas Interactif des Langues Romanes, plusieurs styles de parole sont représentés (voir le corpus IARI Marseille dans le volume des travaux pour les détails) : des phrases simples sont élicitées, une *maptask*[8] (voir annexe 1) est réalisée et un entretien avec une personne est filmé. On retrouve à la fois « corpus oral » et « corpus de parole », ainsi que « corpus de spécialité » pour les points d'enquête et « corpus de référence » pour l'ensemble.

Ballier & Martin (à paraître a) proposent une autre typologie pour caractériser les différents types de corpus d'apprenants, qui me semble beaucoup mieux adaptée : cette typologie est fondée sur les problématiques de recherche sous-jacentes à la constitution du corpus, telles que reflétées dans leur annotation. Ils distinguent ainsi « *mute* spoken corpora (or merely recording-based transcripts), truly *speaking* corpora (with a link to the recording) and *phonetic* corpora ». Cette typologie peut s'appliquer aux corpus oraux en général. Le premier type de corpus

7. *Interactive Atlas of Romance Intonation*. Web page: http://prosodia.upf.edu/iari/.

8. Lors d'une *maptask*, deux participants sont mis en situation de communication avec une tâche précise : ils ont tous deux une carte (simpliste, voir l'exemple du corpus IARI dans le volume des travaux) et un itinéraire est présent sur l'une des deux cartes. Le guide (qui a l'itinéraire) doit indiquer le chemin à l'autre locuteur (le suiveur). Les locuteurs n'ont pas accès à la carte de l'autre. Sans que les locuteurs ne le sachent, les deux cartes sont en fait légèrement différentes, ceci afin de créer des situations de communication intéressantes. Un exemple de *maptask* est donné en annexe 1. On trouvera la *maptask* de IARI dans le volume des travaux.

intéresse les chercheurs qui travaillent uniquement sur les scripts, sans recours au son. C'est le cas du corpus ICE-Ireland (Kirk 2003), collecté sous la houlette de John Kirk, qui n'a pas les droits de diffusion des fichiers son. Le deuxième type de corpus est caractérisé par l'accès possible au son mais on ne dispose pas d'annotation en mots ou en syllabes alignée au signal. C'est le cas du corpus PAROLE[9] (Hilton 2008) par exemple. Enfin, les corpus phonétiques sont annotés de façon beaucoup plus précise, avec des transcriptions alignées au signal, et la possibilité d'étudier des phénomènes acoustiques et prosodiques. Ballier & Martin (à paraître a) donnent les exemples de LeaP (Milde & Gut 2002) ou d'Anglish (Tortel 2008) et aussi d'AixOx.

Je pense pour ma part que ce sont les circonstances particulières dans lesquelles la parole a lieu qu'il faut prendre en compte. Je n'entends pas ici les phonostyles au sens de Léon (1993), mais les différents styles de parole que l'on peut chercher à enregistrer pour un corpus oral, rejoignant Ballier & Martin (à paraître a) en ce que les questions de recherche sous-jacentes sont essentielles. Voici quelques exemples de styles de parole (la liste des styles est non exhaustive et quelques corpus en anglais sont donnés à titre d'exemples, mais il en existe d'autres bien entendu) :

- Lecture de textes : AixOx
- Lecture de phrases : IViE[10] (Grabe 2004)
- Listes de mots : PAC
- Parole semi-spontanée (radio, cours): MARSEC, ICE, BNC[11] (Burnard & Aston, 1998)
- Parole spontanée contrôlée :
    - Maptask (IViE, IARI)
    - Story-telling from memory[12] (IViE)

9. http://www.umr7023.cnrs.fr/sites/sfl/IMG/pdf/PAROLE_manual.pdf
10. Intonational Variation in English
11. British National Corpus
12. Cette tâche consiste en général à faire lire ou visionner ou écouter une histoire au locuteur, qui doit ensuite la restituer à sa façon.

  - Interviews (MARSEC, PAC)
- Parole spontanée non contrôlée : PAC, COEC, ICE, CHILDES [13] (McWhinney 2000).

J'ai parlé plus haut de la nécessité d'une réflexion en amont de tout travail de constitution de corpus. Il est important de savoir ce sur quoi l'on veut travailler pour savoir quel style de parole on va préférer. Pour ne donner que quelques exemples, notre étude sur l'accentuation des dissyllabes (Turcsan & Herment 2013) se prêtait à la lecture puisque nous avons fabriqué des logatomes (voir Herment & Turcsan 2013). Le travail sur l'intonation des structures de focalisation (voir mes travaux avec L. Leonarduzzi) nécessitait par contre de la parole spontanée, contrôlée ou non. L'étude de l'intonation du Donegal entreprise sur le corpus PAC (Turcsan & Herment, à paraître) a été réalisée à la fois sur la lecture du texte et sur des extraits de conversation. Pour le corpus IARI, une map-task a été enregistrée.

Une autre distinction est faite par certains auteurs entre « corpus de validation » et « corpus heuristique » :

> Un corpus peut servir deux propos au moins : ou bien il donne à un ensemble de faits une structure qui dégage la vue du linguiste dans une situation où celui-ci poursuit une question mais n'a pas de solution, ou alors il teste la validité empirique d'hypothèses qui existent pour des raisons indépendantes. On peut donc parler de corpus heuristiques dans le premier, et de corpus de validation dans le second cas : l'un montre le chemin, l'autre démontre qu'une hypothèse est bonne (ou mauvaise). (Scheer 2004 : 153-4)

Là encore, la frontière me semble floue. Parmi les corpus que j'ai montés, le Aix-Di-Nonce peut être considéré heuristique (*cf.* Herment & Turcsan 2013) : en fabriquant des logatomes dissyllabiques et les faisant lire par des locuteurs dans des phrases porteuses, nous avons cherché à comprendre comment marchait l'accent de mot sur les dissyllabes et nous sommes partis à la rencontre de l'inconnu, même si des théories sous-tendaient la constitution de nos logatomes. Les corpus heuristiques s'inscrivent donc dans une démarche ascendante, *bottom-*

13. Le CHIld Language Data Exchange System : http://childes.psy.cmu.edu/.

up : « on part des observations pour en établir d'abord une classification, puis une théorie dont le ou les modèles vont rendre compte de la classification » (Martin 2009 : 94). Dans les corpus phonologiques tels que PAC, l'approche est plutôt celle d'un « corpus de validation » et il s'agit d'une démarche descendante, *top-down* : on part « d'une hypothèse "raisonnable" inspirée par une première observation des données, quitte à la revoir et la modifier si de nouvelles données viennent infirmer la validité de la première » (Martin 2009 : 94). C'est ce que l'on fait lorsque l'on fait lire les listes de mots et le texte dans PAC. Mais qu'en est-il d'un corpus de parole spontanée ? C'est là que la frontière est poreuse et que l'on peut tout à fait utiliser le corpus dans les deux sens. On peut privilégier l'approche hypothético-déductive (descendante) et chercher à valider une hypothèse. On peut aussi simplement observer et déduire. Tout va dépendre encore une fois de ce que l'on cherche à faire avec le corpus. C'est la raison pour laquelle un corpus peut tout à fait être à la fois heuristique et de validation. La distinction ne me semble donc pas satisfaisante non plus.

La réflexion sur le style de parole et sur le type de corpus est donc déterminante, comme nous venons de le voir, mais un autre facteur extrêmement important et souvent laissé de côté est celui de la qualité sonore car il va aussi finalement déterminer beaucoup de paramètres.

1.1.3. Qualité sonore

La qualité du son est essentielle puisqu'elle va induire la possibilité de faire certaines analyses ou pas.

Si l'on a des bruits de fond trop importants ou des chevauchements entre locuteurs trop fréquents par exemple, il sera impossible d'effectuer des analyses acoustiques. C'est le cas du corpus ICE-GB. Se voulant un corpus de parole spontanée, il contient des enregistrements qui ont été réalisés dans des endroits publics comme des bars. On entend de la musique ou le brouhaha de fond typique d'un tel lieu. D'autres enregistrements ont pris place dans une cuisine : on entend le bruit d'une personne qui fait la vaisselle ou encore le cliquetis des couverts. Certains parlementaires ont été enregistrés et leurs voix sont parfois couvertes par

les protestations ou les rires des autres membres du Parlement. Il est donc impossible de faire quelque analyse acoustique que ce soit sur ce corpus. Il n'en reste pas moins extrêmement intéressant car il est annoté syntaxiquement (Wallis *et al.* 1999), et c'est la raison pour laquelle nous l'avons utilisé pour plusieurs de nos travaux sur les relations entre prosodie et syntaxe (Herment 2011, Herment & Leonarduzzi 2012, à paraître, Leonarduzzi & Herment 2013a, 2013b).

Toujours au niveau de la qualité sonore, l'environnement est important comme nous venons de le voir, et l'appareil avec lequel les enregistrements sont faits est aussi déterminant : si l'appareil est de mauvaise qualité, le son sera inexploitable. Même si l'appareil est de qualité professionnelle, il faut être vigilant et penser à tout, comme au bruit du transformateur quand l'appareil est branché sur l'électricité. C'est ce qui est arrivé à Gabor Turcsan pour l'enregistrement d'un point d'enquête PAC en Irlande du Nord : le micro était trop près du transformateur et un bruit parasite est audible. L'un des ingénieurs du Laboratoire Parole et Langage, Thierry Legou, a dû intervenir pour effacer ce bruit afin que nous puissions faire des mesures acoustiques correctes.

Une des solutions consiste alors à enregistrer des locuteurs dans un environnement idéal tel qu'une chambre sourde ou un studio d'enregistrement. La qualité sera irréprochable. Cependant, on va se trouver confronté au problème de la situation de communication. Que cherche-t-on à étudier ? S'il s'agit de parole lue par exemple, l'enregistrement en chambre sourde présente tous les avantages. C'est la solution pour laquelle nous avons opté pour le corpus AixOx puisqu'il s'agit uniquement de lectures. Or même en choisissant une solution *a priori* idéale pour la qualité sonore, nous avons rencontré des problèmes de bruit de fond dans AixOx pour les enregistrements faits à Oxford, bruit dû certainement à une soufflerie ou un ventilateur. Quant au corpus Aix-Di-Nonce, qui consiste aussi en lectures uniquement, il a été enregistré dans le studio d'enregistrement des laboratoires de langues de l'Université Aix-Marseille (alors Université de Provence), dans lequel la qualité sonore est excellente.

Si l'on veut travailler sur la parole spontanée par contre, il va être difficile de recréer une situation de communication naturelle dans une chambre sourde ou

un studio d'enregistrement. L'environnement feutré de la pièce close et la présence ostentatoire des micros rend la situation hautement artificielle. Il est très compliqué d'éliciter une conversation informelle dans un tel environnement. Même dans un lieu familier comme la maison des locuteurs (comme c'est souvent le cas dans les points d'enquête PFC et PAC par exemple), le paradoxe de l'observateur (Labov 1972) est inévitable : afin d'obtenir de la parole aussi naturelle que possible, le chercheur doit observer la façon dont les gens parlent quand ils ne sont pas observés. Plusieurs techniques peuvent être utilisées pour essayer de résoudre ce paradoxe méthodologique : on peut par exemple faire appel à un membre de la communauté du (ou des) locuteur(s) pour faire les interviews. C'est l'idée des réseaux sociaux appliquée par les Milroy lors de leur enquête à Belfast (Milroy 1980), et dont le corpus PAC s'inspire pour sa méthodologie. C'est aussi ce que j'ai fait avec le corpus COEC, enregistré en 1999 dans le studio d'enregistrement de la faculté : j'ai choisi quatre locuteurs que je connaissais et qui travaillaient tous à l'université. Ils se prêtent à une conversation informelle. Ils sont un peu crispés au début mais se détendent rapidement et leur parole devient finalement très naturelle. Ce sont des lecteurs de l'université, habitués à cette salle, ainsi qu'une enseignante de phonétique et phonologie, elle aussi à l'aise face à un micro. Ce corpus pêche cependant par le trop grand nombre de locuteurs. Du fait qu'ils sont quatre, il y a des chevauchements assez souvent et beaucoup de rires. Mais la qualité de l'enregistrement est très bonne, même si les locuteurs n'étaient pas équipés de micro-casques. J'ai même pu effectuer des mesures d'intensité pertinentes, malgré la réticence bien connue à mesurer l'intensité et à comparer les données entre locuteurs sans micro-casques (voir la discussion dans Herment à paraître b). Bertrand *et al.* (2008) ont encore mieux réussi avec le CID (Corpus of Interactional Data) : ils ont filmé des conversations très naturelles et spontanées entre deux locuteurs à chaque fois (équipés de micro-casques), dans la chambre sourde du Laboratoire Parole et Langage. La raison pour laquelle le pari a été gagné est que les locuteurs étaient tous membres du Laboratoire et donc familiers des lieux. Les locuteurs ont aussi été assemblés par paires sciemment : ils se connaissaient très bien et avaient donc

matière à se raconter et à discuter. Le choix des locuteurs est donc extrêmement important et joue une part déterminante dans la constitution de corpus.

1.1.4. Choix des locuteurs

Ce choix est bien entendu dicté par l'objectif poursuivi par la constitution du corpus. Le CID est un exemple intéressant pour le côté naturel du corpus. Le but était l'étude de l'interaction et il était donc nécessaire d'obtenir une parole la plus spontanée possible dans des conditions d'enregistrement optimales. Le choix des locuteurs peut être déterminé par d'autres facteurs, extrêmement divers : la langue bien entendu, mais aussi le genre, l'âge, la classe sociale, la variété parlée, la pathologie, etc. Le choix des locuteurs va parfois déterminer le style de parole. On peut difficilement donner une tâche précise à des enfants très jeunes par exemple. Ainsi le corpus CHILDES (sur lequel l'un de mes étudiants de Master 2 a travaillé, *cf.* Oliviéro 2012) est constitué d'enregistrements et le plus souvent de vidéos d'interactions entre un enfant et l'un de ses parents dans des situations de tous les jours à la maison : les repas, les jeux, etc. Si intéressant que ce corpus fût pour un tas d'analyses sur l'acquisition du langage ou encore sur le « *child-directed speech* », il est très difficile de faire des analyses acoustiques sur un tel corpus et en particulier d'étudier l'intonation car beaucoup de bruits extérieurs viennent parasiter les enregistrements.

Pour donner un autre exemple très différent, l'Atlas Interactif des Langues Romanes s'intéresse à une comparaison de l'intonation inter-langues mais aussi intra-langues avec les accents régionaux. Le choix des locuteurs est donc très important. Des consignes précises sont données bien entendu. Les locuteurs doivent être nés et avoir vécu à l'endroit dont la variété est étudiée toute leur vie ou au moins la plus grande partie de leur vie. Le chercheur qui collecte les données doit néanmoins être vigilant. Pour le point d'enquête Marseille que j'ai réalisé en 2013, plusieurs locuteurs répondaient aux critères établis, mais certains n'avaient pas un accent méridional particulièrement marqué. J'ai donc logiquement choisi ceux qui me semblaient les plus représentatifs. Le genre

féminin était également préféré, et l'âge des locutrices indiqué, une locutrice devant être âgée de plus de 65 ans et les deux autres avoir entre 25 et 45 ans.

Dernièrement les corpus d'apprenants se sont beaucoup développés (voir Ballier et Martin 2013 ou Herment *et al.* à paraître). Les locuteurs sont choisis en fonction de leur langue maternelle (L1) et de la langue qu'ils apprennent (L2). Le niveau de l'apprenant est également à prendre en compte et pour cet aspect le CECR, Cadre Européen Commun de Référence (Council of Europe 2001), fournit une aide précieuse. Le corpus AixOx contient une large partie de données d'apprenants (les deux tiers), regroupés en deux niveaux selon le CECR afin de permettre des comparaisons : les utilisateurs indépendants (B1/B2) et les utilisateurs autonomes (C1/C2). Nous avons rencontré pour ce corpus une autre difficulté d'ordre purement logistique, mais qui fait partie des choses à prendre en compte lors de la constitution d'un corpus : tout simplement trouver des locuteurs disponibles. Dans le cas d'AixOx, les critères pour choisir les locuteurs étaient bien définis par le niveau en L2 qu'ils devaient avoir. Or il s'est avéré très compliqué de trouver des locuteurs indépendants (B1/B2). Nous avons enregistré nos étudiants avancés pour le niveau C1/C2, des lecteurs de l'Université pour les anglophones natifs, nos jeunes collègues de la faculté pour les francophones natifs, mais nous n'avions personne « sous la main » pour le niveau B1/B2. Nos étudiants non-spécialistes d'anglais de la faculté des lettres et sciences humaines se sont avérés avoir un niveau trop élevé. Il a donc fallu avoir recours à notre réseau social. Trouver des personnes ayant entre 20 et 30 ans avec un niveau B1/B2 n'a pas été la chose la plus ardue. Le plus gros problème a été de trouver un moment où ils pouvaient se déplacer et venir au Laboratoire pour être enregistrés dans la chambre sourde aux heures d'ouverture du laboratoire. Les personnes que nous avons contactées travaillaient pour la plupart et ne pouvaient venir que le soir ou le week-end, quand la chambre sourde était fermée. Après toutes les questions méthodologiques que l'on se pose pour constituer un corpus, ces considérations logistiques peuvent paraître futiles mais elles ne le sont nullement car elles conditionnent tout bonnement la collecte du corpus. Selon le nombre de locuteurs, plusieurs semaines voire mois peuvent s'avérer nécessaires

pour réunir tous les locuteurs. Quant au *learner's corpus of reading texts*, il s'agit d'un corpus de textes lus par des apprenants français d'anglais, et là encore ce sont nos étudiants (de première année de licence) qui se sont prêtés au jeu. Les enregistrements ont eu lieu dans le studio des laboratoires de langues. J'utilise ce corpus pour certains de mes cours, un en particulier dans lequel je traite de correction phonétique. Mais ce corpus ne peut malheureusement pas bien être utilisé à des fins scientifiques car aucune information sur les locuteurs n'est disponible. Ces informations sont nécessaires et font partie de la documentation nécessaire à tout corpus.

1.1.5. Documentation

> Sans une documentation jointe, un corpus est mort-né. L'un des dangers de la facilité actuelle à rassembler des textes électroniques est précisément que les objectifs du regroupement ainsi que ceux des annotations effectuées ne soient pas enregistrés : le corpus cesse d'être utilisable dès que se perd la mémoire de ces choix. (Habert *et al.* 1997 : 156)

Comme je l'ai dit plus haut, pour les corpus oraux, il est évident que le travail ne s'arrête pas une fois les données enregistrées. Pour que le corpus soit utilisable, il faut qu'il soit transcrit, au moins orthographiquement, j'y reviendrai. Il faut aussi qu'un certain nombre d'informations permettant « d'assurer la traçabilité et la poursuite des recherches » (Baude *et al.* 2006 : 25) figure quelque part. Ces informations, appelées métadonnées, concernent plusieurs paramètres :

- le locuteur : sexe, âge, origine géographique, niveau d'études, langues parlées, etc. On fait en général remplir une fiche aux locuteurs sur laquelle ils donnent toutes ces informations. La fiche peut être plus ou moins complète, en fonction des informations qui semblent les plus importantes au chercheur. Cependant, dans la perspective d'un partage des données et de leur réutilisation possible par d'autres scientifiques, il me semble que plus la fiche est complète, mieux cela vaut. Certaines informations non pertinentes pour une étude peuvent l'être pour une autre. C'est la raison pour laquelle j'ai choisi de faire remplir une fiche très détaillée, celle utilisée dans le protocole PAC, pour le corpus AixOx

par exemple. Cette fiche est donnée en annexe 2. Pour IARI, une autre fiche de renseignements a été donnée aux locuteurs (annexe 3) moins complète celle-ci.

- la situation d'interlocution : elle peut être une métadonnée importante dans le cas des interviews et des conversations en particulier : les informations sur les relations entre les différents acteurs de la conversation ou de l'interview sont déterminantes pour classifier le niveau formel ou informel de la parole. Une interview dans le cadre d'une recherche de travail (l'équivalent de l'ANPE) comme l'on en trouve dans ICE sera complètement différente des interviews de PAC où les interlocuteurs se connaissent bien (faisant partie du même réseau social, voir plus haut).
- la situation d'enregistrement : le lieu et le cadre des enregistrements doivent être précisés. Comme je l'ai dit précédemment, un confinement en chambre sourde peut être stressant pour un locuteur qui n'y est pas habitué. Un enregistrement dans une école (comme dans IViE) peut poser des problèmes de bruits parasites (brouhaha d'une cour d'école, retentissement de la sonnerie). Les tâches ou exercices à effectuer et les consignes données doivent aussi être mentionnés.
- le matériel utilisé et le support : lorsque l'on fait des enregistrements, le choix du matériel est crucial pour la qualité sonore. En studio ou en chambre sourde, la qualité sera normalement excellente (voir 1.1.2), il faut néanmoins toujours préciser le type de matériel utilisé : micro, carte son, ordinateur, logiciel. Pour les enregistrements hors laboratoire, on dispose aujourd'hui d'excellents matériels dont il faut donner les caractéristiques. Il faut faire bien attention à ne pas enregistrer au format mp3, qui « déforme l'enregistrement par codage. Le format PCM (wav) représente plus fidèlement le signal issu du microphone » (Martin 2009 : 31), et permettra ainsi de faire des analyses acoustiques plus fiables.
- Les détails techniques, comme mono ou stéréo, la fréquence d'échantillonnage et le format de conversion, font aussi partie des

métadonnées à fournir. Il n'est pas nécessaire d'enregistrer en stéréo. Cela augmente inutilement la taille des fichiers. « La fréquence d'échantillonnage optimale pour l'enregistrement de la parole est de l'ordre de 16000 Hz (16000 échantillons numériques du signal par seconde) » (Martin 2009 : 31). Le format de conversion « détermine le niveau de bruit apporté par la conversion analogique/numérique de l'enregistreur et donc le rapport signal/bruit résultant. […] Un format de 16 bits est tout à fait adéquat » (Martin 2009 : 32).

Outre les éléments décrits ci-dessus, la documentation est censée également apporter des informations sur les conventions d'annotation (Habert *et al.* 1997 : 156). J'y reviendrai en 1.2.

Comment faut-il encoder ces informations ? Les linguistes et informaticiens travaillant dans le traitement automatique des langues (TAL) et les experts du groupe EAGLES préconisent des normes d'encodage, comme par exemple la TEI[14] (*cf.* Ide & Véronis 1995), qui est un encodage XML. Je n'utilise pas encore ces types d'encodage et mes métadonnées se trouvent dans un ou des fichier(s) annexe(s) partageable(s). Le logiciel d'aide à l'enregistrement ROCME[15], conçu sous la férule d'Emmanuel Ferragne, code les métadonnées en XML et me permettrait de mettre mes données à jour.

Un autre point extrêmement important concerne le cadre juridique, que je ne développerai pas (*cf.* Baude *et al.* 2006 pour une discussion approfondie de ces aspects), mais qu'il faut quand-même mentionner et dont il ressort surtout qu'il est absolument nécessaire d'obtenir un consentement éclairé de nos locuteurs. Deux exemples sont donnés en annexe 4 et 5. Le premier est celui que nous utilisons au Laboratoire Parole et Langage pour tous nos enregistrements et que nous avons donc utilisé pour AixOx et Aix-Di-Nonce (pour COEC et learner's corpus of reading texts, nous ne disposons malheureusement pas de ces documents : il y a 10 ans, ces questions d'ordre juridiques n'étaient pas encore

14. *Text Encoding Initiative*
15. http://www.ddl.ish-lyon.cnrs.fr/Download/Index.asp?Action=Edit&Langue=FR&Page=ROCme

bien connues). Le deuxième est le consentement utilisé pour IARI. Les pratiques actuelles de recueil de consentement ne sont pas encore uniformisées. Elles sont plus avancées lorsque le corpus est multimodal ou met en scène des enfants, par exemple pour le projet COLAJE[16] d'Aliyah Morgenstern.

La confidentialité des données est également à prendre en compte et concerne l'approche éthique du chercheur. Même si les locuteurs ont donné leur consentement éclairé, il apparaît que dans certains cas ils parlent de choses ou de personnes sans penser que l'enregistrement peut être diffusé. Cela peut poser des problèmes éthiques, il faut donc y faire extrêmement attention, au risque de ne pas pouvoir exploiter les données.

Une documentation complète et encodée dans les formats suggérés par les experts permet de faire des requêtes, qui sont indispensables pour une utilisation optimale du corpus. Ballier & Martin (à paraître a) distinguent ainsi « base de données » et « corpus », la première étant un corpus documenté dans lequel on peut faire des requêtes.

1.1.6. SLDR, mise en forme et curation des données

Avec le développement technologique et l'essor de la linguistique de corpus, des bases de données numériques de plus en plus larges voient le jour. Il est essentiel que ces données soient sauvegardées. Nous avons aussi vu plus haut qu'un corpus doit pouvoir être réutilisable. Il est crucial de partager les données pour faire avancer la recherche. Le Laboratoire Parole et Langage héberge le SLDR[17], qui « est un centre numérique de qualité certifiée (*Trusted Digital Repository*) permettant aux laboratoires et aux chercheurs indépendants de partager librement leurs données orales/linguistiques tout en assurant leur archivage pérenne » (Bell & Bénard 2013). Il faut donc penser à déposer son corpus sur le SLDR, afin d'en assurer l'archivage. Pour ce faire, les données doivent être mises en forme correctement et la première chose à faire est

16. http://colaje.scicog.fr/
17. *Speech and Language Data Repository* : http://www.sldr.org

d'anonymiser les données. La façon de nommer les fichiers son est un point extrêmement important lorsque l'on monte un corpus, ne serait-ce que pour s'y retrouver rapidement et pour permettre une réutilisation plus facile, mais qui est souvent négligé. Il faut pourtant y faire extrêmement attention si l'on veut éviter une perte de temps. On se retrouve parfois obligé de renommer tous les fichiers parce que l'on complète le corpus ou bien parce que le nom ne convient finalement pas. C'est ce qui nous est arrivé pour AixOx. Le nom donné aux fichiers ne permettait pas de savoir s'il s'agissait d'un locuteur ou d'une locutrice. Nous avions au départ un seul groupe d'apprenants et avons décidé ensuite d'en faire deux. Il fallait donc le niveau de l'apprenant soit reflété dans le nom des fichiers. Nous avons donc renommé tous nos fichiers, ce qui a été très long et fastidieux car nous avions commencé les annotations et avons dû renommer tous les fichiers son mais aussi tous les fichiers d'annotation les accompagnant. Rédiger des scripts sous Perl s'est révélé quasiment aussi long que de renommer les fichiers manuellement.

Le travail énorme d'aide à la mise en forme et à la préparation des données que proposent les ingénieurs du SLDR s'appelle la curation de données. Nous, linguistes, nous attelons à la collecte des données avec tout le travail de réflexion mentionné plus haut et les théories sous-jacentes à cette réflexion. Nous effectuons ensuite les analyses. Les « curateurs » s'occupent matériellement des données, des métadonnées, de l'archivage, du cadre juridique, des demandes des utilisateurs. La curation est essentielle : elle permet d'améliorer la qualité des données, de les rendre accessibles, partageables et réutilisables. Elle permet enfin d'en améliorer la longévité. Le chercheur qui constitue des corpus et accepte de les partager avec la communauté scientifique ne peut pas se passer de curateurs ni d'une plateforme comme le SLDR. L'avancée technologique dont nous sommes les témoins ces dernières années nous permet de bénéficier de la mise en œuvre progressive de bonnes pratiques et nous devons en profiter.

1.1.7. Proposition de typologie complétée

Devant toutes les contraintes décrites précédemment, il semble bien difficile de proposer une nouvelle typologie. Celle que défendent Ballier & Martin (à paraître a) pour les corpus d'apprenants est applicable aux corpus de locuteurs natifs. Leur typologie est fondée sur les différents types d'annotation et de requêtes que l'on peut faire. J'ai montré pour ma part que les styles de parole ont une importance capitale en ce qu'ils déterminent grandement la qualité sonore et le choix des locuteurs. On pourrait alors proposer de compléter la typologie de Ballier & Martin en s'appuyant sur les styles de parole, comme suggéré en 1.1.2 :

- lectures
- semi-spontané
- spontané contrôlé
- spontané non contrôlé.

On aurait donc par exemple des corpus phonétiques de lectures, ou bien des corpus muets de spontané non contrôlé, ou encore des bases de données phonétiques de spontané contrôlé et même des corpus phonétiques de lectures, spontané contrôlé et non contrôlé comme PAC.

1.2. Transcription et outils

Comme on l'a vu plus haut (1.1.1), un corpus selon la définition de Gibbon *et al.* (1997 : 79) est un ensemble d'enregistrements dotés d'une documentation et d'annotations. L'annotation minimale pour un corpus oral est la transcription orthographique : « des données audio brutes ne peuvent pas être analysées sans un travail préalable de transcription et de segmentation » (Delais-Roussarie 2003 : 98). Quant à la segmentation, elle s'automatise de plus en plus, mais la transcription orthographique reste la base nécessaire de toute segmentation pour l'interprétation des données.

1.2.1. Transcription orthographique

Le problème de la transcription orthographique fait débat et occupe une grande place dans la littérature (voir Delais-Roussarie & Post à paraître pour un panorama). Il se pose principalement pour les corpus de parole spontanée ou semi-spontanée. De façon largement consensuelle, les grands corpus oraux sont transcrits en orthographe standard afin d'en faciliter la lecture, l'homogénéité et la recherche de données. Il existe aujourd'hui des outils comme PRAAT (Boersma & Weenink 2001), Winpitch (Martin 2005) et SPPAS[18] (Bigi 2012), sur lesquels je reviendrai plus bas, qui facilitent la transcription dans le sens où l'on peut écouter le son et écrire simultanément les mots dans des fichiers comme les TextGrid de PRAAT. Mais la transcription reste encore largement manuelle car les logiciels de transcription existants ne sont pas encore assez performants pour gérer les chevauchements par exemple, certaines liaisons, certaines ambiguïtés grammaticales ou syntaxiques (*cf.* Blanche-Benveniste & Jeanjean 1987), ou culturelles et contextuelles (*cf.* Delais & Post à paraître pour une description des outils d'annotation les plus distribués). Dans le cadre du projet PAC par exemple, la personne qui a assisté aux enregistrements est la mieux placée pour transcrire car elle connaît le contexte d'interlocution et cela facilite souvent l'interprétation des données sonores.

Le problème des conventions de transcription est épineux car la transcription engage toujours une théorie (Ochs 1979). Faut-il se limiter à transcrire ce que les locuteurs disent ou bien faut-il noter les tours de parole, les changements de locuteurs, les hésitations, les faux-départs, les mots tronqués, etc. ? Doit-on noter la ponctuation, qui représente déjà en elle-même une interprétation de la prosodie ? Doit-on marquer les pauses ? Ceci ne concerne que la transcription orthographique, et on peut se demander s'il faut aussi noter les rires, les toux, les bruits extérieurs, les gestes, etc. On part alors sur une transcription multimodale.

Encore une fois, on va souvent opter pour tel ou tel choix en fonction de ce que l'on veut faire avec le corpus. L'essentiel est de préciser ces choix dans la documentation. En général les changements de locuteurs et les tours de parole, les chevauchements, les hésitations, les faux-départs, les mots incomplets, les

18. *SPeech Phonetization Alignment and Syllabification*

phénomènes extralinguistiques comme rires, toux, etc. ou bruits extérieurs (aboiement, klaxon, ou autre) sont retenus. Mais les conventions varient en fonction des corpus, en particulier pour l'annotation des pauses, les phénomènes de sandhi[19] et la ponctuation. Par exemple dans COEC les pauses sont marquées par le signe =, dans AixMarsec par # et dans AixOx les pauses longues par # et les pauses courtes par +. La frontière entre pauses longues et courtes n'est pas clairement définie et reste très subjective dans AixOx. Pour le corpus PAC[20], une ponctuation simplifiée est adoptée et les phénomènes extralinguistiques sont donnés entre parenthèses. Les mots répétés sont notés, séparés par une virgule, un slash indique qu'un mot est tronqué. Par contre les élisions de syllabes, quelles qu'elles soient, ne sont pas transcrites. Les expressions dialectales sont transcrites en alphabet SAMPA[21] et mises entre parenthèses.

Pour les corpus lus comme AixOx, on peut penser que c'est très simple, la transcription correspondant aux textes à lire. C'est en fait loin d'être aussi aisé, tout particulièrement lorsque l'on a affaire à des apprenants. Déjà avec des natifs, la transcription sert de base mais elle doit être corrigée manuellement avec tous les faux départs, hésitations, mots éventuellement incomplets, erreurs de lectures, etc. Avec des apprenants, et en particulier les apprenants qui n'ont pas un niveau très élevé, comme le groupe des locuteurs indépendants (B1/B2) dans AixOx, le problème se corse. Hormis les nombreuses hésitations, pauses pleines ou silencieuses et faux-départs, on entend des erreurs très fréquentes de mauvaise prononciation d'un mot simplement parce qu'il n'est pas connu, comme le mot *obligée* prononcé /oblige/ (fs 1), ou encore *Greenstead* prononcé /grinstid/ (fs 2). On entend aussi souvent en anglais L2 des /h/ insérés dans la prononciation alors qu'ils ne sont pas présents dans la graphie (fs 3). En français L2, les liaisons sont souvent problématiques : on peut entendre des liaisons erronées comme « en_Hollande », d'autres qu'un natif ne ferait certainement pas comme pas « d'abord Paris, puis_ensuite Bruges, Francfort et Rome », d'autres qu'un natif ferait « est-ce que

19. Pour l'anglais et le français, il s'agit principalement des liaisons, enchainements et insertions de schwa ou /r/.

20. *Cf.* les conventions de transcription sur http://www.projet-pac.net

21. L'alphabet SAMPA (*Speech Assessment Methods Phonetic Alphabet*) est un alphabet en caractères ASCII, voir Wells *et al.* (1992).

vous_avez des tarifs spéciaux » et qui ne sont pas réalisées par l'apprenant, enfin des cas ambigus comme « quels hôtels » prononcé « quel hôtel ». Que faut-il transcrire dans ces cas-là ? Comme je le disais plus haut, tout va dépendre de ce à quoi va servir le corpus et de ce que l'on veut en faire. Dans le cas d'AixOx, nous avions besoin d'une transcription phonétique dans le cadre du projet AMENNPRO[22] et nous avons opté pour une phonétisation automatique. Dans cette perspective, une transcription orthographique riche est nécessaire pour éviter que l'automatisation ne comporte trop d'erreurs. Nous avons utilisé le logiciel SPPAS (voir 1.2.2) et une transcription orthographique enrichie que le logiciel pouvait déchiffrer (*cf.* Bigi *et al.* 2012). Ainsi les conventions utilisées sont les suivantes : les pauses longues sont annotées # et les pauses courtes '+', les bruits extérieurs '*', les rires '@', les pauses pleines 'euh' pour le français, 'uh' ou 'uhm' pour l'anglais. Les mots tronqués sont transcrits avec un '-' accolé au mot. On note tous les faux-départs, et hésitations. Et les mots mal dits sont transcrits comme ils sont prononcés : ainsi *obligée* prononcé /oblige/ sera transcrit *obligué* et *Greenstead* prononcé /grinstid/ sera transcrit *Greensteed,* dont l'orthographe se rapproche le plus de la prononciation réalisée. Les liaisons erronées sont notées, mais pas les liaisons possibles, qui sont intégrées comme une variante dans le dictionnaire. Le cas de « quels hôtels » prononcé comme un singulier sera transcrit « quel hôtel » afin que le logiciel ne cherche pas le /z/ en vain.

Voici deux exemples de transcription enrichie dans AixOx (avec en gras les passages intéressants) :

> *Ma femme a une série de* ***réu*** *+ de réunions + le mois prochain entre neuf heures et dix-sept heures # quatre jours de suite et chaque fois + dans une* ***vi*** *différente # d'abord paris + puis ensuite + Bruges Francfort Rome + et pour finir + Hambourg # quels + vols du soir vous lui + conseill****erez*** *+ et quels hôtels # elle a horreur des* ***grands*** *chaînes elle recherche plutôt une ambiance familiale #* fs 4

22. AMENNPRO (*Automated Metrics for the Evaluation of Non-Native Prosody*) est un projet franco-britannique Alliance PHC (Partenariats Hubert Curien) que j'ai porté pour l'Université d'Aix-Marseille avec Anastassia Loukina pour l'Université d'Oxford (2010-2012, financement de 3500 euros par an pour chaque université).

Could you please tell me the best connections to Sheffield from East ***Greensteed****? # I need to arrive by* ***uh*** *# 10* ***uh*** *+* ***13***[23] ***a.n.*** *# on Saturday #* fs 2

Cette transcription enrichie est un gain de temps très important car elle évite le recours à des corrections manuelles trop nombreuses. En effet, si le logiciel ne trouve pas les sons qui correspondent à la graphie, il va faire des erreurs, voulant absolument trouver le son, et cela va avoir des répercussions sur la suite ou sur ce qui précède, entrainant des erreurs en cascade. De la même façon, transcrire une consonne épenthétique dans le cas du fichier son fs 3 donné comme illustration plus haut évitera que le logiciel ne fasse des alignements erronés.

Je n'ai parlé dans cette sous-section que de la transcription orthographique. On peut annoter un corpus de bien d'autres façons, souvent avec l'aide d'outils précieux.

1.2.2. SPPAS et PRAAT

La manière dont un corpus est encodé va conditionner son utilisabilité car tous les logiciels ne lisent pas les mêmes formats. Il est impensable aujourd'hui de ne pas utiliser un logiciel de traitement de la parole pour travailler sur les corpus oraux. Je me sers du logiciel PRAAT pour mes travaux : j'ai été parmi les premiers chercheurs à l'utiliser pour ma thèse, suivant les conseils de Daniel Hirst. Je me rappelle m'être inscrite dans les 100 premiers adhérents à la liste PRAAT en 2000, il y en a aujourd'hui près de 2700. PRAAT est en effet le logiciel le plus utilisé dans le monde de la recherche sur la parole. Le grand intérêt de cet outil est qu'il est multiplateforme : il peut être utilisé sous Linux, Windows et Mac. Du fait que PRAAT est le logiciel le plus utilisé dans notre communauté scientifique, les autres logiciels qui voient le jour depuis ces dernières années font en sorte de pouvoir lire les fichiers TextGrid générés par PRAAT, comme par exemple WinPitch[24], ANVIL[25] ou ELAN,[26] des outils d'annotation multimodale pour la vidéo. C'est également le cas de SPPAS, qui permet en outre d'exporter les TextGrid en CSV (format excel) pour faire des statistiques par exemple.

23. 30 dans le texte.
24. http://www.winpitch.com
25. http://www.anvil-software.org/
26. http://tla.mpi.nl/tools/tla-tools/elan/

PRAAT permet de lire des fichiers audio (on peut aussi en enregistrer à partir de PRAAT) et de les annoter dans un fichier TextGrid qui peut contenir plusieurs tires. Il est ainsi possible de regrouper et de visualiser plusieurs niveaux d'annotation dans un seul fichier. Les niveaux d'annotation vont bien entendu varier en fonction des besoins de l'analyse. Dans Aix-Marsec par exemple, on trouve 8 tires correspondant à des annotations différentes : la segmentation en phonèmes[27], en syllabes, en unités rythmiques selon Abercrombie, puis selon Jassem, la transcription orthographique incluant les TSM (*cf.* 1.3.2) alignée en mots, le découpage en unités intonatives mineures et majeures, le codage Intsint et les valeurs de fréquence fondamentale correspondant aux tons Intsint d'au-dessus :

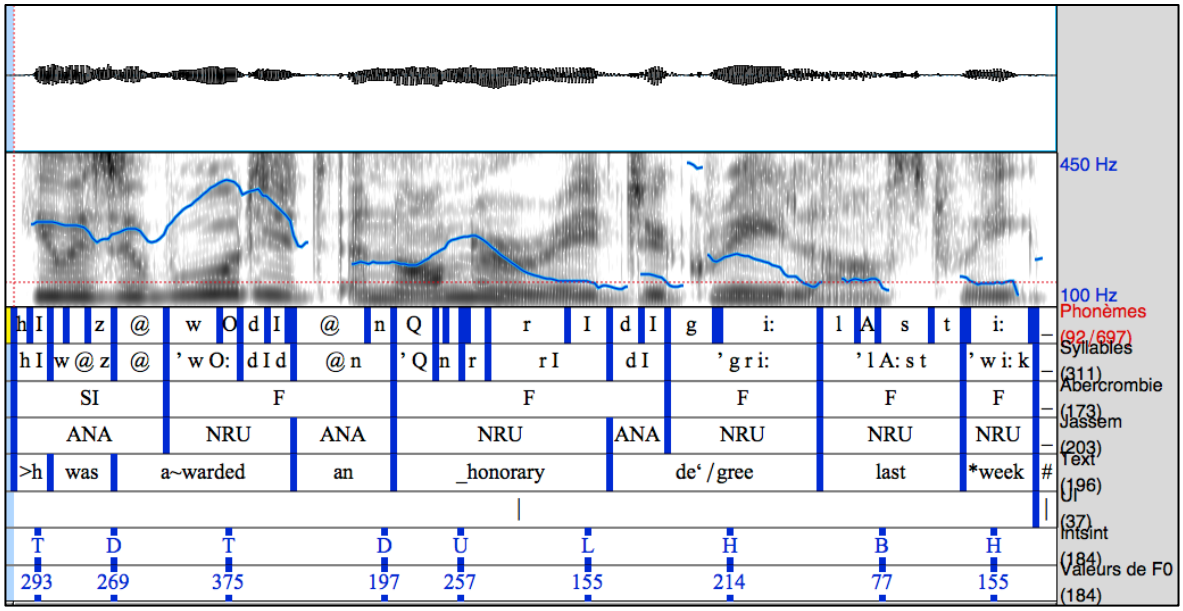


Figure 1 : niveaux d'annotation dans Aix-Marsec fs 5

Dans AixOx, les niveaux d'annotation sont différents : la figure 2 ci-dessous montre 9 tires : le découpage en unités interpausales (avec la transcription orthographique), le découpage en mots (tire nommée *tokens*), en syllabes, en syllabes mais avec les 6 classes de phonèmes composant la syllable[28], en structure syllabique (consonne/voyelle), le découpage en phonèmes, les valeurs de F0

27. Les phonèmes sont transcrits ici en alphabet SAMPA.

28. F pour fricatives, G pour les semi-voyelles (*Glides*), L pour liquide, N pour nasale O pour occlusive et V pour voyelle.

calculées à partir de Momel, le codage de tons Intsint, et les valeurs de F0 recalculées à partir du mapping Momel-Intsint-Momel (*cf.* De Looze *et al.* 2013).

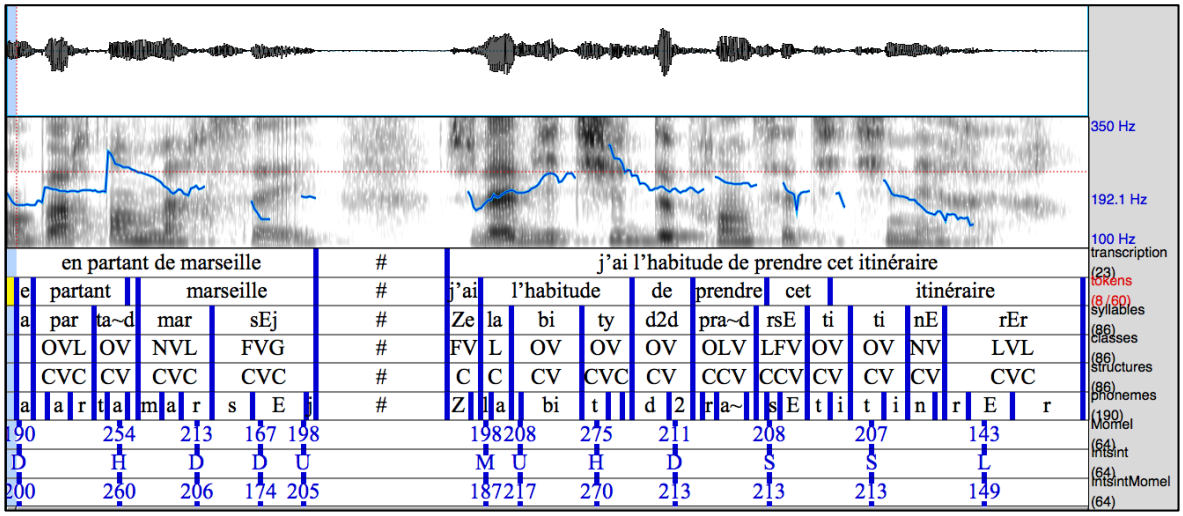


Figure 2 : niveaux d'annotation dans AixOx fs 6

Les tires dans AixOx ont été générées grâce au logiciel SPPAS à partir du son et de la transcription orthographique (*cf.* Herment *et al.* à paraître). SPPAS a été développé au Laboratoire Parole et Langage par Brigitte Bigi en 2010 et cette dernière ne cesse d'en améliorer les fonctions depuis. Nous l'avons tout d'abord utilisé pour sa fonction de phonétisation et d'alignement des phonèmes. Il est très performant sur le français, un peu moins sur l'anglais britannique car les modèles acoustiques qu'il intègre sont largement américains : le modèle acoustique pour l'anglais provient d'une base de données libres, VoxForge[29], constituée d'enregistrements de tout locuteur souhaitant contribuer sur le site. Précisons bien que SPPAS fournit une transcription phonémique. Dans le cas de *Greensteed* donné plus haut, SPPAS transcrira phonétiquement /gri:nsti:d/, avec les voyelles anglaises, et non /grinstid/, qui est la prononciation de l'apprenant français. Cela pose donc problème dans le cas des corpus d'apprenants, puisque les locuteurs d'un niveau moins avancé ont souvent les voyelles et les consonnes de leur L1, ou un son se rapprochant qui n'est pas vraiment celui de la langue cible, ni celui de la L1. Afin de minimiser les erreurs, il est quand-même recommandé de transcrire un mot dont la prononciation se rapproche le plus de celle effectivement réalisée par le locuteur. La solution serait

29. http://www.voxforge.org

d'intégrer dans SPPAS un nouveau dictionnaire de prononciation avec des formes produites par les apprenants. Cela semble ardu à mettre en place pour les apprenants puisqu'outre le fait qu'il faut prendre en compte la langue maternelle et qu'il faudrait donc un nouveau dictionnaire pour chaque nationalité d'apprenants, la prononciation change selon le niveau de l'apprenant. Il est par contre envisageable de travailler sur les variétés dialectales avec SPPAS, puisque la variation inter-locuteurs est moins grande. A partir du recensement des phonèmes d'une variété d'anglais, comme cela est fait dans le cadre du projet PAC par exemple, ou bien à partir d'une variété déjà bien documentée comme l'anglais australien (voir Harrington *et al.* 1999, Cox 2000 entre autres), on pourrait intégrer un dictionnaire de la variété que l'on veut étudier. Le besoin d'un dictionnaire de prononciation pour chaque variété d'anglais traitée vaut également pour les variétés britannique standard et *General American*. Il est intéressant de constater que grâce à la transcription orthographique enrichie, SPPAS est capable de phonétiser et d'aligner du français L2 et de l'anglais L2, même sans l'ajout d'un dictionnaire spécifique. Il reste des erreurs bien entendu, et des corrections manuelles s'imposent, en particulier pour les phonèmes de la L1 plaqués sur la L2. L'oubli d'une pause peut aussi créer des erreurs de segmentation dans SPPAS, il faut donc y être extrêmement attentif.

SPPAS intègre désormais Momel (Hirst *et al.* 2000, voir 1.4.2 pour davantage de détails) et Intsint, on peut donc aussi obtenir les trois dernières tires que l'on voit sur la figure 2 si l'on souhaite faire des analyses de l'intonation, ce qui est bien-sûr très précieux pour moi.

Ayant obtenu des résultats satisfaisants pour la phonétisation et l'alignement dans AixOx, nous avons pensé que SPPAS pourrait être un outil utile dans le cadre du projet PAC et des variations dialectales. Ainsi Brigitte Bigi et moi avons animé un atelier annotation lors du dernier colloque PAC (à Aix-en-Provence les 30 et 31 mai 2013), au cours duquel Brigitte Bigi a présenté les différentes fonctions de SPPAS et j'ai quant à moi montré l'intérêt de telles annotations automatiques pour l'analyse segmentale et suprasegmentale de nos données. En tant que coordinatrice du groupe prosodie de PAC (avec Caroline Bouzon et Stephan Wilhelm), j'ai proposé que nous réfléchissions à intégrer ces annotations dans le protocole PAC pour les conventions

d'annotations. Cela avait déjà été fait, mais pas mis en place, il y a quelques années, uniquement pour la prosodie, non pour la transcription (voir Bouzon *et al.* 2004). Suite à l'atelier tout récent du projet LVTI -Langue, Ville, Travail, Identité- (voir Giraudo *et al.* 2013) qui s'est tenu en juillet 2013 à Toulouse, il est question d'utiliser SPPAS également pour PFC.

1.3. Annotation prosodique

Il existe plusieurs systèmes de transcription prosodique, dont quelques-uns sont brièvement décrits dans l'introduction (*cf.* Delais & Post à paraître pour une revue des systèmes existants) : j'ai mentionné les Tonetic Stress Marks (TSM), ToBI, Intsint et IViE. Certains corpus sont déjà annotés prosodiquement, comme Aix-Marsec qui contient une tire avec la transcription orthographique assortie des TSM (*cf.* figure 1) ou AixOx qui est annoté avec Intsint (*cf.* figure 2). Ces annotations peuvent être très utiles, mais doivent être manipulées avec précaution. En effet, toute annotation implique une part de subjectivité et une théorie sous-jacente.

1.3.1. Subjectivité de l'annotation et influence de la théorie sous-jacente

Pour certaines de mes études sur les relations entre prosodie et syntaxe, j'ai réalisé des analyses auditives de la prosodie (voir chapitre 3). Cela présente des inconvénients que certains relecteurs de conférences internationales n'ont pas manqué de nous reprocher : il est vrai qu'il s'agit d'une analyse subjective. Mais le système de transcription ToBI, prôné ces dernières années dans le monde de la recherche en prosodie (*cf.* Martin 2009 : 83), ne l'est-il pas également ? Wightman (2002), prenant appui sur deux études pour l'anglais (Pitrelli *et al.* 1994 et Syrdal & McGorg 2000) et une pour l'allemand (Grice *et al.* 1996) montre que l'accord inter-annotateurs pour ToBI n'est pas aussi bon que l'on voudrait nous le faire croire :

> In [20][30], for example, pairwise agreement on the specific pitch accent label failed to exceed 50% for six of the eight label types. Likewise, agreement on the specific edge tone label failed to exceed 50% for six of the nine label types. Note that this was under the almost ideal circumstances of very highly and uniformly trained labelers working in laboratory conditions with full access to time-aligned pitchtracks, spectrograms, and waveforms.

30. Pitrelli *et al.* 1994

La question de la subjectivité est très pertinente, mais peut-on y échapper ? Il est clair que dès que l'homme met le nez (ou plutôt l'oreille) dans la machine, une certaine subjectivité va rentrer en ligne de compte et l'on va constater des écarts inter-annotateurs. Martin (2009 : 79-82) propose une réflexion très intéressante sur la relation entre perception et interprétation. Il explique que le transcripteur n'étant pas un auditeur naïf, il va pouvoir entre autres possibilités

> réécouter à loisir les segments à transcrire (parfois à vitesse ralentie par resynthèse), être influencé par une théorie sous-jacente à sa démarche pour prendre une décision difficile sur tel ou tel détail de la transcription, être influencé par la grille de perception de sa langue maternelle, voire par le volume de reproduction sonore. (Martin 2009 : 80)

Je ne suis pas sûre que l'on ne constate pas également quelques écarts chez le même annotateur qui aurait transcrit le même morceau de parole à quelques mois d'intervalle. Il serait intéressant d'en faire l'expérience.

Tout comme il n'y a pas de transcription orthographique sans théorie, il n'y a pas d'annotation prosodique sans théorie. ToBI par exemple est totalement dépendant d'un cadre phonologique précis. Intsint ne déroge pas à la règle : par exemple le choix des 8 tons qu'il utilise relève d'une réflexion théorique en amont. Comme le précisent Ballier & Martin (à paraître a), cela est acceptable d'un point de vue scientifique, du moment que l'approche théorique qui a servi à établir les conventions d'annotation est transparente.

Je me réclame pour ma part de l'école britannique, ainsi que je l'ai expliqué dans l'introduction de ce document de synthèse, et je propose de montrer qu'au sein d'un même modèle, des divergences peuvent apparaître quand-même, qui sont dues à des discordances théoriques, ainsi que la sous-section suivante propose de le montrer.

1.3.2. Système d'annotation et exemple de discordance avec les TSM

Je souhaite exposer ici en détails la conception théorique qui sous-tend ma propre façon d'annoter l'oral. A cause d'une mauvaise qualité acoustique, on est parfois obligé de faire des analyses uniquement auditives des corpus. Cela présuppose une théorie sous-jacente. L'école britannique fait largement usage des TSM : deux aspects de l'intonation y sont rendus, les tons et les frontières. Les

syllabes accentuées sont marquées d'un signe diacritique qui indique quel mouvement mélodique commence sur ou tout près de ces syllabes. Trois tons simples (*fall*, *rise* et *level*) et deux tons complexes (*fall rise* et *rise fall*) existent. D'autres signes diacritiques sont aussi utilisés pour marquer des excursions inhabituelles de F0 ou encore les pauses. La parole est divisée en unités intonatives et deux types de frontières sont distinguées : les frontières mineures et majeures. Les frontières mineures apparaissent en fin de phrase syntaxique. Les symboles sont reproduits dans le tableau 1 ci-dessous.

_	low level
~	high level
<	step-down
>	step-up
/'	(high) rise-fall
'/	high fall-rise
\	high fall
/	high rise

,	low rise
'	low fall
,\	(low rise-fall – not used)
\,	low fall-rise
*	stressed but unaccented
|	minor intonation unit boundary
||	major intonation unit boundary

Tableau 1 : Tonetic Stress Marks dans Aix-Marsec

Ainsi on peut déchiffrer la prosodie de la phrase suivante :

*The ~Overseas De,velopment *Minister | is _visiting Ethi '/opia this week*end | to *see the *famine re'lief ope*ration ||* fs 7

En ce qui me concerne, je me démarque de la théorie sous-jacente aux TSM tout d'abord au niveau des unités intonatives, en ne considérant, à l'instar de Cruttenden (1998) ou Grabe *et al.* (2001), qu'un seul type de frontière. Je pense en effet qu'en discours spontané, les phrases ne sont plus du tout les mêmes qu'en discours oralisé comme Aix-Marsec. La syntaxe de l'écrit vole en éclat en parole spontanée et la notion même de phrase est remise en cause (voir les travaux du GARS[31] sur la syntaxe de l'oral en français sous l'égide de Claire Blanche-

31. Groupe Aixois de Recherches en Syntaxe.

Benveniste et José Deulofeu ; Sabio 2011 entre autres). La notion de frontière majeure et mineure n'a donc plus lieu d'être.

Pour les tons, je distingue les tons neutres, *fall* et *rise*, des tons non neutres, *high fall* et *fall rise*. Je m'écarte ici légèrement de la tradition britannique et de la série des 7 tons de Cruttenden (1986), dont il reconnaît lui-même qu'elle n'est pas entièrement satisfaisante, pour suivre l'approche tri-tonale de Gussenhoven (1983, 1984), qui considère que les trois tons principaux de l'anglais sont *fall*, *rise* et *fall rise*. Il prend en compte la gamme mélodique, c'est-à-dire l'amplitude et la hauteur du mouvement mélodique, définissant le *high rise* et le *high fall* par opposition au *low rise* et *low fall*. Je distingue moi aussi des tons statiques hauts et bas, qui correspondent au *high level* et *low level* des TSM. Ces tons sont en général perçus sur la première syllabe accentuée de l'UI (voir chapitre 2), et il s'agit plus souvent d'un ton haut, ceci étant dû au réajustement typique (*reset*, voir 2.2.2) d'un début d'UI. Le noyau est le dernier accent mélodique de l'UI (voir entre autres Cruttenden 1998, Wells 2006) et les syllabes qui suivent le noyau sont donc désaccentuées (dans le sens du *deaccenting* de Ladd 1980) et ne portent pas d'accent. Par contre les mots lexicaux (par opposition aux mot grammaticaux) qui précèdent le noyau sont porteurs d'accents de phrase selon moi (voir chapitre 2).

Si je reprends l'exemple fs 7, je suis donc en désaccord pour les frontières d'UI, que je note | uniquement, et également pour les mots *see* et *famine*. Ils sont associés à une étoile qui veut dire *stressed but unaccented*, c'est-à-dire qu'ils gardent leur accent lexical, mais ne sont pas accentués au niveau de la phrase. Ils le sont selon moi et selon la perception que j'ai de l'énoncé. Je ne suis pas d'accord non plus avec le ton *low level* sur *visiting*, que je perçois comme un *high level*. C'est *is* qui est sur un ton très bas, comme le montre la courbe de F0 figure 3 : on voit un ton statique plus haut sur la première syllabe *visiting*.

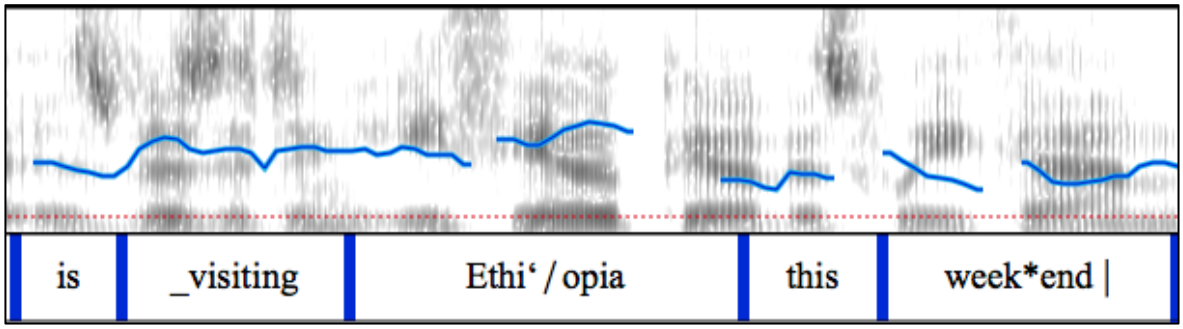


Figure 3 : courbe de F0 pour un extrait de fs 7

Prenons un deuxième exemple, tiré de Aix-Marsec toujours :

In ~recent _months | it's men who've `/benefited | from the *slackening `pace | of _unem`ployment || fs 8

Plusieurs discordances se font jour à nouveau dans cet exemple. Outre les frontières d'UI majeure et mineure, le découpage en UI est différent du mien : je n'aurais divisé qu'en 3 UI :

In recent months | it's men who've benefited | from the slackening pace of unemployment |

Les experts placent une frontière d'unité intonative après *pace* sur laquelle je ne suis pas d'accord. Il y a bien une chute sur *pace* mais qui n'est pas nucléaire à mon avis et relève seulement d'un accent de phrase (*cf.* 2.2.5.). Ici, le contexte joue : sémantiquement, il ne me semble pas très cohérent de couper entre un nom et son complément si la longueur des UI ne l'impose pas. Ceci dit, et pour en revenir à mon propos, tout est question d'annotation car que l'on pose une frontière d'UI ou que l'on considère qu'un accent de phrase peut être prononcé avec un accent mélodique, la prononciation ne change pas. Il ne s'agit donc pas d'un problème de perception mais de théorie sous-jacente et l'on en revient à ce que dit Martin (2009 : 80, cité plus haut) : une théorie sous-jacente peut être influente pour une prise de décision quant à l'annotation de tel ou tel détail. Concernant les accents de phrase, je suis en désaccord avec l'étoile sur *slackening* (pour les mêmes raisons que pour *see* et *famine* dans l'exemple précédent). *Slackening* est la tonique de tête (voir 2.2.4) selon moi, avec un ton statique relativement haut. Enfin, au niveau des mouvements mélodiques, l'annotation de la première UI relève également de

divergences théoriques : pour moi une unité intonative comporte nécessairement un noyau et un seul, qui est porteur d'un mouvement mélodique. Ce ne peut donc être un ton statique. Sur la première UI, les TSM donnent un statique haut sur *recent* et un statique bas sur *months*. Je perçois une petite montée sur *months*. Sous PRAAT, la montée n'est pas franche :

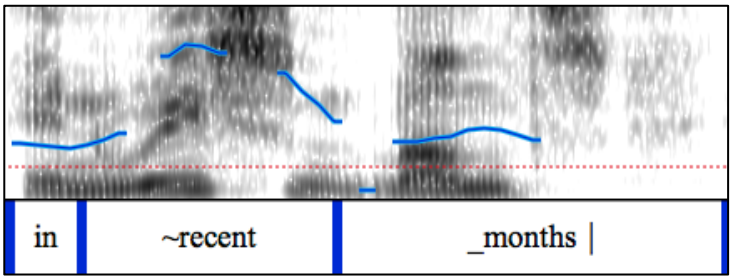


Figure 4 : courbe de F0 pour un extrait de fichier fs 8

Ceci prouve qu'encore une fois ce n'est pas une différence de perception qui est en jeu, mais bien de convention d'annotation.

Pour les travaux que je mène, je n'ai pas besoin pour l'instant d'une annotation aussi complexe que les TSM. Je note les frontières d'UI (|), je souligne la syllabe nucléaire, et j'indique à la fin de l'UI le schéma R, F, HF, ou FR pour les tons. Pour l'analyse du corpus PAC Donegal (Turcsan & Herment à paraître), j'ai distingué le ton RF pour *Rise-Fall* (qui correspond à un type de *fall*, le *delayed fall*, dans la conception de Gussenhoven), car il est récurrent dans cette variété d'anglais.

L'énoncé fs 8 sera donc transcrit comme suit :

In recent months R | it's men who've benefited FR | from the slackening pace of unemployment F |.

1.4. Le primat du perceptuel et l'expérience de l'écoute

1.4.1. Annotation automatique sur grands corpus

L'annotation automatique représente un gain de temps considérable. Il faut cependant être prudent et une vérification et une correction manuelles sont

nécessaires dans le cadre de corpus de moyenne taille. Il va sans dire qu'il est impossible de procéder à des corrections sur des corpus énormes tels que le BNC, dont la partie orale contient 10 millions de mots (Coleman *et al.*, 2011). Il me semble néanmoins délicat de fonder des analyses entières sur des corpus annotés automatiquement et dont les annotations sont bien souvent erronées. Prenons l'exemple du corpus Aix-Marsec. L'alignement en phonèmes a été réalisé de façon automatique. Nous avons travaillé sur les voyelles hautes (Herment & Turcsan à paraître) et lors de nos mesures, nous sommes rendu compte que de nombreuses erreurs étaient présentes. Dans le TextGrid correspondant au fichier A703G (fs 9 et TextGrid disponibles sur le CD-ROM), pris au hasard, l'étiquette du /iː/ de *Atlit* englobe les sons /l/ et /t/. De même avec le mot *these* juste après dont le /iː/ englobe /ð/ et /z/. Le /ɪ/ final de *policy* est aussi trop long. Il prend un bout du /s/ qui précède et un bout du /ə/ qui suit. Ce ne sont que trois exemples dans un fichier pris au hasard, et je n'ai regardé que les /iː/ et /ɪ/. Lors de notre étude, nous avions donc dû être vigilants et procéder à des corrections manuelles sous PRAAT avant de faire les mesures.

1.4.2. Réglages des seuils de F0 et variations micro-mélodiques

Un autre danger que représente l'utilisation non avisée des outils d'analyse est la mauvaise interprétation qui peut en découler. Prenons l'exemple d'une courbe de fréquence fondamentale dans PRAAT. Bien souvent, les chercheurs ou même les enseignants non spécialistes d'intonation vont utiliser ce logiciel s'ils ont un doute sur la mélodie d'un segment. Ils ne savent pas si ça monte ou si ça descend par exemple. La solution de facilité consiste alors à ouvrir le son dans PRAAT et à regarder la courbe de fréquence fondamentale. Or si l'on ne sait pas bien utiliser le logiciel, voici le genre d'erreurs qui peuvent se produire : les figures 5 et 6 ci-dessous représentent exactement la même portion de son, mais avec des réglages différents dans PRAAT :

Now it's slow progress but they are making progress fs 10

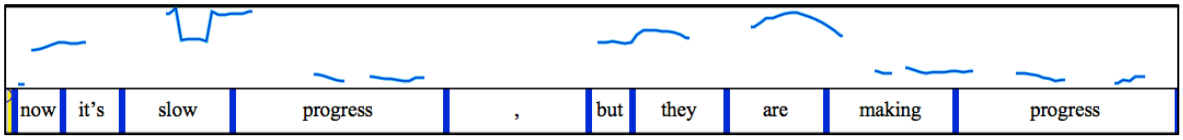


Figure 5 : fs 10 avec plancher à 50 Hz et plafond à 240 Hz

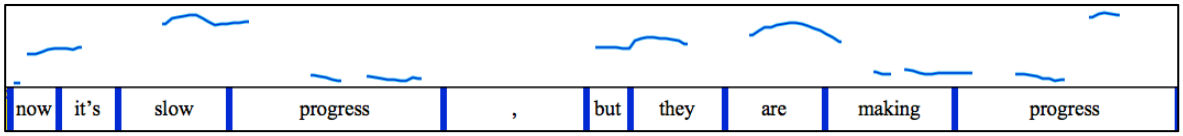


Figure 6 : fs 10 avec plancher à 50 Hz et plafond à 275 Hz

Si l'on fait le premier réglage des *pitch settings* de PRAAT avec un plafond à 250 Hz (le locuteur a une voix assez grave), on croit que *progress* remonte à la fin et on se demande ce qu'il se passe sur *slow* (figure 5). La courbe en forme de dent sur *slow* indique que le réglage n'est pas bon. Il faut donc l'affiner : en réglant le plafond à 275 Hz comme à la figure 6, la courbe de F0 sur *slow* est correcte, par contre on a un ton très haut sur la deuxième syllabe de *progress.* Ceci ne peut être qu'une erreur sinon on entendrait très bien une montée aussi abrupte et le /e/ de *progress* serait très aigu pour ce locuteur. Un troisième réglage avec un plafond à 249 Hz et un plancher à 40 Hz est montré figure 7 :

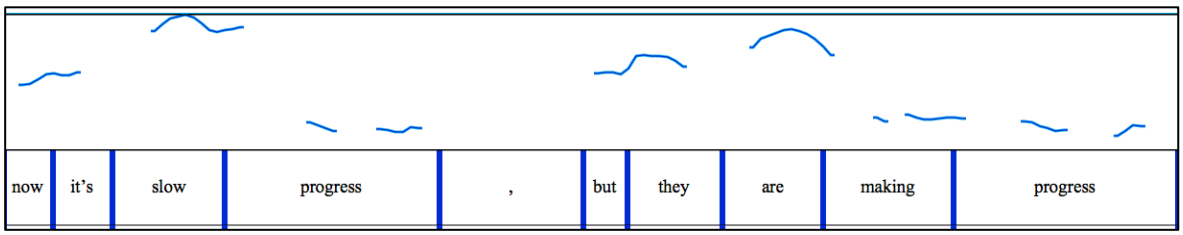


Figure 7 : fs 10 avec plancher à 40 Hz et plafond à 249 Hz

L'erreur micro-mélodique sur *progress* est gommée. Il s'agissait en fait du /gr/. Sur cette fenêtre, on retrouve le morceau de courbe de fréquence fondamentale sur la deuxième syllabe du mot. Est-ce que *progress* monte, comme un coup d'œil rapide pourrait le faire croire ? On entend clairement que non à l'oreille, et si on zoome et que l'on sélectionne uniquement *progress*, on voit que la petite remontée a lieu sur le /s/ et non sur la voyelle :

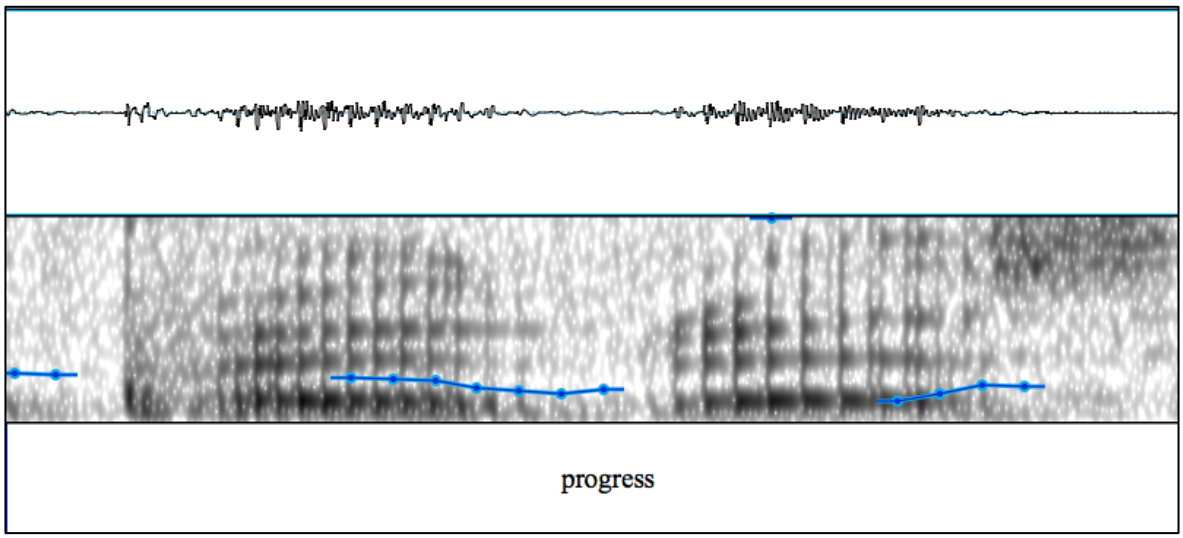


Figure 8 : spectrogramme de *progress*

Progress est donc bien prononcé sur un ton statique grave sans aucune remontée, car il est en queue d'énoncé (*tail*), en position désaccentuée à droite du noyau *are.*

L'algorithme Momel (Hirst *et al.* 2000) permet de lisser la courbe de F0 et d'en gommer les variations micro-mélodiques pour ne garder que la composante macro-mélodique :

> [T]he irregularities introduced by the voiced portions of the microprosodic component (*i.e.* during the voiced consonants) are not perceived as contributing *per se* to the overall pitch pattern of the utterance. (Hirst 2005 : 176)

La figure 9 ci-dessous reproduit l'exemple de Hirst (2005 : 177) d'une courbe de F0 brute lissée grâce à Momel :

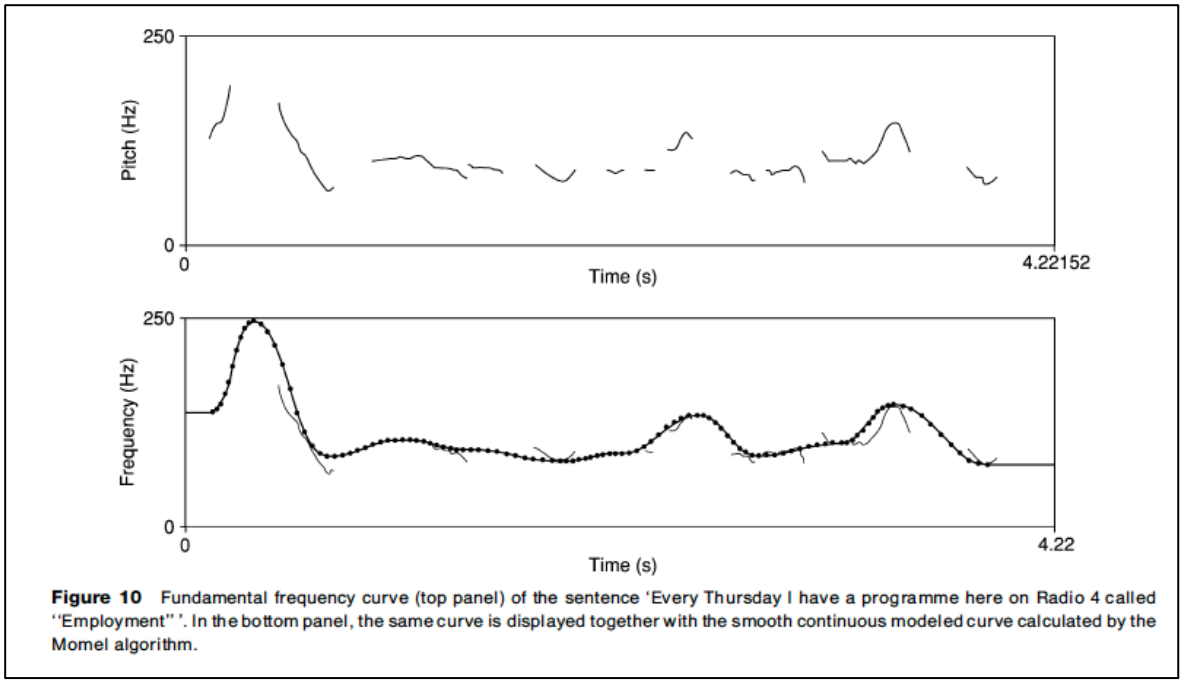


Figure 9 : courbe brute de F0 en haut et courbe lissée en bas pour la phrase *Every Thursday I have a programme here on radio 4 called 'Employment'*

La courbe du bas reproduit ce que nous percevons. On observe le même phénomène que pour *progress* de petite remontée à la fin de cet énoncé, visible sur la courbe brute. On voit que la courbe lissée l'a complètement gommée et qu'il s'agit bien d'un ton descendant.

Un autre exemple est celui que je commenterai en détails en 2.2.4 :

The red ones are my guess, and the blue pins are yours. fs 11

La différence est frappante pour le mot *ones* entre deux réglages différents, montrés aux figures 10 et 11 :

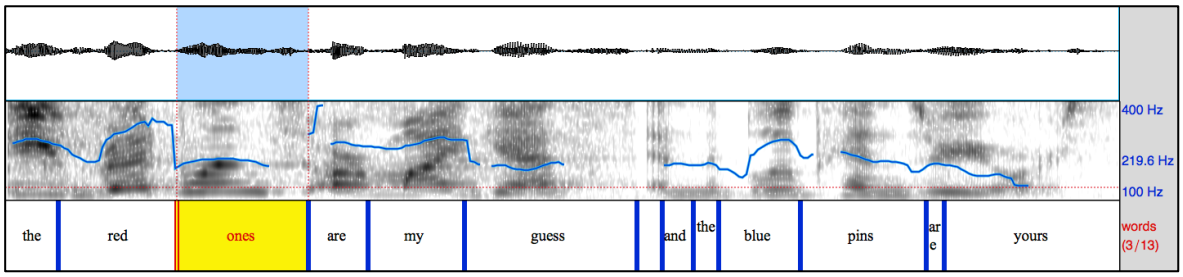


Figure 10 : fs 11 plancher 100Hz, plafond 400 Hz

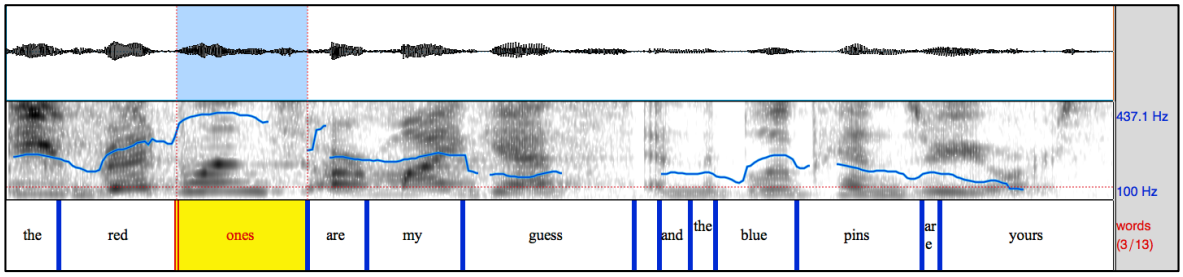


Figure 11 : fs 11 plancher 100 Hz, plafond 500 Hz

Encore une fois le décalage entre la courbe affichée et ce que l'oreille perçoit permet de se rendre compte que le réglage est mauvais. On entend très bien que *ones* est prononcé sur un ton très haut, le premier réglage ne peut donc être le bon.

Les deux exemples donnés dans cette sous-section illustrent deux erreurs très fréquentes lorsque l'on utilise PRAAT :

- celles dues à mauvais réglage du seuil de F0. J'ai montré qu'en affinant les réglages des seuils de F0 dans PRAAT, on pouvait tenter de gommer ces micro-variations. Il n'est cependant pas absolument nécessaire de le faire à chaque fois et de manière aussi précise. Il faut simplement être averti et savoir par exemple qu'un morceau de courbe de F0 qui apparaît dans l'aigu (ou dans le grave) de manière complètement décalée par rapport au reste de la courbe n'est pas à prendre en compte dans l'interprétation de cette courbe. Cela est d'autant plus important que l'on ne peut parfois pas faire disparaître ces variations, dans le cas où l'on a deux locuteurs (ou davantage) sur la même piste dont la tessiture est très différente (adulte/enfant, homme /femme ou deux personnes du même sexe ayant des timbres de voix très éloignés).
- Celles dues à une mauvaise interprétation des variations micro-mélodiques, comme la légère remontée que l'on voit sur le /s/ de *progress* qui n'est pas à interpréter comme une montée.

1.4.3. Erreurs d'interprétation de la courbe

Les problèmes d'interprétation de la courbe m'amènent à définir un deuxième type d'erreurs d'ordre phonologique celles-ci, à savoir les erreurs d'interprétation de la courbe en termes de catégorisations préétablies en contours.

Prenons la courbe de la 12 ci-dessous qui montre une portion de l'énoncé suivant :

Protestants however are a tiny minority in Argentina and the delegation won't be including the Roman Catholic fs 12

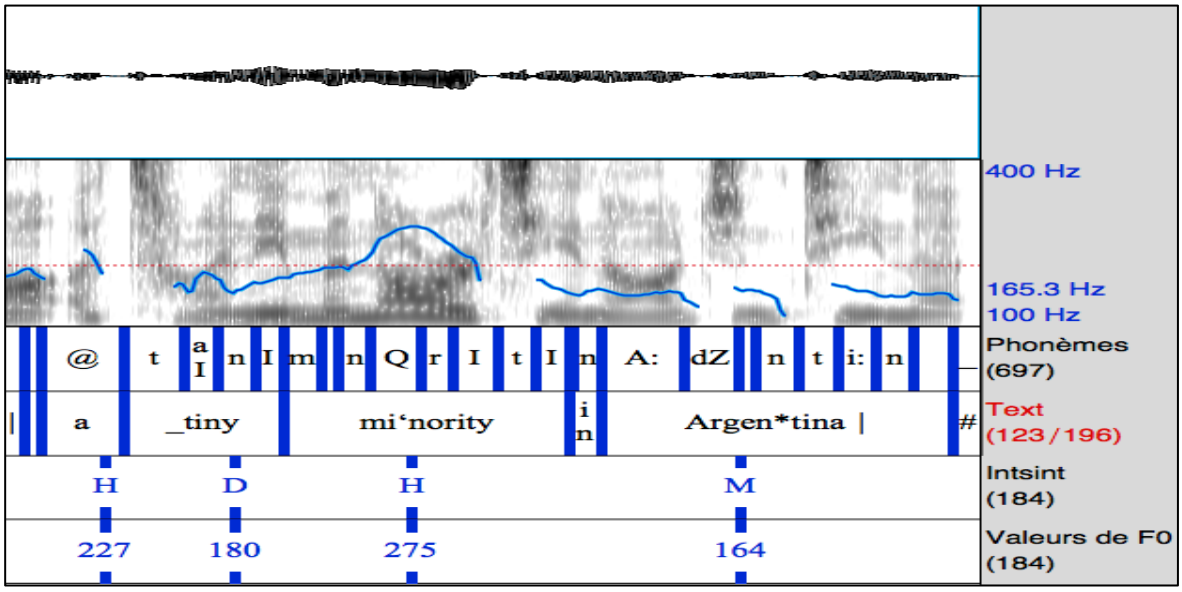


Figure 12 : oscillogramme, spectrogramme, courbe de F0 et quatre tires dont les tons intsint pour un extrait de fs 12.

D'aucuns pourraient penser que l'on a un *rise-fall* sur *minority* s'ils ne regardent que la courbe. C'est l'erreur classique du néophyte qui croit ce qu'il voit mais n'écoute pas son oreille. On a bien ici une remontée avant la chute, mais si le locuteur veut éviter de faire une rupture, un hiatus, il est bien obligé de remonter pour pouvoir redescendre. Ce que l'auditeur perçoit est bien une chute et non un *rise-fall*, comme la TSM[32] sur la tire nommée *text* le montre. Le symbole ' veut dire chute (*low fall*). C'est l'alignement des tons qui est important : la montée

32. *Cf.* l'introduction.

commence à partir de la deuxième syllabe de *tiny*, continue sur la première syllabe *minority* et la chute commence sur la syllabe accentuée de *minority* (l'antépénultième *-nor-)* pour continuer sur tout le mot. Le ton H sur la tire Intsint est d'ailleurs aligné sur le /ɒ/ accentué de *minority*. Si l'on avait eu un *rise-fall*, la montée aurait commencé sur la syllabe nucléaire : on sait que le mouvement mélodique nucléaire a lieu sur la syllabe nucléaire ou à partir de la syllabe nucléaire (Cruttenden 1998, Wells 2006 entre autres). Les syllabes qui précèdent le noyau (qu'elles soient accentuées ou pas) et éventuellement les consonnes qui précèdent la voyelle de la syllabe nucléaire sont souvent les supports des montées ou des descentes prénucléaires. Il faut donc bien regarder sur quelle syllabe a lieu le mouvement que l'on observe sur la courbe.

Il est impossible de comparer avec un *rise-fall* produit par la même locutrice. Ce ton est en général employé lorsque le locuteur est surpris ou impressionné (Roach 2009), ou bien lorsqu'il veut se montrer un peu provocateur (Wells 2006). L'ironie peut aussi être une des attitudes véhiculées par ce ton (Cruttenden 1998). Notre locutrice étant journaliste, elle n'a pas l'occasion d'employer ce ton dans le corpus étudié. Je propose donc de le comparer avec un *rise-fall* comme l'on en trouve en anglais d'Irlande du Nord (*cf.* Turcsan & Herment à paraître). L'alignement est différent, comme le montre la figure 13 ci-après :

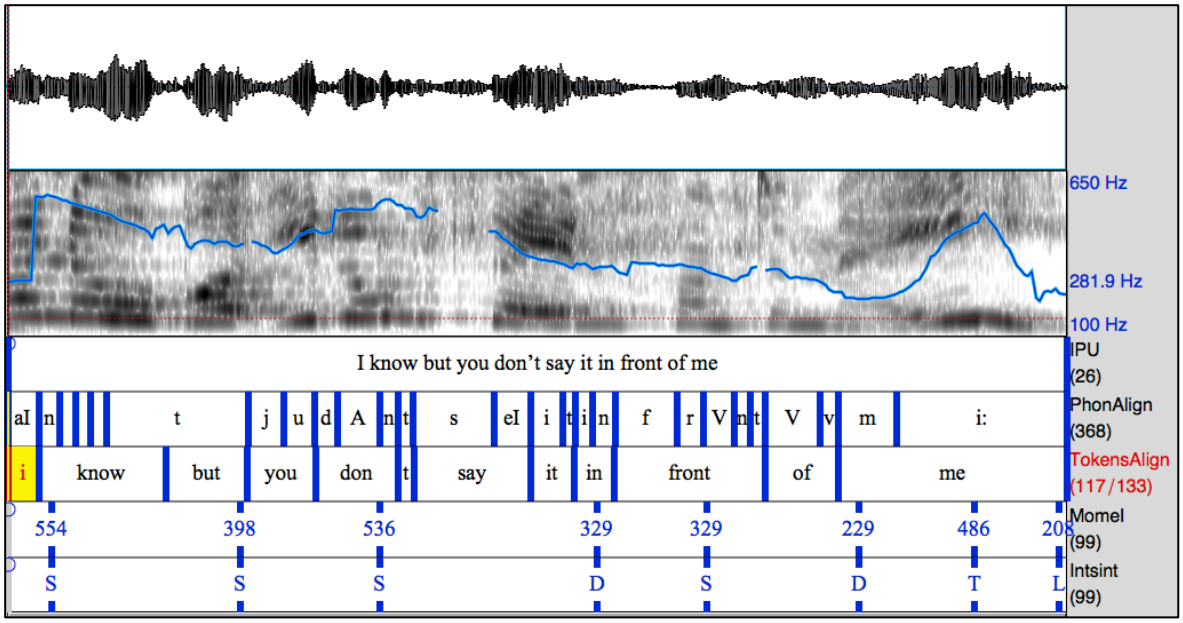


Figure 13 : extrait de la conversation PAC Donegal, locutrice Helen. fs 13

La locutrice prononce un *fall-rise* sur *me* qui ressemble à première vue et pour un œil non expert à celui de la figure 12 plus haut. Or si l'on y regarde de plus près on voit que l'on a un *downstep* avant la montée, marqué par le ton Intsint D au début du mot. Ce *downnstep* permet d'amorcer la montée : le locuteur part de plus bas. On a ensuite la montée, puis la redescente, marquées par les tons Intsint H et L au milieu de /iː/ puis à la fin du /iː/ respectivement. Le *rise-fall* entier a donc lieu sur la syllabe nucléaire. Il se trouve qu'il s'agit de la dernière syllabe de l'unité intonative. S'il y avait eu plusieurs syllabes après la syllabe nucléaire, la montée aurait eu lieu sur la syllabe nucléaire et la redescente sur les syllabes suivantes. En anglais du Nord de l'Irlande, on trouve un *rise-fall* avec la syllabe nucléaire en ton statique bas (ton B pour *Bottom* sur la tire Intsint), puis montée et redescente sur les syllabes de la queue de l'énoncé (*tail*) (*cf.* Rahilly 1997), comme on le voit sur la figure 14 ci-dessous :

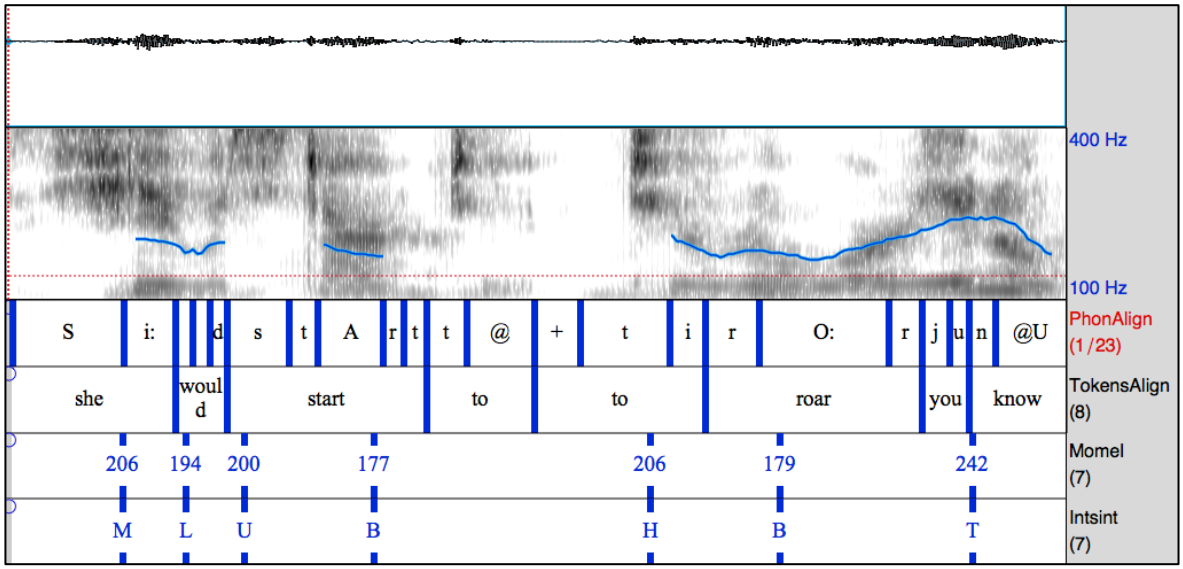


Figure 14 : extrait de conversation PAC DONEGAL, locutrice Majella fs 14

Ce contour est typique de la région du Donegal (Dalton & Ní Chasaide 1997). Il s'agit d'un prosodème dialectal, et même trans-variétal qui n'est pas utilisé en anglais britannique standard.

Il est à noter qu'il faut aussi se méfier de la tire Intsint : on n'a pas de ton L ou B pour marquer la redescente. On la voit sur la courbe de F0, mais les tons ne la donnent pas, contrairement à ce que l'on voit sur la figure 13. Cela ne veut pas dire pour autant que l'on n'a pas une redescente. Il manque un ton D, qui apparaît en fait juste après dans la séquence sonore, sur la première syllabe de *cock-a-doodle-doo*, qui est la queue de l'unité intonative, en statique grave. Ainsi si on coupe trop tôt et que l'on ne regarde pas l'unité intonative entière, on peut être amené à se tromper. Cela confirme une fois encore que sans l'aide de l'oreille, et sans une bonne connaissance préalable de l'outil utilisé, on peut interpréter les courbes et les tons de façon complètement erronée.

1.4.4. Conclusion : progression pédagogique pour la perception et la catégorisation des contours

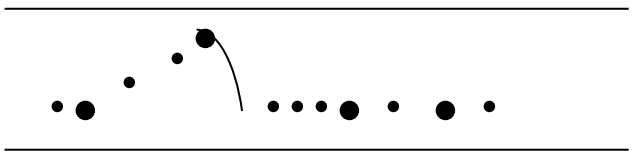
Ballier & Martin (à paraître a) mettent en garde contre l'annotation automatique, et préconisent une vérification visuelle sur le logiciel :

> The advantages of automatic transcription tools for prosody to annotate bigger corpora are obvious, but the difficult nature of acoustical fundamental frequency analysis calls for a careful visual checking of results. (Ballier & Martin à paraître a)

Je vais même plus loin en ajoutant qu'il faut une vérification auditive. Les exemples détaillés ci-dessus montrent bien qu'il est crucial de se fier à son oreille lorsque l'on utilise un logiciel. Beaucoup croient que grâce à PRAAT ou WinPitch (ou autre), ils vont enfin savoir quel est le mouvement mélodique d'un énoncé. Or si l'on n'a pas un minimum de connaissances théoriques et d'expérience de l'intonation, les logiciels de traitement de la parole peuvent mener à des erreurs gravissimes. Il n'est pas rare d'entendre des chercheurs dire des choses fausses sous prétexte que la courbe de PRAAT le montre. J'insiste donc à nouveau sur les dangers que peuvent représenter les logiciels pour un utilisateur non averti.

C'est la raison pour laquelle, dans la lignée de l'école britannique à laquelle j'appartiens, il me semble crucial de commencer l'enseignement de l'intonation à l'aide des transcriptions tonétiques interlinéaires : les unités intonatives sont délimitées, les contours globaux pour chacune d'elle sont notés grâce à la position du noyau, et les syllabes accentuées et inaccentuées sont marquées.

Si je reprends l'exemple de la figure 12, voici la transcription que l'on obtiendra :



a tiny minority in Argentina

La catégorisation des contours est donc implicite dans ce mode de transcription. On a ci-dessus par exemple un *high fall*, précédé de sa montée préparatoire typique (Wells 2006).

On pourra ensuite passer à une analyse visuelle des contours grâce au logiciel de traitement de la parole choisi, en confrontant une transcription

interlinéaire préalable à ce que l'on voit sur le logiciel, afin de comprendre les erreurs potentielles d'interprétation de la courbe.

La dernière étape consistera à appliquer des transcriptions automatiques de l'intonation comme Intsint, mais elle ne se dispense toujours pas de l'aide de l'oreille puisqu'il est recommandé pour une utilisation optimale d'Intsint de procéder à une vérification par resynthèse afin de s'assurer que l'énoncé obtenu de la sorte est perceptivement équivalent à l'original.

Le logiciel aide l'expert de façon incontestable et je suis la première à l'utiliser. Mais le logiciel ne remplacera jamais l'expérience de l'écoute.

CHAPITRE 2

La proéminence prosodique, du suprasegmental au segmental

Ce chapitre correspond à la première partie du volume des travaux, mais je ferai également référence à des travaux classés dans le deuxième axe. L'ordre chronologique est respecté ici (ce ne sera pas toujours le cas) car cette partie inclut les premiers articles que j'ai écrits après ma thèse. Celle-ci portait sur les corrélats acoustiques et prosodiques de l'emphase. Mes premiers travaux sont donc logiquement ancrés dans une perspective acoustique et traitent de l'emphase au sens large. En effet, j'avais choisi lors de ma thèse de travailler sur un corpus d'oral spontané et afin d'effectuer des analyses acoustiques sur des passages emphatiques et non emphatiques, j'avais demandé à des auditeurs anglophones de pointer les passages qui leur semblaient emphatiques dans des extraits de corpus que je leur faisais écouter. Une définition volontairement vague de l'emphase leur était donnée (voir Herment 2006) : « being made prominent in some way », ou encore « not neutral », « with a special involvement on the part of the speaker », ceci afin d'éviter une démarche circulaire. Les passages emphatiques pouvaient ainsi être des syllabes, des mots ou des groupes de mots. Mes travaux sur l'emphase s'intéressent donc à des segments de discours dont la taille varie. J'ai

étudié différentes unités rythmiques : l'unité intonative, l'unité tonale[33] (voir Herment & Hirst 2003) et le mot. On remarque que je pars de l'unité la plus grande pour descendre vers la plus petite. La question de la direction de l'approche, du suprasegmental vers le segmental ou bien inversement, se pose déjà ici, je vais y revenir. Les autres articles que j'ai inclus dans la première partie du volume des travaux portent sur la production de l'accent de mot dans des logatomes dissyllabiques (Turcsan & Herment 2013) ainsi que sur les voyelles hautes /iː/, /ɪ/, [i] et /uː/, /ʊ/, [u] (Herment 2010 et Herment & Turcsan, à paraître). Tous les niveaux prosodiques sont donc étudiés, depuis l'unité intonative jusqu'au phonème. C'est la raison pour laquelle j'ai choisi d'abandonner le terme d'emphase pour adopter celui de proéminence. Il paraît inopportun en effet d'employer emphase pour la proéminence accentuelle, ou pour la différence entre un /uː/ en position inaccentuée et un [u]. Le terme d'emphase, même si la définition en est vague et varie d'un auteur à l'autre, tout comme celle de proéminence du reste, semble plus approprié au niveau suprasegmental uniquement, alors que proéminence peut s'appliquer à tous les niveaux prosodiques, y compris celui du phonème. De surcroît, l'idée d'importance particulière est contenue dans le terme emphase : les dictionnaires en donnent comme définition en anglais « stress given to a word or words when speaking to indicate particular importance » (Oxford), « special importance or significance » (Collins), « special attention or importance » (Longman), « special forcefulness of expression that gives importance to something singled out ; prominence given to a syllable, word, or words, as by raising the voice or printing in italic type » (The American Heritage Dictionary). Les mots *special* ou *particular* reviennent dans chaque définition. C'est aussi ce que nous disons dans Herment & Hirst (2003) : « l'emphase est une mise en relief inattendue qui marque une implication de la part du locuteur plus importante qu'à la normale ». C'est ce qui

33. L'unité tonale est constituée d'un mot accentué (et en inclut donc le cas échéant l'anacruse) suivi de tous les mots inaccentués jusqu'au mot accentué suivant (voir Hirst 1999 et Hirst *et al.* 2000).

distingue selon moi emphase et proéminence : l'emphase est en fait une proéminence particulière. Je retiens donc le terme de proéminence.

Je propose dans ce chapitre de récapituler les propriétés acoustiques et linguistiques des différentes entités prosodiques que mes analyses sur l'emphase m'ont conduite à étudier de manière plus particulièrement fine. Je souhaite revenir sur les propriétés d'un certain nombre d'entités telles que le noyau, la tonique de tête et l'accent de mot, ce qui me permettra également de nuancer l'opposition catégorielle entre mots à contenu sémantique et mots grammaticaux. Je descends ensuite jusqu'à la syllabe pour en distinguer les différents degrés de proéminence. Chemin faisant, j'illustre ma méthode d'analyse prosodique des énoncés par le décorticage de certains cas ambigus ou plus délicats à analyser. La question de la détection des proéminences conduit à en interroger la hiérarchisation au sein des énoncés : je propose donc un modèle de hiérarchie des proéminences sur deux niveaux distincts : le niveau du mot à l'échelle de l'UI et le niveau de la syllabe à l'échelle du mot. Les corrélats acoustiques et prosodiques des différents degrés de proéminence sont discutés tout au long du chapitre et j'en résume les tendances à la fin, en nuançant avec la notion d'emphase.

2.1. Du segmental au suprasegmental : la force de l'inertie de la tradition

J'ai commencé par travailler sur le niveau suprasegmental pour ma thèse. Je me suis plus récemment intéressée au niveau des phonèmes et à l'accent de mot. La direction du suprasegmental vers le segmental semble ainsi logiquement privilégiée. Pourtant, c'est dans la direction inverse que nous enseignons traditionnellement la phonétique à Aix-en-Provence. On retrouve encore une fois le terme tradition dans le titre de cette sous-section. Il s'agit cette fois de la tradition de l'enseignement.

Gimson (1962) est représentatif de cette tradition de livres et de manuels sur la prononciation de l'anglais dans lesquels les sons sont traditionnellement traités en premier (en général ce sont d'abord les voyelles puis les consonnes, mais pas

toujours, voir Carr 2013 par exemple), pour laisser la place à l'accent de mot, au rythme, aux phénomènes de discours suivi et seulement en dernier lieu (ou parfois pas du tout) à l'intonation. Il s'agit bien entendu d'une vision caricaturale de l'ouvrage de [Gimson 1962] Cruttenden (2008), qui est bien plus complet qu'il n'y paraît ici. Cette tradition est maintenue par bien des auteurs, avec de légères variantes, mais suivant la même approche globale des phonèmes vers l'intonation, c'est-à-dire du segmental au suprasegmental. L'on peut citer de façon non exhaustive Lilly & Viel ([1977]1998b), Roach ([1983]2009), Catford ([1988]2001), Ginésy (1995), Watbled (1996), Carr ([1999]2013), Viel (2003), Collins & Mees ([2003]2013), Deschamps *et al.* (2004), Herry-Bénit (2011).

Il est frappant de constater que l'autre approche, du suprasegmental au segmental, n'est que très peu présente dans les manuels de phonétique ou de prononciation anglaises classiques. Duchet (1994) et dans son sillage Huart (2002, 2010) et Fournier (2010) considèrent que l'accent de mot est le « pivot du système » (Huart 2010 : 1). Fournier (2010 : 6) précise que « les syllabes accentuées sont ainsi la signature des mots : c'est essentiellement sur elles que repose l'identification des mots en anglais ». C'est par là qu'il faut commencer : « apprendre et enseigner l'anglais exigent le respect de son système accentuel qui est à l'origine de pratiquement toutes les difficultés de production et de reconnaissance des francophones » (Huart 2010 : 9). C'est aussi l'idée de Fraser (2001) :

> In the communicative approach we concentrate heavily in the early stages of pronunciation teaching on helping learners understand how syllables function in English, and especially on how stress is used. This forms the basis of further work on phonemes and prosody. (Fraser 2001 : 30)

Duchet (1994) traite d'abord de l'accentuation, puis des formes réduites et pleines et pour finir de la valeur phonétique des lettres. Il ne traite pas du tout de l'intonation, ni Fournier (2010), qui traite de l'accentuation aussi en tout premier lieu. Enfin, Huart consacre la première partie de son ouvrage de 2002 à la prosodie pour ne s'intéresser aux sons qu'ensuite. Dans la version révisée de 2010, c'est l'accent et le rythme qui font l'objet du premier chapitre mais l'intonation est

reléguée au troisième chapitre, après l'étude des phonèmes. Schottman (2012) part du rythme et des réductions de syllabes inaccentuées pour aller vers l'accent de mot, les règles de correspondance graphie-phonie, et donc les sons, puis l'intonation. Etrangement, c'est plutôt au niveau de l'anglais pour non-spécialistes que cette pratique est plus répandue. Diana (2010) propose par exemple l'ébauche d'un programme de phonétique pour les spécialistes d'autres disciplines qui commence par la prosodie (l'accent de mot et l'intonation) pour s'intéresser ensuite aux phonèmes puis aux phénomènes de discours suivi. Tous ces auteurs ne partent pas du niveau suprasegmental le plus haut, celui des unités intonatives, puisqu'ils partent d'abord du rythme (de l'accent de mot), mais il n'en reste pas moins que c'est à partir de phénomènes suprasegmentaux que tout s'articule.

Cette dernière approche, pourtant la moins populaire, me semble tout à fait justifiée au niveau de l'enseignement, comme la recherche tend à nous le démontrer.

2.2. Du suprasegmental au segmental

2.2.1. Acquisition et logique de la perception

Si l'on part de la logique tout à fait implacable de l'acquisition, il semble évident que c'est du suprasegmental qu'il faut partir. Comme je le dis déjà dans les tout premiers mots de ma thèse :

> L'enfant dans le ventre de sa mère entend ce qui se dit autour de lui mais l'enveloppe dans laquelle il se trouve joue le rôle d'un filtre faisant que l'enfant ne perçoit pas les mots, mais la prosodie. C'est grâce à la prosodie qu'il reconnaît sa langue maternelle et c'est pour cela que, lorsqu'il va apprendre une autre langue, c'est sa prosodie qu'il aura le plus de mal à modifier, celle-ci étant ancrée chez lui pratiquement depuis sa conception. […] Tout petit, lorsque l'enfant ne sait pas encore parler et commence à peine à comprendre les mots, il sait si on le gronde ou non, il est rassuré si on lui parle doucement. Il semble capable de percevoir du sens sans comprendre les mots. (Herment-Dujardin 2001 : 12)

Il en va de même pour un apprenant : même s'il ne comprend pas encore tous les mots et n'arrive pas encore à déchiffrer clairement tous les phonèmes, la prosodie l'aide à comprendre le sens global des énoncés. C'est ainsi qu'il me semble essentiel d'insister sur l'importance de la perception des unités prosodiques suprasegmentales, pas seulement en cours de phonétique, mais pour un exercice de compréhension par exemple, et au-delà de cela, lors de n'importe quelle situation d'écoute. Ce type d'arguments pédagogiques milite également pour un modèle configurationnel de l'étude de la prosodie, comme je vais tenter de le montrer ci-après.

2.2.2. Le repérage des unités intonatives

Lors d'une conférence, d'un cours, ou même d'une émission radiophonique, ce sont les mots portant les accents nucléaires qui vont être perçus saillants par les auditeurs. Ces mots-là sont porteurs d'information nouvelle, ou en tout cas d'information qui revêt une importance particulière pour le locuteur. C'est le concept d'informativité (*informativeness*) que nous avons développé dans Herment & Leonarduzzi (2012) et Leonarduzzi & Herment (2013a) :

> The notion of informativeness is to be distinguished from that of old/new information […]. An item is informative when it is relevant at this point of discourse. Conversely, it is uninformative when it is not relevant at this point of discourse. This notion sometimes matches that of old/given information but not always. An item can be given, but important for the speaker at a particular point of discourse: it is informative. (Leonarduzzi & Herment 2013a)

Nous montrons que dans certaines structures syntaxiques particulières comme les clivées, les dislocations à droite et les extrapositions, cette notion d'informativité est marquée par *tonality*, c'est-à-dire le découpage en unités intonatives (UI[34]).

34 . J'utiliserai désormais UI pour les unités intonatives. Comme expliqué dans l'introduction, j'ai décidé de suivre la tradition britannique, je parle donc d'unité intonative et non d'IP (intonation phrase). On aura cependant remarqué que dans certains de nos travaux en anglais, la terminologie IP est utilisée (Herment & Leonarduzzi 2012 et Leonarduzzi & Herment 2013a). Cela est en effet paradoxal, mais nous avons fait le choix d'employer le terme américain à dessein. Devant la popularisation de la théorie auto-segmentale et l'émergence du système de notation ToBI dans le monde de la

L'unité intonative est définie comme un segment de parole qui est prononcé avec un contour intonatif cohérent en terme de rythme et de mélodie (voir par exemple Crystal 1969 ou Cruttenden 1997). Dans le cas particulier de nos structures syntaxiques (*cf.* chapitre 3), le fait que le locuteur prononce la structure en la divisant en plusieurs UI montre que les différentes parties de la structure, et pas seulement celle qui est focalisée syntaxiquement comme dans les clivées en *it* par exemple, ont une importance particulière pour lui. Le lien avec l'enseignement coule de source : l'une des techniques permettant à un étudiant (natif ou apprenant d'ailleurs, l'enjeu est le même) de repérer les points importants dans le discours d'un locuteur (l'enseignant ou le conférencier par exemple) est d'identifier les mots sur lesquels un accent nucléaire, donc un accent mélodique (*pitch accent*) saillant, est prononcé.

L'expérience de l'enseignement montre que le découpage en unités intonatives ne pose pas de problème au niveau de la production. C'est au niveau de la perception que la chose se corse. Des indices phonétiques et phonologiques se révèlent alors précieux. Crystal (1969) donne quatre critères principaux :

- la présence de pauses (silencieuses ou remplies) ;
- une anacruse en début d'UI : l'anacruse est une série de syllabes proclitiques inaccentuées (Jassem 1952) ;
- l'allongement de la syllabe finale de l'UI ;
- un changement de direction de la mélodie ou un changement de hauteur mélodique (*step-up* ou *step-down*) non nucléaire, sur des syllabes inaccentuées souvent, qui signale en général la fin d'un groupe intonatif.

On peut ajouter deux autres critères :

- *pitch reset :* le réajustement mélodique qui consiste en un changement de hauteur mélodique en début d'UI ;
- et bien entendu la présence d'un syllabe nucléaire (porteuse d'un mouvement mélodique nucléaire).

---

recherche en prosodie, l'emploi du terme IP s'est révélé un choix stratégique face à certains évaluateurs de grandes conférences.

A l'exception de la pause, qui peut aussi être une marque d'hésitation ou de planification du discours, les indices donnés ci-dessus sont largement consensuels dans la littérature (voir entre autres Cruttenden 1997 : 36-7 ou Couper-Kuhlen 1986 : 75).

En général la division en UI correspond à la structure syntaxique, comme dans l'exemple ci-dessous, tiré du corpus Aix-Marsec (Auran *et al.* 2004, *cf.* chapitre 1). Une journaliste présente les nouvelles, il s'agit donc d'un monologue avec script. La division en UI suit strictement la syntaxe (les frontières d'UI sont marquées par le symbole | ; la ponctuation est délibérément omise) :

More news about the Reverend Sun Myung Moon | founder of the Unification Church | who's currently in jail | for tax evasion | he was awarded an honorary degree last week | by the Roman Catholic University of la Plata | in Buenos Aires | Argentina | fs 15

Il n'en va pas de même en discours spontané. Brown *et al.* (1980 : 46) déclarent qu'ils ont bien du mal à identifier les groupes intonatifs d'un corpus d'anglais oral enregistré à Edimbourg. Les hésitations, répétitions, faux départs et structures syntaxiques atypiques rendent la perception du découpage ardu en discours spontané, comme le précise Cruttenden :

> When we consider spontaneous speech (particularly conversation) any clear and obvious division into intonation-groups is not so apparent because of the broken nature of much spontaneous speech, including as it does hesitations, repetitions, false starts, incomplete sentences and sentences involving a grammatical caesura in the middle. (Cruttenden 1986 : 36)

L'exemple ci-dessous, extrait du COEC (Herment 2010, *cf.* chapitre 1), illustre parfaitement l'incohérence entre la division en UI et la structure syntaxique : on entend une frontière d'UI entre un verbe et son objet, ce qui n'arrive normalement pas en discours préparé.

It's very difficult to find somewhere | where you don't have to pay | agency's fees and everything else | fs 16

Dans ces cas très particuliers, le locuteur n'a souvent pas eu le temps de planifier son discours, et il rajoute quelque chose, comme c'est très probablement

le cas ici. Le fait d'entendre deux noyaux là où l'on pourrait s'attendre à n'en entendre qu'un ne gêne cependant pas la compréhension, bien au contraire. L'auditeur comprend que *pay* est aussi important que *agency's fees* (les noyaux des deux dernières UI) dans l'esprit de celui qui parle.

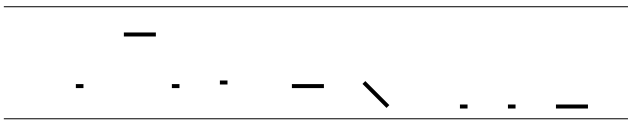
2.2.3. Détection des noyaux

Dans une unité intonative, plusieurs accents sont réalisés, grâce aux paramètres prosodiques que sont le mouvement mélodique, la longueur des phonèmes et l'intensité. Un mot de l'UI est considéré comme le plus important par le locuteur, qui place sur sa syllabe accentuée (pour l'anglais) la proéminence prosodique la plus forte, grâce à un mouvement mélodique : un changement brusque de hauteur mélodique ou bien le début d'un mouvement ascendant ou descendant. C'est ce qu'on appelle en anglais *pitch prominence*. Cette syllabe est la syllabe nucléaire, le noyau (*nucleus* ou *tonic* en anglais), celle qui porte la proéminence mélodique la plus forte (Cruttenden 1986 : 44), le ton mélodique distinctif (voir entre autres Tench 1996 : 12-13 ou Halliday & Greaves 2008 : 42-44), ce mouvement que l'on appelle *tone* en anglais si l'on s'en réfère aux 3T (voir l'introduction). Une UI ne peut contenir qu'une seule syllabe nucléaire.

La place du noyau est évidemment déterminante pour la compréhension : tous les mots qui suivent le noyau et qui sont donc désaccentués car déjà connus ne sont pas porteurs de mouvements mélodiques nucléaires. Cela est souvent très simple dans le cas d'un schéma intonatif globalement descendant : la chute va avoir lieu sur la syllabe nucléaire et les syllabes suivantes seront prononcées sur des tons statiques graves. Prenons la phrase

I've only been told part of the truth

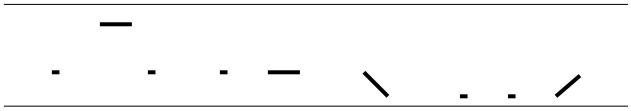
dont le noyau serait *part.* On peut la représenter schématiquement par une transcription scalaire comme ci-dessous, prononcée avec la chute sur le noyau *part* et les syllabes suivantes dans le grave :



I've only been told part of the truth

On perçoit en général aisément que le noyau n'est pas sur le dernier mot de l'UI, *truth,* mais sur un mot le précédant, cela marquant que *truth* est déjà connu et présent à l'esprit des co-énonciateurs. Ainsi c'est *part* et non *truth* qui représente l'information essentielle de cet énoncé.

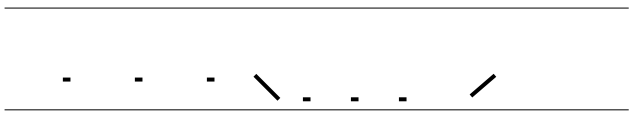
Ce n'est cependant pas toujours aussi aisé et d'autres schémas mélodiques, plus particulièrement le ton creusé (*fall-rise*), vont être bien plus problématiques : si la queue d'énoncé (*tail*) comporte plusieurs syllabes, alors la remontée n'a typiquement lieu que sur la dernière syllabe de la queue. Cette remontée pourrait être prise pour un mouvement mélodique nucléaire :



I've only been told part of the truth

Or la montée ne peut pas être sur un noyau car il n'y a clairement qu'une seule unité intonative (la syntaxe ne permet pas d'en faire deux ici, on ne peut pas couper entre le nom *part* et son complément), on ne peut donc avoir qu'un seul mouvement nucléaire. On ne pourrait pas avoir deux mouvements mélodiques nucléaires dans la même UI. Il s'agit donc bien d'un ton creusé (un *fall rise*) qui commence sur *part.*

On va cependant trouver des cas ambigus, là où par exemple une frontière d'unité intonative aurait été possible. Prenons l'exemple de Nicaise & Gray (1998 : 39) :



It would be awful if we failed

Lorsque cette phrase est prononcée comme ci-dessus, avec une chute du médium dans le grave (*low fall*) sur la première syllabe de *awful*, suivie de syllabes en statiques graves et une montée du grave dans le médium (*low rise*) sur *failed*, Nicaise & Gray estiment que l'on a deux groupes intonatifs, l'un en chute et le suivant en montée :

It would be awful F | if we failed R |

On pourrait pourtant considérer que l'on n'en a qu'un, avec le noyau sur *awful* et un ton creusé, qui commence donc sur le noyau (chute sur *aw-*) et dont la remontée a lieu uniquement sur la dernière syllabe (*failed*). Nicaise & Gray (1998 : 40) précisent d'ailleurs à propos de cette même phrase que « le choix de diviser en deux groupes intonatifs est un choix libre de la part du locuteur ». C'est certes un choix de la part du locuteur, et des indices sont clairement décelables à l'écoute. Je pense en effet que si le locuteur opte pour deux groupes intonatifs, une pause est probablement produite entre *awful* et *if*, et la proposition commençant par *if* est ce que l'on appelle parfois un *afterthought*, d'où la petite pause, qui indique que l'on rajoute quelque chose. Si l'on n'a pas cette mélodie typique de *l'afterthought*, qui débute en statique grave[35], et que l'on a un schéma montant classique, alors les syllabes prétoniques *if we* ne sont pas prononcées aussi bas et repartent d'un peu plus haut. C'est ce *reset*, le réajustement typique d'un début d'unité intonative, qui ferait percevoir une deuxième unité intonative. On comprend bien à la lecture de l'analyse de cet exemple, que la perception des frontières d'unités intonatives et donc des noyaux n'est pas simple.

La question du repérage des noyaux ne se pose pas que dans les cas de phrases complexes, mais interroge aussi l'opposition mots outils / mots à contenu

35. C'est le schéma mélodique que certains auteurs ont nommé *take off* (voir par exemple Ginésy 1995, à la suite de O'Connor & Arnold 1961).

sémantique. Les mots porteurs de l'accent nucléaire sont en général les mots à contenu sémantique (*content words*) par opposition aux mots grammaticaux ou mots outils (*tool words*). Les mots grammaticaux sont inaccentués en contexte neutre. En contexte non neutre, un mot grammatical peut tout à fait porter un accent de phrase, ou même un accent nucléaire. Il sera alors considéré emphatique. Nous avons en effet montré (Herment & Hirst 2002) que lorsque l'on délexicalise une phrase (en remplaçant les phonèmes de l'anglais par d'autres phonèmes et en resynthétisant tout en gardant les mêmes critères prosodiques), les mots grammaticaux qui étaient perçus emphatiques car non réduits par exemple, ne le sont plus autant dans la phrase délexicalisée car ils n'étaient pas porteurs de critères prosodiques fortement révélateurs de l'emphase. Dans l'exemple suivant

Now it's slow progress but they are making progress fs 10

le mot *are* est prononcé avec sa forme pleine /ɑː/. Il est donc accentué au niveau de la phrase et c'est même le noyau de l'UI. Il est perçu emphatique à 100% par les auditeurs, alors qu'il n'est plus perçu qu'à 30% dans le stimulus délexicalisé (*cf.* Herment 2006). Il est pourtant noyau et porteur d'un mouvement mélodique nucléaire. Ceci prouve que le mouvement mélodique n'a pas besoin d'être très ample pour faire ressortir ce mot : le fait qu'un mot grammatical (un auxiliaire ici) normalement inaccentué soit porteur d'un accent (marqué par sa prononciation non réduite) suffit à le faire ressortir, à le rendre proéminent (je reviendrai en 2.2.6 sur cette proéminence due à la qualité de la voyelle et en 2.3.4.1 sur cet exemple particulier).

Je montre aussi que le sémantisme du mot peut jouer un rôle non négligeable dans sa mise en relief (*cf.* 2.3.4.2), même s'il n'est pas le seul paramètre :

> Il nous semble nécessaire de distinguer 2 types de mots, que nous appelons les mots sémantiquement marqués, et les mots neutres. La 1ère catégorie de mots est emphatique par nature en quelque sorte, et ces mots n'auraient pas besoin d'un marqueur prosodique. On les trouve pourtant souvent associés à un

mouvement mélodique et à un allongement [...]. Ceci est l'indice à nouveau d'une combinaison de tous les paramètres. (Herment 2006)

Ainsi le noyau présente une proéminence mélodique (*pitch prominence*), mais celle-ci peut être moindre si d'autres critères tels que la catégorie grammaticale ou le sémantisme entrent en ligne de compte.

2.2.4. Saillance et tonique de tête

De la même façon que pour le noyau, les mots porteurs d'un accent de phrase, c'est-à-dire les accents qui se trouvent avant l'accent nucléaire, auront également une importance au niveau du message. Ce sont aussi les mots à contenu sémantique, mais tout comme pour le noyau, les mots grammaticaux peuvent porter un accent de phrase en contexte non neutre. Parmi ces accents, le premier de l'unité intonative, que l'on appelle la tonique de tête (*head* ou *onset* selon les auteurs) est plus saillant que les autres (Brazil *et al.* 1980, Couper-Kuhlen 1986).

Nous avons vu que la syllabe nucléaire était marquée par un accent mélodique (et éventuellement par d'autres marques). La tonique de tête est quant à elle traditionnellement représentée par un ton statique lorsque l'on utilise la notation scalaire comme en 2.2.3 ci-dessus, ou bien par des points et non des mouvements mélodiques avec la transcription tonétique interlinéaire comme dans Cruttenden (1997) ou Wells (2006) par exemple. Les tons statiques sont en effet des syllabes qui sont proéminentes car placées sur un ton haut ou bas par rapport aux autres, mais aussi parce qu'elles sont perçues proéminentes à cause d'une durée plus longue et d'une intensité plus forte que les syllabes adjacentes.

Un exemple de transcription tonétique interlinéaire est reproduit figure 15 ci-dessous :

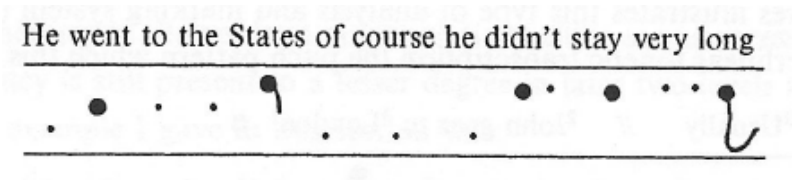


Figure 15 : exemple d'une transcription tonétique interlinéaire (Cruttenden 1997 : 37)

Dans cet exemple, une frontière d'unité intonative est repérable entre *States* et *course* (elle ne peut pas se situer après *course* car dans ce cas il y aurait eu un réajustement et le *he* serait reparti d'un peu plus haut, comme nous l'avons vu en 2.2.2 et 2.2.3). Ainsi si l'on prend la deuxième UI, on a deux accents de phrase autres que le noyau, qui sont représentés par les gros points et qui sont des tons statiques selon la transcription : *did-* et *stay*. Le premier accent de phrase de l'UI, est donc représenté de la même façon que le deuxième. Le premier accent est pourtant toujours plus saillant que les autres, se démarquant souvent en ce qu'il est généralement plus haut ou plus bas. La tonique de tête n'est traditionnellement jamais représentée par un mouvement mélodique, c'est sa hauteur qui la distingue des autres accents et non le mouvement.

Les toniques de tête ont souvent un rôle important dans le message. On peut tout à fait mettre un pronom en tonique de tête pour marquer un contraste par exemple, ou encore un auxiliaire comme un *do* qui n'est pas nécessaire au niveau de la structure syntaxique et qui a pour fonction de renforcer le lien prédicatif (voir 3.3.2.5) comme dans l'exemple suivant tiré de Herment (2010) :

They did have trains fs17

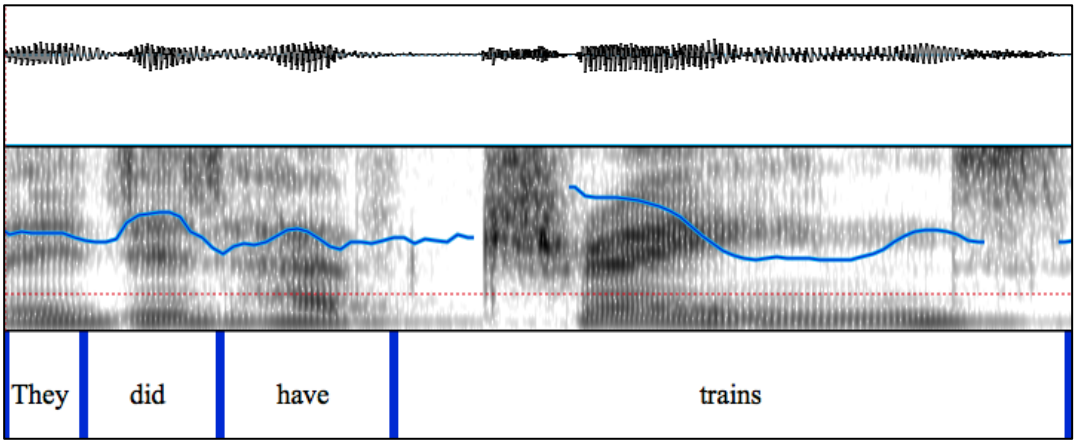
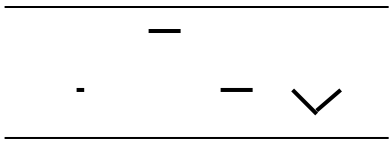
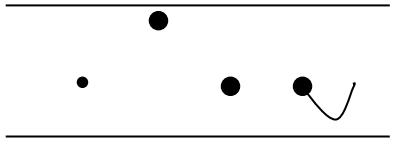


Figure 16 : exemple d'un do auxiliaire placé en tonique de tête (fs 17)

On voit nettement sur cette figure (qui est la reproduction d'une fenêtre du logiciel PRAAT, Boersma & Weenink 2001) que l'on a un ton plus haut sur *did* que sur *have*. Les représentations scalaire et tonétique interlinéaire pour cet énoncé seraient donc les suivantes :



They did have trains



They did have trains

Le fait que *do* soit placé en tonique de tête et *train* en noyau de l'UI est essentiel pour le sens de l'énoncé : dans ce cas précis, la relation prédicative est renforcée grâce à *do*. Si *do* avait été placé en noyau, comme cela est tout à fait possible, cela aurait signifié que le locuteur contredit ce qui a été énoncé avant (*cf.* Herment 2009, 2010, 2011). On trouve aussi des *do* auxiliaires en forme affirmative qui ne sont ni tonique de tête, ni noyau, mais de simples accents de phrase :

I really do think | it's better to work for a voluntary organisation like VSO fs 18

Dans ce cas, j'ai émis l'hypothèse que l'auxiliaire était présent pour renforcer le sémantisme de l'adverbe qui le précède (*cf.* Herment 2010).

Il est donc essentiel de distinguer la tonique de tête des autres accents de phrase. Elle a une valeur importante au niveau du message, à tel point que les auditeurs non expérimentés ont souvent tendance à la confondre avec le noyau, surtout en contexte plutôt neutre. Dans ces cas de figure, le noyau est porteur d'un mouvement mélodique qui n'est pas forcément très ample, alors que la tonique de tête va être perçue plus proéminente car elle est en général beaucoup plus haute. L'exemple ci-dessous, dont le fichier son est tiré du mémoire de M2 d'une de mes étudiantes (Féron 2013), est emblématique de cette confusion possible, et aussi de la difficulté à percevoir le découpage en UI pour un auditeur non chevronné :

The red ones are my guess, and the blue pins are yours. fs 11

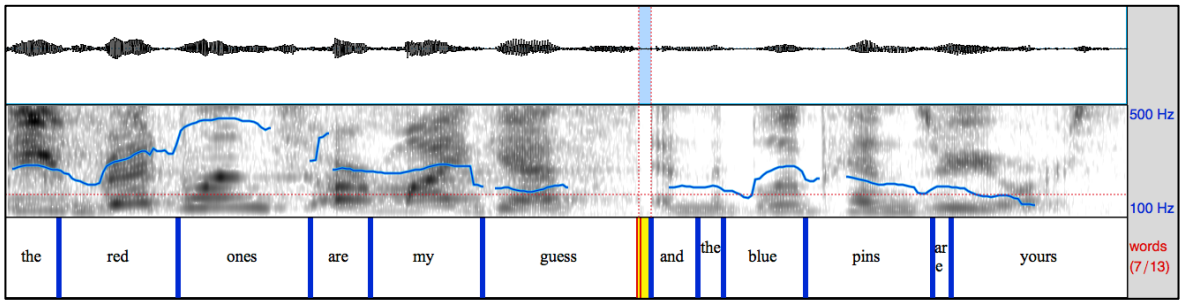
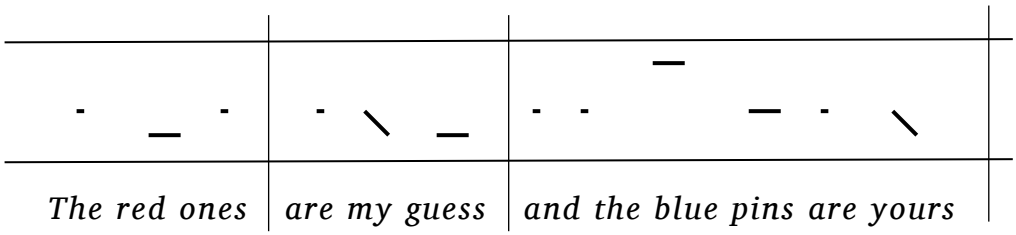


Figure 17 : oscillogramme, spectrogramme et courbe de F0 pour fs 11

Cet énoncé comprend en fait trois unités intonatives, avec une frontière après *ones* et une frontière après *guess*. On a l'impression d'entendre une pause entre *ones* et *are*. Or on voit bien sur le spectrogramme à la figure 17 ci-dessous qu'il ne s'agit pas d'une pause mais d'un allongement du /z/ (qui semble être un /s/ en fait) de *ones*. Ceci est un premier indice de frontière d'UI. Le deuxième indice pour cette frontière est le réajustement (*reset*) qui suit *ones* : *are* commence beaucoup plus bas que *ones*. Enfin, le dernier indice est la présence d'un contour global montant sur *the red ones*, avec un noyau sur *red*, qui est donc plutôt prononcé en statique, médium ici, la remontée sur la queue d'UI *ones* et un nouveau contour cohérent, descendant cette fois, sur *are my guess*. La frontière entre la deuxième et la troisième UI est nettement plus audible, avec la présence d'une pause (ou d'un coup de glotte plus vraisemblablement du fait des explosions visibles sur le spectrogramme), qui est sélectionnée sur la figure 17, une anacruse formée par les syllabes *and the* prononcées rapidement, et un contour cohérent sur la troisième UI. Quant aux accents, il n'est pas évident non plus de repérer les noyaux et les toniques de tête. Dans la deuxième UI, le noyau est sur *my* : il est porteur d'une chute et le mot suivant est désaccentué, en statique grave comme on le voit bien sur la figure 17. Dans la troisième UI, *blue* est tonique de tête car il est prononcé en statique plus haut que les autres mots de l'UI et parce que le reste du schéma présente un contour globalement descendant jusqu'au dernier mot *yours*, porteur de la chute terminale et donc noyau de l'UI. Je sais par expérience que dans de tels énoncés la tonique de tête (*blue* ici) est souvent perçue comme le noyau. Si le

noyau avait été sur *blue*, ce qui suit aurait été en statique grave. On voit ici encore que la tonique de tête *blue* est essentielle dans le message puisque *blue* est le pendant de *red.* La représentation scalaire pour cet énoncé est donc la suivante :



Ma représentation sur le texte serait celle-ci :

The red ones R | are my guess, F | and the blue pins are yours. F|

Les contours prosodiques sont tout à fait en adéquation avec le sémantisme et les contrastes induits par la locutrice : *red* et *blue* d'une part et *my* et *yours* d'autre part. Le premier élément de chaque contraste est donné en noyau (*red* et *my*) et en deux unités intonatives séparées afin de bien en montrer l'importance, mais liées par le *rise* de la première. Une fois le décor posé, marqué par la chute sur la fin de la deuxième UI, la locutrice prononce le reste en une seule UI, en donnant le deuxième élément du premier contraste en tonique de tête (*blue*) et le second en noyau car en fin d'UI (*yours*).

Pour conclure cette sous-section sur la tonique de tête, on peut dire que cette dernière est quand-même bien marquée prosodiquement, pour preuve la fréquente confusion entre tonique de tête et noyau. Si la tonique de tête est en général plutôt prononcée sur un ton statique, elle est aussi parfois porteuse d'un mouvement mélodique, mais moins ample que celui qui a lieu ou débute sur le noyau et qui peut même être parfois perçu comme un statique :

> We need to remember that physical events and perceived linguistic structures are not necessarily directly related. [...] A 'level' tone in the British system may be anything but level if we look at the F0 contour on a computer screen. (Wichmann 2000 : 13)

L'énoncé ci-dessous tiré de Wells (2006 : 8) en est un très bon exemple fs19 .

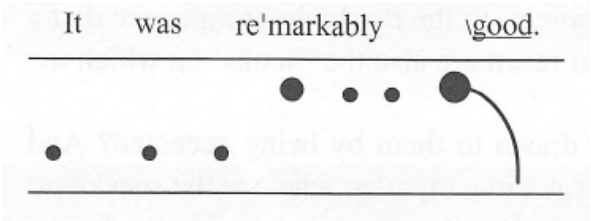


Figure 18 : exemple de transcription tonétique interlinéaire (Wells 2006 :8)

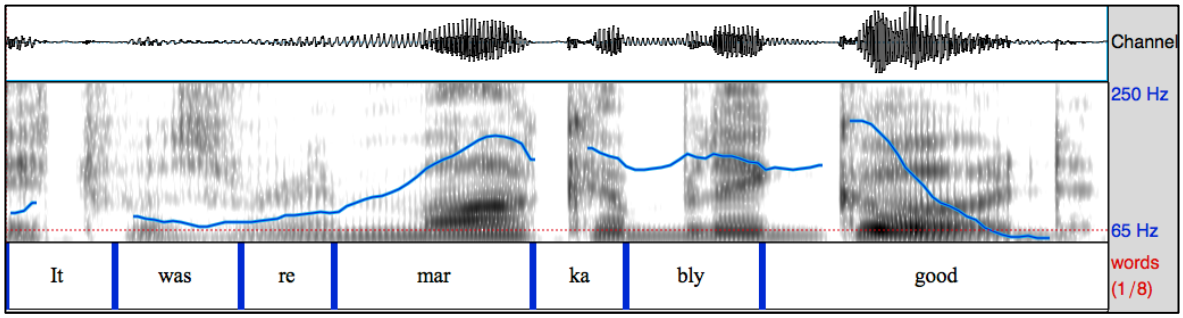


Figure 19 : visualisation de la mélodie dans PRAAT de l'énoncé de la figure 18 (fs 19)

On trouve la transcription de Wells reproduite en figure 18 et la figure 19 montre la courbe mélodique dans PRAAT. La courbe de fréquence fondamentale ne correspond pas exactement à la transcription tonétique interlinéaire que donne Wells : on voit une montée sur *remarkably* alors que Wells donne un ton statique. Cela est tout à fait normal, il faut bien monter quelque part pour aller placer la syllabe *-mar-* dans le statique aigu. L'oreille ne perçoit pas une montée mais un statique car le ton haut est placé sur la voyelle, la montée a lieu sur le /m/. Les voyelles étant les noyaux des syllabes, ce sont elles qui portent les tons si elles sont longues. Si elles sont courtes, on peut avoir un ton haut sur la voyelle suivi d'un ton bas sur la voyelle de la syllabe suivante et cela sera perçu comme une descente. De même pour une montée, on peut avoir un ton bas sur la voyelle de la syllabe accentuée puis une montée sur la suite, alors on perçoit une montée. Ici on ne peut pas percevoir un *rise* car la montée a lieu avant la voyelle.

Les transcriptions interlinéaires sont donc tout à fait précieuses et il faut se méfier de ce que l'on peut voir sur les logiciels de traitement de la parole lorsque l'on ne possède pas un certain bagage théorique et pratique (voir chapitre 1). Ces transcriptions sont en effet schématiques et représentent finalement plutôt ce que

l'on perçoit de l'intonation, c'est-à-dire une courbe lissée en quelque sorte (voir Momel, chapitre 1) :

> An utterance may be sprinkled with sudden high frequencies at high vowels (e.g. /i/) and silent spots at devoiced stop consonants (e.g. /p/) (Cruttenden 1997), but somehow the listener looks past these perturbations and perceives what seem to be reasonably smooth frequency contours. (Edwards 2001 : 332)

Pour ne pas confondre tonique de tête et noyau, il faut aussi bien garder à l'esprit que le noyau est la syllabe porteuse du dernier mouvement mélodique de l'UI. Les apprenants ont souvent tendance à l'oublier et à accorder davantage d'importance à la tonique de tête qu'au noyau, ceci étant peut-être dû à une surreprésentation des premières proéminences prosodiques chez l'apprenant.

2.2.5. Accent de phrase et accent de mot : quels niveaux de proéminence ?

Les accents de phrase sont les syllabes proéminentes qui se trouvent entre la tonique de tête et le noyau. Ils sont représentés par des tons statiques dans les transcriptions que nous avons vues plus haut (voir *stay* figure 15), mais ces représentations sont schématiques. Un petit mouvement mélodique peut également être produit sur les accents de phrase, et parfois perçu, comme on peut le voir sur *pins* dans l'exemple de la figure 17. Ils correspondent souvent aux mouvements globaux des patrons intonatifs et sont donc non distinctifs, mais importants au niveau de la hiérarchie prosodique des proéminences (*cf.* 2.3).

Ces accents sont perceptibles grâce à des corrélats prosodiques comme l'intensité et la durée, plutôt que grâce à la mélodie. De nombreux auteurs ont travaillé sur la proéminence syllabique et sur les notions de *stress*, *accent* et proéminence (voir Herment-Dujardin 2001 : 37-42 pour une synthèse) et le manque de consensus est criant. Si la durée et l'intensité ressortent quand-même comme les corrélats de l'accent de mot, Sluijter (1995) et Sluijter & van Heuven (1993, 1996), Campbell (1995) ont montré que la différence la plus marquante entre *stressed* et *unstressed syllables* est une augmentation de l'intensité dans les fréquences hautes des voyelles accentuées. Cet effet nommé « balance spectrale » ou « tilt spectral » est cependant très difficile à mesurer, en particulier en discours

spontané. Campbell & Beckman (1997) affirment quant à eux qu'en anglais l'effet de balance spectrale n'est discriminant que dans le cas des syllabes nucléaires, mais pas entre *stressed* et *unstressed* syllables. Une étude plus récente de Kochanski *et al.* (2005) sur les syllabes proéminentes (au sens de *stressed*) dans différentes variétés d'anglais a montré que « speakers primarily marked prominence by patterns of loudness and duration ». Ainsi la durée et l'intensité semblent être les paramètres les plus importants pour l'accent de mot. Les toniques de tête présentent également des phonèmes allongés et une intensité plus forte, mais c'est davantage la hauteur qui les distingue des autres accents de phrase.

Les accents de phrase ont toujours lieu sur des syllabes accentuées aussi (*stressed*), de la même façon que le noyau ou la tonique de tête. La syllabe d'un mot à contenu monosyllabique est considérée comme une syllabe accentuée (*stressed*) au même titre que la syllabe accentuée d'un mot polysyllabique. On en vient donc à l'accent de mot. En anglais les mots de plus d'une syllabe présentent tous une syllabe qui est plus proéminente que les autres, la syllabe accentuée. Si les accents de phrase ne peuvent avoir lieu que sur des accents de mot, l'inverse n'est pas vrai et toutes les syllabes accentuées au niveau lexical ne sont pas porteuses d'un accent de phrase. Dans une UI dont le noyau n'est pas le dernier mot à contenu sémantique par exemple, tous les mots qui se trouvent après le noyau sont désaccentués, comme dans l'exemple suivant (tiré de Leonarduzzi & Herment 2013b) où le noyau dans la première UI est sur *work* et où *winning* et *prize* ne portent pas d'accent de phrase :

> *But has it had an effect on your work winning the prize F | ii in the sense that you know you can write a best-selling prize-winning novel F |* fs 20

Dans l'exemple de Cruttenden à la figure 15, c'est le mot *course* qui ne comporte pas d'accent de phrase car il est placé avant la tonique de tête, en position de pré-tête (*pre-head*). Ces mots sont en général prononcés plus rapidement et le tempo entre en jeu ici.

On distingue donc deux niveaux différents, celui de la phrase et celui du mot. Le français ne possédant qu'un seul mot, accent, pour l'accent de phrase et

l'accent lexical, il est plus aisé d'employer les mots anglais, car l'on peut distinguer *accent*, que j'emploie pour faire référence à l'accent de phrase, et *stress*, qui désigne pour moi l'accent de mot. On peut donc différencier *stressed syllables* et *accented syllables*. Mais la chose n'est pas aussi simple qu'il y paraît. On distingue d'une part les syllabes accentuées au niveau du mot et porteuses d'un accent de phrase : on peut les appeler *accented stressed syllables.* Nous avons vu en 2.2.4 qu'elles pouvaient comporter un petit mouvement mélodique mais étaient caractérisées plutôt par un ton statique, et par un allongement de la durée des phonèmes et une intensité plus forte. On a ensuite les syllabes accentuées au niveau du mot mais pas au niveau de la phrase, *unaccented stressed syllables*, qui sont marquées quant à elles aussi par la durée et l'intensité mais dans une mesure moindre que les précédentes. Les transcriptions interlinéaires (*cf.* l'introduction) sont ambiguës à cet égard car elles représentent par des tirets ou des gros points les accents lexicaux (*stressed syllables*), même s'ils ne sont pas porteurs d'accent de phrase (*accent*).

Qu'en est-il des monosyllabes et des accents secondaires ? J'entends par accent secondaire un accent de mot qui se trouve à gauche de l'accent primaire dans un mot, selon les traditions des dictionnaires de prononciation (Wells 2008 : 783, Jones 2011 : xvi). En ce qui concerne les monosyllabes à contenu sémantique, on considère que leur syllabe unique est accentuée comme un accent de mot primaire d'un mot polysyllabique. Ainsi dans l'énoncé de la figure 17

The red ones are my guess, and the blue pins are yours fs 11

le mot *pins* est porteur d'un accent de phrase, tout comme le mot *effect* par exemple dans fs 20 ci-dessus.

Quant aux accents secondaires, ils sont à un niveau de proéminence inférieur à celui des accents primaires au niveau du mot. Qu'en est-il au niveau de la phrase ? Doit-on les considérer comme des accents de phrase ? A ma connaissance aucune étude ne traite de cela et la chose n'est pas simple. Wells (2008 : 783) semble considérer qu'un accent secondaire ne sera porteur d'un accent de phrase que s'il y a glissement accentuel (*stress-shift)*, c'est-à-dire que si l'accent secondaire devient accent primaire pour des raisons de rythme en

discours suivi. Cela voudrait dire que les accents secondaires ne sont pas porteurs d'accents de phrase. Mais si la syllabe porteuse de l'accent secondaire se trouve dans un mot accentué au niveau de la phrase, sera-t-elle plus saillante que dans un mot non porteur d'accent de phrase ? L'accent secondaire d'un mot accentué au niveau de la phrase sera-t-il plus saillant que l'accent primaire d'un mot non accentué au niveau de la phrase ? Mon sentiment est qu'il est très difficile de répondre à ces questions et qu'une étude mériterait d'être envisagée (j'y reviendrai au chapitre 4). La réflexion sur les accents secondaires me conforte dans l'idée que les unités rythmiques sont cruciales et qu'il faut distinguer deux niveaux dans les proéminences : les proéminences sur les mots à l'échelle de l'UI, et les proéminences sur les syllabes à l'échelle du mot. Car si la question se pose pour les accents secondaires, alors pourquoi ne se poserait-elle pas pour les syllabes inaccentuées des mots porteurs d'accent de phrase et des mots non porteurs d'accent de phrase ? Il se révèle donc très complexe de tenter d'imbriquer les deux niveaux. A défaut d'étude sur cette question épineuse, je ne sais donc pas comment placer les accents secondaires dans la hiérarchie des proéminences. Il est vrai que peu de mots en anglais possèdent finalement des accents secondaires, mais il n'en reste pas moins qu'un problème se pose. Je propose donc de placer l'accent secondaire dans un mot accentué au niveau de la phrase sur le même plan qu'un accent de mot primaire non accentué au niveau de la phrase. Quant à un accent secondaire dans un mot non accentué au niveau de la phrase, il sera à mon avis moins proéminent et se trouvera donc plutôt à un niveau inférieur à *unaccented stressed syllable.* Mais j'en reste à de simples hypothèses qui mériteraient d'être testées (voir chapitre 4).

2.2.6. Les syllabes inaccentuées et la qualité de la voyelle

Reste les syllabes inaccentuées, qui représenteraient alors le degré zéro de proéminence. Or ce n'est pas tout à fait le cas. Il est bien connu qu'une syllabe accentuée peut être réduite ou non. Ceci dépend de la qualité de la voyelle : une voyelle peut changer de timbre selon qu'elle est accentuée ou pas, dans les mots à contenu tout comme dans les mots outils. Les voyelles que l'on trouve en position

réduite sont /ə/, /ɪ/ et /ʊ/, la dernière étant beaucoup plus rare. Prenons l'exemple (emprunté à Watbled 1996 : 94) de *courage* /ˈkʌrɪdʒ/ *et courageous* /kəˈreɪdʒəs/. On note les alternances vocaliques dans les deux mots entre /ʌ/ et /ə/ et entre /ɪ/ et /eɪ/ selon qu'ils sont sous accent ou non. Ces alternances sont automatiques. Pour les mots outils, on peut réduire mais ce n'est pas obligatoire, même si c'est très courant : si l'on reprend l'exemple

The red ones are my guess, and the blue pins are yours fs 11

on entend la réduction de *and* /ən/ et de *are* /ə/ dans la dernière UI.

Dans la perspective perceptuelle dans laquelle je me place, la qualité de la voyelle s'est révélée cruciale dans la détermination des niveaux de proéminences. Nous en avons fait l'expérience pour l'étude sur la place de l'accent dans les dissyllabes (Turcsan & Herment 2013). Nous avons fait lire à des anglophones natifs des logatomes dissyllabiques insérés dans des phrases porteuses et avons ensuite déterminé à l'écoute où le locuteur plaçait l'accent lexical. La tâche a été aisée dans la plupart des cas, et les rares occurrences pour lesquelles nous avons hésité étaient celles dans lesquelles le locuteur n'avait fait aucune réduction, comme dans le mot *calbains* ci-dessous, prononcé /kælbeɪnz/[36].

My Mum likes these sonnels. She often calbains when she's tired. fs 21

La figure 20 montre l'oscillogramme, le spectrogramme, la courbe de F0 (en bleu) et la courbe d'intensité (en jaune) pour le mot *calbain* :

36. Je ne marque délibérément pas l'accent de mot.

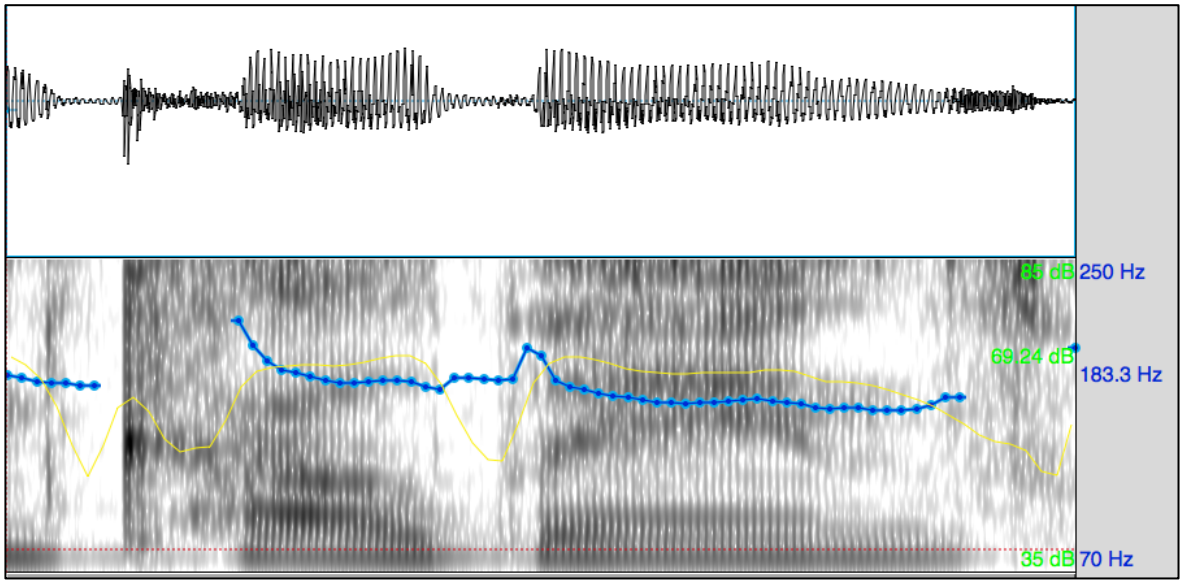


Figure 20 : oscillogramme, spectrogramme et courbe de F0 pour le mot *calbain* (fs 21)

Comme on peut le voir, aucun corrélat ne nous permet de savoir où la locutrice a placé l'accent de mot : F0 et intensité sont au même niveau sur les deux syllabes. L'oscillogramme confirme que l'amplitude du son est à peu près équivalente pour les deux syllabes. Les formants 1 et 2 non parallèles sur la deuxième syllabe montrent que la locutrice a bien produit une diphtongue. Les deux sons vocaliques sont relativement longs et les deux plosives initiales de syllabes ne nous apportent pas grande information non plus.

Le degré zéro de proéminence au niveau du mot est donc la réduction de la voyelle de la syllabe inaccentuée. On parle souvent de la voyelle mais dans les faits les consonnes peuvent jouer un rôle également, parce qu'elles sont moins longues dans ces syllabes-là, ou parce que l'aspiration est moindre pour les plosives sourdes par exemple. C'est exactement la même chose pour les monosyllabes, où l'on va avoir la réduction du mot, dans le cas des mots outils par exemple.

Tout de suite après le degré 0, le degré 1 de proéminence serait alors la forme pleine de la voyelle dans une syllabe non accentuée au niveau du mot, comme les syllabes finales des mots *expert* /ˈ*ekspɜːt*/, *menu* /ˈ*menjuː*/ ou *educate* /ˈ*edjukeɪt*/. Ceci correspond à que certains auteurs appellent l'accent tertiaire post-tonique (voir Wells 1990 par exemple). A l'étude du son [i], nous avons montré

que l'on peut considérer qu'il existe un niveau intermédiaire (Herment & Turcsan, à paraître). La voyelle [i], surnommée dans la communauté des phonéticiens le *happy* [i] (par analogie au phénomène connu sous le nom de *happY tensing*), n'est pas un phonème. Elle est en fait un allophone à la fois de /ɪ/ et de /iː/ selon sa position dans le mot. Ce son n'apparaît qu'en position inaccentuée, dans différents environnements que l'on peut résumer ainsi (*cf.* Herment 2010c pour davantage de détails) :

- en finale (*happy* /ˈhæpi/), ou dans les préfixes *be-, de-, re-, pre-* (*begin* /biˈgɪn/) : elle est alors allophone de /ɪ/ et correspond au phénomène de tension appelé *happY tensing* ;
- devant voyelle (*piano* /piˈænəʊ/ : elle est alors allophone de /iː/ et c'est un phénomène d'abrègement qui prend place.

Des mesures acoustiques nous ont permis d'affirmer que ce son [i] n'est en fait pas moins court qu'un /iː/, par contre c'est au niveau de la tension qu'une différence est marquée : il se trouve à mi-chemin entre /ɪ/ et /iː/. Nous avons également étudié le son [u], mais moins en détail. Sa position est un peu plus complexe (*cf.* Herment 2010c), mais il présente probablement les mêmes caractéristiques acoustiques. On voit dans le mot *educate* /ˈedjukeɪt/ que l'on trouve [u] en position inaccentuée comme un allophone de /ʊ/.

Au niveau de la proéminence, nous argumentons pour une position intermédiaire qui serait semi-réduite et nous proposons le schéma suivant :

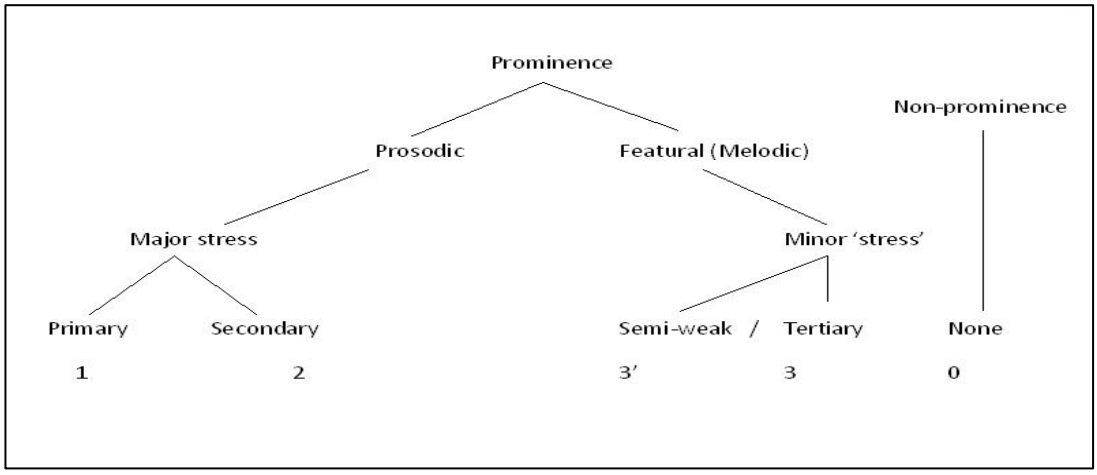


Figure 21 : relation de proéminence au niveau du mot (Herment & Turcsan, à paraître)

Ainsi un mot comme *periodicalize* /ˌpɪəriˈɒdɪkəlaɪz/ aurait un accent primaire sur /ɒd/, un accent secondaire sur /pɪər/, un que l'on pourrait qualifier de tertiaire post-tonique sur /laɪz/ et un accent qui serait intermédiaire, le 3' dans le schéma ci-dessus, sur /i/. Les autres syllabes ont le degré zéro de proéminence. Nous donnons donc un degré de plus par rapport aux quatre traditionnels que donne par exemple Wells (1990).

2.3. Tentative de modèle de hiérarchie des proéminences

2.3.1. Niveaux de proéminence

J'ai tenté de montrer dans les deux sous-sections précédentes qu'il paraissait plus logique de partir du suprasegmental pour arriver au segmental. Cette logique dérive principalement d'une approche perceptuelle, mais aussi d'une interprétation cognitive de la structure prosodique : dans quel ordre se font les proéminences ? Comment les range-t-on et comment les hiérarchise-t-on ? Selon la théorie du stockage-concaténation de Martin (2009), les hiérarchies sont analysées séquentiellement, mises en mémoire et concaténées afin que l'auditeur puisse reconstituer l'énoncé et son contenu :

> les événements prosodiques n'apparaissent pas à l'auditeur d'un seul coup au même instant. Au contraire, ils se produisent séquentiellement dans le temps et sont perçus l'un après l'autre en déclenchant pour chacun de ces

événements un processus partiel de décodage de l'information porté par les séquences syllabiques. (Martin 2009 : 113)

L'ordre de perception des proéminences est forcément l'ordre linéaire, temporel, mais les proéminences sont stockées, comparées les unes aux autres et donc finalement hiérarchisées en fonction des corrélats acoustiques bien connus que sont la fréquence fondamentale, la durée et l'intensité, mais aussi d'autres corrélats prosodiques tels que l'accentuation et la qualité vocalique. D'après ce que j'ai développé dans la section 2.2, on peut aisément hiérarchiser les proéminences sur une échelle qui compte 8 degrés (si l'on compte le degré 0) selon le modèle donné dans le tableau 2 ci-dessous (écrit en anglais pour plus de facilité due au vocabulaire). Il s'agit des proéminences perçues au niveau de l'unité intonative, en partant du suprasegmental vers le segmental :

7	tonic
6	head
5	accented syllables (primary stresses in accented words)
4	unaccented stressed syllables (primary stresses in unaccented words and secondary stresses in accented words ?)
3	unaccented stressed syllables (secondary stresses in unaccented words)
2	unstressed unreduced syllables
1	unstressed semi-reduced syllables
0	unstressed reduced syllables

Tableau 2 : échelle de degrés des proéminences dans l'unité intonative

Ces degrés ne me semblent cependant pas satisfaisants car comme expliqué en 2.2.5, si l'on distingue les accents primaires sur les mots accentués ou non, les accents secondaires sur les mots accentués ou non, alors pourquoi ne pas distinguer les syllabes inaccentuées (*unstressed*) sur les mots accentués ou non ? Prenons par exemple cet énoncé dans lequel la syllabe accentuée de *Jonathan* serait nucléaire :

Why hadn't Jonathan thought of this before ?

Si l'on considère que la syllabe accentuée de *before* (degré 4) est moins proéminente que celle de *hadn't* (degré 5), alors pourquoi ne considèrerait-on pas que la syllabe inaccentuée de *before* est aussi moins proéminente que celles de *Jonathan* ? Pourtant, la syllabe inaccentuée de *before* peut se prononcer avec un /i/, et serait donc au degré 1 alors que les syllabes inaccentuées de *Jonathan*, toutes deux prononcées avec un /ə/ sont au degré 0. Je crois que ces questions ne sont pas véritablement pertinentes car il faut distinguer plusieurs niveaux rythmiques. On ne se situe ici qu'au niveau de la syllabe finalement. Or c'est à un niveau rythmique supérieur à la syllabe qu'il faut aussi se placer.

2.3.2. Unités rythmiques

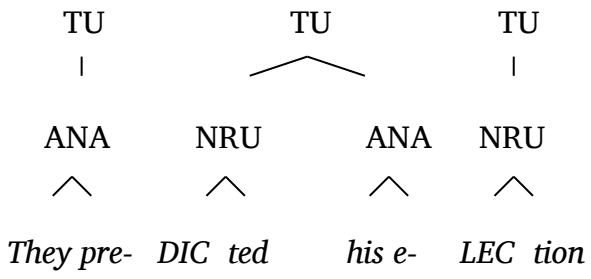
En matière de perception, les évidences méritent d'être rappelées car certaines étapes semblent indispensables, comme le fait Hewings dans son manuel :

> To make sense of what we hear, we need to be able to divide the stream of speech up into units (for example tone units, words or individual sounds) and to interpret what they mean. (Hewings 2007 : 16).

Comme je l'ai dit en 2.2, je pense en effet qu'au niveau de l'apprenant, mais aussi de l'auditeur en général, même natif, la distinction entre les différentes UI est essentielle : l'auditeur repère les noyaux pour comprendre l'énoncé. Sur le plan de la production, le placement des noyaux pose d'ailleurs problème aux apprenants, en particulier ceux de langue maternelle romane comme le français, ainsi que l'explique Cruttenden (2008) :

> The movability of the nucleus to different places in intonational phrases (particularly one-phrase sentences) should be regarded as one of the prime features of an English accent, making it easily intelligible to native speakers. Thus a tendency to constant final nucleus placement taking no account of (de-accentable) old information (common among learners with a background in Romance languages and in tone language, notably those of Africa and Asia) should be counteracted. (Cruttenden 2008 : 329)

J'ai déjà expliqué que je prends dans mes travaux pour unité de base l'unité intonative selon l'école britannique[37]. Cette unité (qui est à peu près l'équivalent de l'IP) fait consensus dans la littérature. D'autres unités rythmiques, inférieures à l'UI, sont utilisées par de nombreux auteurs, comme le pied, à la suite d'Abercrombie (1964) puis Halliday (1967). Il est constitué d'une syllabe accentuée suivie de toutes les syllabes inaccentuées jusqu'à l'accent suivant. Le pied est en fait inspiré de la théorie de Jassem (1952), qui avait proposé deux unités distinctes pour la structure rythmique de l'anglais : l'unité rythmique étroite (NRU, *narrow rhythm unit*) et l'anacruse (ANA, *anacrusis*). La première consiste en une séquence de syllabes contenant la syllabe accentuée d'un mot suivie des syllabes non accentuées du même mot et la deuxième est une suite de syllabes inaccentuées proclitiques. Au niveau mélodique, l'association de l'unité rythmique étroite et de l'anacruse suivante forme une unité tonale pour Jassem, et un pied pour Abercrombie. L'exemple suivant est tiré de Hirst (1998 : 59) :



Ces unités rythmiques sont donc indépendantes de la frontière de mot. Cela me semble problématique pour la perspective que j'ai choisie. Je ne remets nullement en cause ici la pertinence de ces unités pour d'autres champs d'investigation et d'autres études, mais dans l'approche de corpus qui est la mienne et dans l'approche perceptuelle liée à l'enseignement que je défends dans ce travail de synthèse, je crois que l'on ne peut pas passer outre les mots. Déjà dans ma thèse le problème était apparu puisque le corpus sur lequel je travaillais était étiqueté en mots. Dans les travaux qui ont suivi mon doctorat, j'ai continué à

37. *Cf.* la sous-section sur les constituants prosodiques en introduction.

travailler sur des corpus : l'annotation minimale est l'annotation orthographique (*cf.* chapitre 1), en mots donc. D'autre part, il me semble que lorsque l'on se place du point de vue de l'apprenant ou de l'auditeur, ce sont les mots que l'on essaie de capter, avec l'aide du rythme certes. Les résultats donnés à la figure 5 de Herment (à paraître b) montrent que pour l'emphase, les unités rythmiques qui ressortent dans les arbres de classification sont le mot et l'UI. Enfin, la théorie du stockage des entités se fonde aussi essentiellement sur le mot et la syllabe : selon Ladefoged (2005 : 107), les syllabes et les mots sont les unités qui nous permettent d'interpréter ce que nous entendons : « We can interpret what we hear straightaway by reference to syllables and words that have been stored in our brains ». Nous avons vu tout au long de 2.2 que le niveau de la syllabe s'est révélé important en effet, puisque même si les accents de phrase sont réalisés sur un mot, ils ont lieu sur la syllabe accentuée du mot. J'ai aussi pointé l'importance de la réduction vocalique pour la perception des mots accentués ou non au niveau de l'UI. Je ne place cependant pas la syllabe sur le même niveau hiérarchique que le mot, comme je vais l'expliquer ci-après.

2.3.3. Niveaux hiérarchiques et modèle

Allant toujours dans la même direction du suprasegmental vers le segmental, je crois que l'on peut donc distinguer un premier niveau hiérarchique de proéminence au niveau de l'UI qui a lieu sur les mots dans leur ensemble, comme présenté dans la figure 22 ci-dessous, en anglais à nouveau par commodité terminologique et par souci d'homogénéité avec les autres travaux consacrés au domaine :

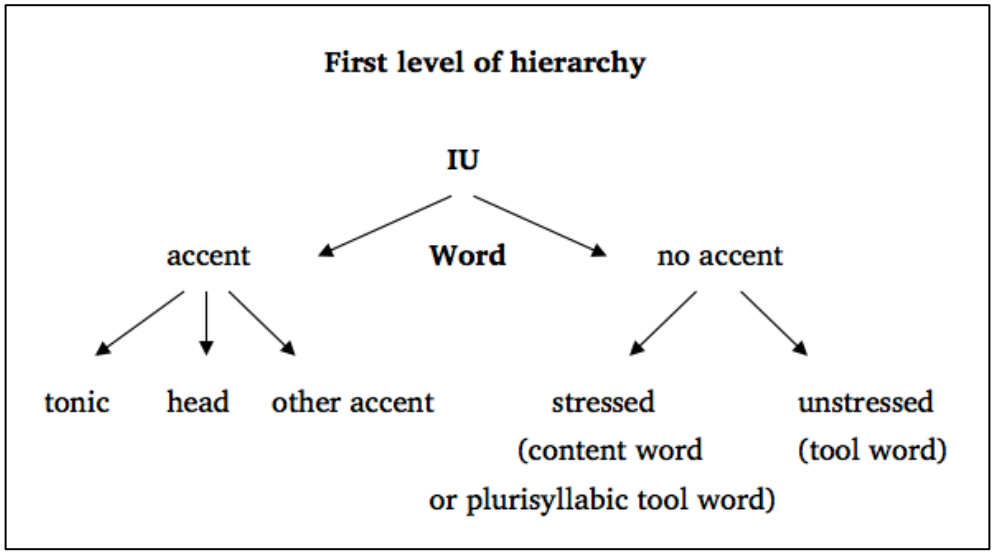


Figure 22 : premier niveau de hiérarchie des proéminences sur les mots à l'échelle de l'UI

La correspondance entre cette figure et les niveaux de proéminence donnés dans le tableau 2 est du fait impossible à faire puisque l'on se situe ici sur les mots et non sur les syllabes. C'est au niveau de l'UI que l'on va distinguer les mots accentués au niveau de la phrase (*accent*) des mots qui ne portent pas d'accents de phrase (*no accent*). Les termes *accented* et *unaccented* s'appliquent donc au mot entier et non à la syllabe. Parmi les mots accentués, on aura nécessairement celui qui contient la syllabe nucléaire (*tonic*), et éventuellement celui qui porte la tonique de tête (*head*) et d'autres accents entre la tonique de tête et le noyau. Tous les mots à contenu sémantique qui se trouvent dans le corps de l'énoncé (*body* ou *head*[38] selon les auteurs), c'est-à-dire entre la tonique de tête et le noyau, sont porteurs d'un accent de phrase. Dans le corps de l'énoncé, les mots qui ne portent pas d'accent de phrase ne peuvent donc qu'être les mots outils. On voit là toute l'importance des accents de phrase pour l'intelligibilité du message. On peut trouver des mots à contenu sémantique qui n'en portent pas, mais ils seront soit en position de pré-tête (*pre-head*), avant la tonique de tête, comme *of course* dans l'exemple de la figure 15 en 2.2.4, soit en position post-nucléaire, dans la queue

38. Pour Wells (2006) par exemple, la tonique de tête s'appelle *onset* et la partie ente *onset* et *tonic* s'appelle *head*. J'utilise quant à moi *head* pour tonique de tête (*cf.* figure 22) car *onset* peut aussi référer à l'attaque de la syllabe, que j'étudie dans d'autres travaux. J'emploie du fait *body* pour la partie qui se trouve entre la tonique de tête et le noyau.

de l'énoncé (*tail*), comme *guess* par exemple à la figure 17. Ces mots-là n'ont pas une importance particulière pour le message. Ils sont déjà connus s'ils sont en queue d'énoncé, ou bien servent à la transition ou à l'organisation du discours s'ils sont en pré-tête (comme *of course*). Ce sont les mots qui correspondent à *no accent* mais *stressed*. Tout comme les termes *accented* et *unaccented*, les termes *stressed* et *unstressed* s'appliquent au mot et non à la syllabe. J'emploie donc le terme *stressed* ici pour signifier que le mot comporte une syllabe qui porte un accent lexical, qu'il s'agisse d'un polysyllabe ou d'un monosyllabe[39] (par opposition aux mots outils monosyllabiques qui ne sont pas considérés comme *stressed*).

Une fois les mots proéminents repérés au niveau de l'UI, on peut parler d'un deuxième niveau de hiérarchie, qui se situe cette fois sur les syllabes, à l'échelle du mot. La figure 23 ci-dessous, en anglais toujours, représente le deuxième niveau :

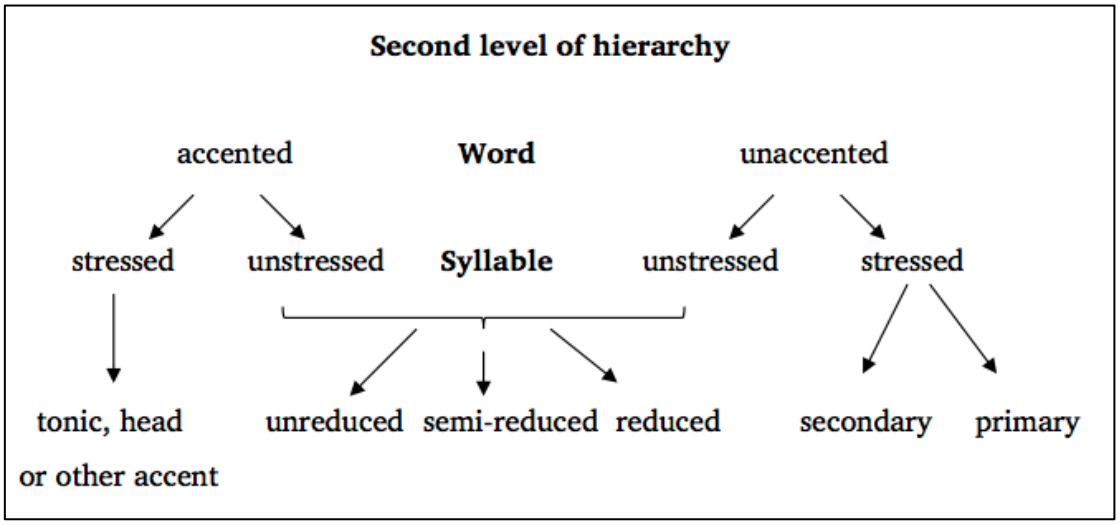


Figure 23 : deuxième niveau de hiérarchie des proéminences sur les syllabes à l'échelle du mot

On retrouve les *accented stressed syllables*, qui peuvent être le noyau, la tonique de tête ou les autres accents de phrase d'une UI, puis les *unaccented stressed syllables*, qui peuvent être des accents primaires ou secondaires dans des mots qui ne

39. *Cf.* 2.2.5 où je dis que je considère la syllabe unique d'un mot à contenu sémantique comme une syllabe accentuée au niveau lexical.

portent pas d'accent de phrase. Enfin, on a les syllabes inaccentuées, pour lesquelles la distinction *accented/unaccented* n'est pas pertinente, comme je l'ai dit plus haut. Pour les syllabes inaccentuées au niveau lexical (*unstressed*) on retrouve les 3 degrés de proéminence 0, 1 et 2 donnés au tableau 2. On peut donc faire correspondre les degrés du tableau 2 avec la figure 23 puisque l'on se situe au niveau de la syllabe :

tonic, head or other accent	unreduced	semi-reduced	reduced	secondary	primary
7 to 5	2	1	0	4 & 3	4

Figure 24 : correspondance avec les degrés de proéminence au niveau de la syllabe

Les syllabes à accent secondaire posent problème ici, comme expliqué en 2.2.5, puisqu'elles correspondent aux degrés 4 et 3.

On ne peut imbriquer les deux figures 22 et 23 puisque l'on ne se situe pas au même niveau de constituance. Ce que je cherche à démontrer est l'importance des proéminences sur les mots à l'échelle de l'UI sans prendre en considération la syllabe.

2.3.4. Corrélats acoustiques et prosodiques de la proéminence

Une fois les niveaux hiérarchiques déterminés, c'est sur leurs corrélats acoustiques et prosodiques qu'il est intéressant de s'arrêter. Je les ai déjà mentionnés dans les sections précédentes, à savoir F0, durée, intensité et qualité vocalique, mais il faut désormais les rattacher plus précisément à chaque niveau de proéminence.

2.3.4.1. F0, durée, intensité et qualité vocalique

Si un consensus se dégage dans la littérature autour des paramètres qui entrent en jeu dans la production et la perception de la proéminence, les différents auteurs ne s'accordent pas quant au paramètre le plus discriminant pour

tel ou tel niveau de proéminence. Cela n'a rien d'étonnant : Martin (2009 : 113) explique à propos de l'interprétation cognitive des proéminences « que les marques prosodiques entrent dans un réseau de contrastes relatifs, et n'ont pas de caractéristiques acoustiques absolues ». Il donne comme exemple la voix chuchotée, au cours de laquelle les contrastes sont plutôt exprimés par d'autres moyens que la F0, les durées en particulier. Mes travaux sur l'emphase et la focalisation (Herment & Hirst 2002, 2003, Herment 2006 et à paraître) ont aussi montré que pour mettre un élément en relief, plusieurs paramètres sont déterminants et co-agissent (voir aussi Cappeau & Hanote 2012) . Nous suggérons dans Herment & Hirst (2003) que l'énergie dépensée pour prononcer un mot est le critère essentiel dans la perception de l'emphase : si les phonèmes du mot sont allongés et que son intensité est plus forte, alors le mot sera perçu comme davantage proéminent. Voici donc une combinaison de deux paramètres que l'on voit à l'œuvre sur le mot *are* dans l'exemple ci-dessous :

Now it's slow progress but they are making progress fs 10

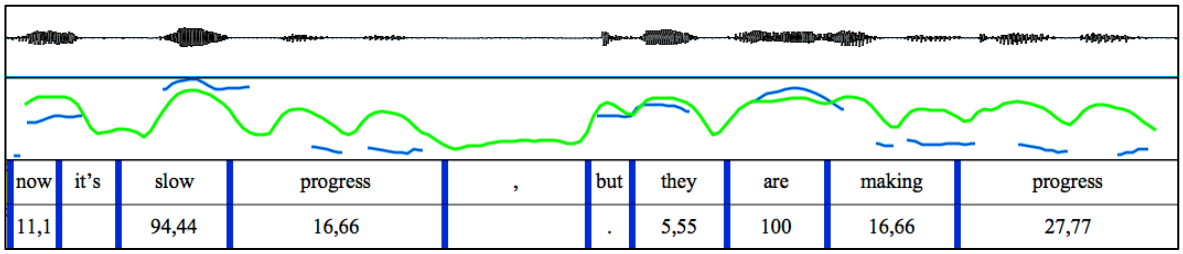


Figure 25 : courbes d'intensité et de F0 pour fs 10

La figure 25 montre deux courbes : la bleue discontinue représente la F0 et la verte est la courbe d'intensité. La tire sous les mots donne le degré d'emphase suite à un test de perception (Herment-Dujardin 2001). Les deux mots perçus les plus emphatiques sont *slow* et *are*, qui sont noyaux des deux unités intonatives. *Are* est très long et on voit que son intensité est forte. La qualité de la voyelle est bien entendu déterminante également : le mot en tant qu'auxiliaire est normalement réduit et prononcé /ə/ en contexte neutre. Le fait qu'il soit réalisé ici avec sa forme pleine /ɑː/ contribue à le rendre très proéminent. Pour le mot

slow, la fréquence fondamentale est très élevée. L'intensité est en général liée à la hauteur de fréquence fondamentale : si la F0 est élevée, l'intensité va être plus forte et contribuer à l'effet de proéminence, comme sur le mot *slow*.

Il est intéressant de constater également que lorsque l'on dit que l'intensité peut être déterminante, on entend une hausse d'intensité. Or j'ai montré dans ma thèse qu'une baisse d'intensité pouvait aussi être une marque de saillance. Je ne résiste pas à redonner l'exemple de la Duchesse de Kent, tiré du COEC. Les locuteurs parlent de leur cérémonie de fin d'études et des personnages célèbres qui étaient présents. Tous ont eu des stars de cinéma ou de télévision et cette locutrice s'est vue remettre le diplôme par un membre de la famille royale. Elle insiste donc sur le côté incroyable de la chose et met en relief *Kent*. On voit bien sur la figure 26 que contrairement à l'exemple de la figure 25, la hauteur mélodique n'entraine pas une hausse de l'intensité (comme plus haut, la courbe de F0 est donnée en bleu et celle d'intensité en vert). En effet, la locutrice chuchote presque le mot afin d'en augmenter l'effet dramatique :

Yeah | *I had the Duchess of* *Kent* | fs 22

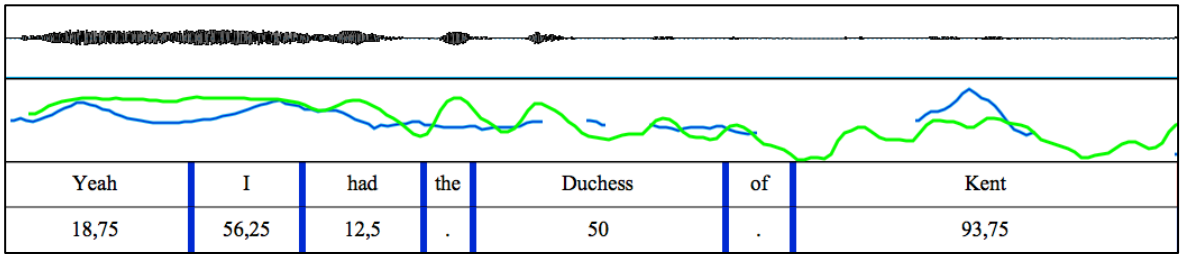


Figure 26 : courbes d'intensité et de F0 pour fs 22

Baisser la voix est un des procédés connus pour attirer l'attention des ses interlocuteurs. Il est à noter que dans cet énoncé à nouveau, la baisse d'intensité n'est pas le seul corrélat de la proéminence sur ce mot. Le mouvement mélodique sur *Kent* est ample, le mot est allongé et je montrerai en 2.3.4.5 (figure 29) que l'attaque du mot est également marquée.

La fréquence fondamentale est le corrélat qui est le plus discriminant tout seul, comme sur *violence* dans l'énoncé suivant, dans lequel la grande chute,

nettement visible sur la figure 27, rend le mot extrêmement emphatique (100% sur la deuxième tire).

Or would you say domestic violence is a private issue? fs 23

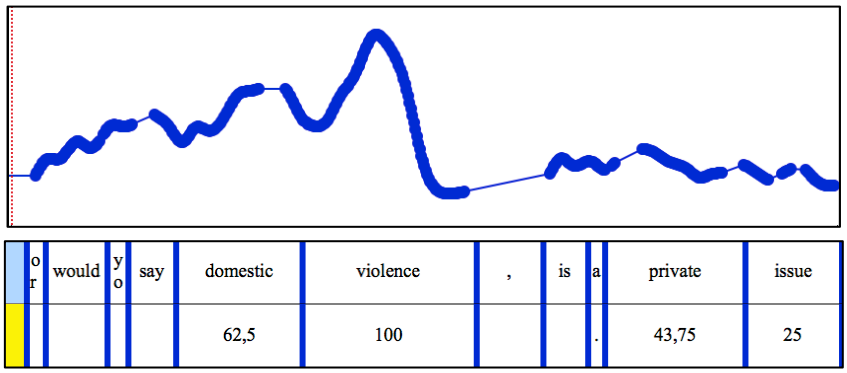


Figure 27 : courbe de F0 pour fs 23

Violence est noyau de l'UI, et ce qui suit est désaccentué. On voit sur la figure 27 que les mots à droite de *violence* sont prononcés plus bas et de façon plus plate que les autres mots de l'UI. La désaccentuation joue certainement un rôle ici aussi.

Ainsi F0, durée, intensité et qualité vocalique sont des paramètres déterminants dans la perception des proéminences, mais sont souvent combinées. Nous avons mis le doigt dans Herment & Hirst (2002) sur l'importance d'autres paramètres.

2.3.4.2. Sémantisme et syntaxe

Le sémantisme, le contexte et la syntaxe se révèlent être des critères pertinents dans la perception des proéminences. Je consacre le chapitre 3 aux relations entre prosodie et syntaxe et montre quels liens ils entretiennent et comment ils interagissent. Je vais donc m'intéresser plus particulièrement dans cette sous-section au sémantisme, avec les mots que j'ai appelés « sémantiquement non neutres » (voir 2.2.3). Nous montrons dans Herment & Hirst (2002) grâce un test de perception avec des phrases délexicalisées qu'ils sont décisifs dans la perception d'une proéminence mais insistons aussi sur le fait qu'ils sont

révélateurs encore une fois de la combinaison de plusieurs paramètres entrant en jeu dans la perception de la saillance. Dans l'exemple ci-dessous les mots *absolutely* et *freezing* en sont de bonnes illustrations :

and it was absolutely freezing fs 24

De par son sémantisme, *freezing* est déjà emphatique. La locutrice ne s'est pas contenté de dire « *very cold* » par exemple. L'ajout de l'adverbe *absolutely*, lui-même non neutre aussi, augmente encore l'effet d'emphase. Ainsi on pourrait s'attendre à ce que la prosodie ne soit pas particulièrement marquée ici. Or on constate que ces deux mots sont associés à un mouvement mélodique et à un allongement, tous deux renforçant encore l'emphase et soulignant qu'une combinaison de paramètres est à l'œuvre. La pause perçue entre les deux mots rajoute aussi à l'impression de proéminence.

2.3.4.3. Pauses

Les pauses peuvent être des marques d'insistance (Tannen 1984, Wells 1986, Selting 1994, Dahan & Bernard 1996). A nouveau, elles ne sont pas le seul paramètre et dans l'exemple ci-dessous, le mot *existing*, qui est entouré de pauses, est aussi prononcé avec une chute ample. Il est encore noyau de l'unité intonative :

Well I think the one thing it achieves | just by existing |is telling people | this cannot be a hidden problem anymore | fs 25

Les pauses ne font pas consensus dans la littérature car elles sont parfois l'expression d'une hésitation. Dans ma thèse, l'étude sur les pauses ne s'est pas montrée révélatrice pour la perception de l'emphase. Même l'expérience avec les stimuli resynthétisés (Herment 2006) n'a pas été concluante : nous avons enlevé des pauses et en avons rajouté. Les résultats montrent que cela ne change rien dans la perception de l'emphase. Il faut cependant nuancer ces propos : les durées des mots précédant les pauses n'ayant pas été abrégées dans nos segments manipulés, on a donc l'effet de la pause sans la pause puisqu'on sait qu'un mot précédant une pause est généralement allongé (Goldman-Eisler 1968, Repp 1992).

Il serait intéressant de pousser les expériences plus loin en faisant par exemple un test de perception sur des phrases porteuses dans lesquelles on insérerait des pauses plus ou moins longues et où le locuteur devrait dire dans quelle phrase le mot est le plus proéminent. La place de la pause, avant ou après, serait également intéressante à tester.

2.3.4.4. Qualité de voix

La qualité de voix (voir Laver 1980, Ní Chasaide & Gobl 2000) peut aussi influer sur la perception d'une proéminence (Tannen 1984), comme dans l'exemple suivant, où le locuteur a une intonation plutôt monotone. Le dernier mot est prononcé avec une voix rauque qui augmente la perception de la proéminence, même si ici encore ce n'est pas le seul paramètre : la F0 plus élevée et l'allongement du mot convergent pour la perception de la proéminence.

I think half way through the flight they give you a progress report and the pilot said well there's a 10-degree difference between Marseille and and London so I thought ohhh fs 26

La qualité de voix est très difficile à analyser et il s'agit certainement d'un phénomène marginal pour la perception de la proéminence en anglais standard (voir Coadou 2007 pour une étude sur la qualité de la voix et les accents régionaux). Ce paramètre méritait tout de même d'être mentionné.

2.3.4.5. Attaque du mot

Enfin, l'attaque du mot mis en relief participe parfois aussi à la perception des proéminences (Lilly & Viel 1998a). Nous montrons dans Herment (2009) que plusieurs paramètres peuvent jouer.

Si le mot commence par une ou plusieurs consonnes, un allongement de celle(s)-ci va augmenter la mise en relief du mot, comme on le voit à la figure 28 : le groupe consonantique initial /sk/ de *scooters* est particulièrement long sur les deux occurrences et le degré d'emphase (deuxième tire) augmente avec la longueur du /sk/ :

… *t to get his k get the scooters his kids' scooters…* fs 27

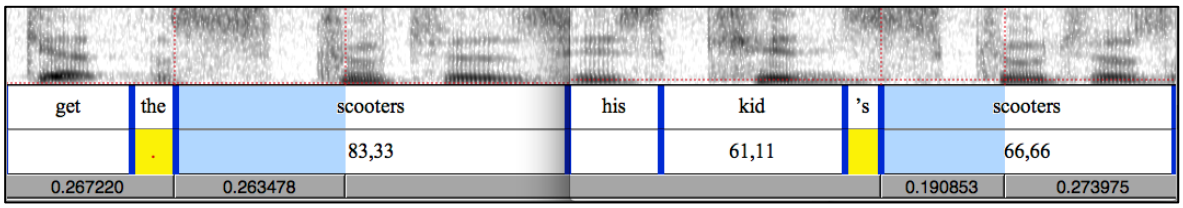


Figure 28 : allongement du groupe consonantique initial de *scooters* fs 27

Lorsque le mot commence par une consonne plosive, la tenue de celle-ci sera plus longue. L'exemple du mot *Kent*, déjà donné en figure 26 pour la baisse d'intensité, illustre bien ce phénomène : la figure 29 montre le spectrogramme, sur lequel on voit que la tenue est prolongée et l'explosion très longue :

Yeah I had the Duchess of Kent fs 22

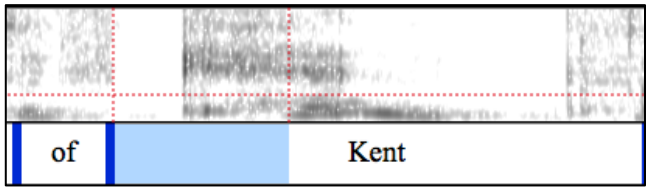


Figure 29 : allongement du /k/ : tenue et explosion plus longues fs 22

Dans le cas d'un mot commençant par une voyelle, l'insertion d'un coup de glotte est fréquente pour en augmenter la mise en relief. On voit nettement une plosive devant *actual* sur le spectrogramme donné en figure 30 :

or actual abuse fs 28

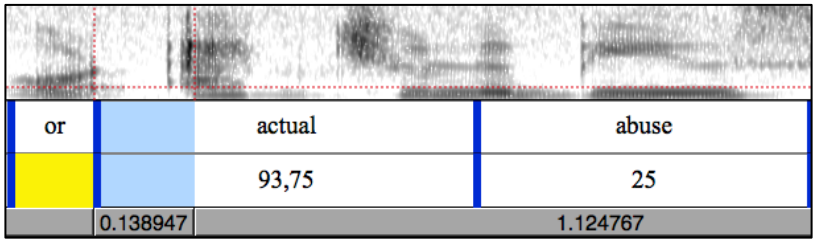


Figure 30 : coup de glotte devant *actual* fs 28

C'est donc une combinaison complexe et subtile de paramètres qui est à l'œuvre. C'est la raison pour laquelle il est si difficile de déterminer des corrélats bien précis pour chaque degré de proéminence.

2.3.4.6. Corrélats acoustiques des proéminences : tendances

Je peux quand-même tenter de relier des corrélats acoustiques et prosodiques à chaque degré de proéminence, en donnant donc plutôt des tendances et tout en gardant bien à l'esprit que c'est la combinaison de tous ces corrélats, avec souvent l'addition d'autres paramètres, qui va faire que l'on perçoit un mot ou une syllabe proéminents. Le tableau 3 ci-dessous donne ces tendances. Il est en anglais à nouveau toujours pour les mêmes raisons et la terminologie reprend celle explicitée et utilisée tout au long du chapitre.

Prominences	**Acoustic correlates**
1st level of hierarchy	
Words bearing the nucleus	pitch movement, duration, intensity
Words bearing the head	pitch height, duration, intensity
Words bearing an accent	duration and intensity, F0 to a lesser degree
Unaccented stressed words	duration and intensity to a lesser degree
Unaccented unstressed words	vowel quality, tempo
2nd level of hierarchy	
Tonic syllable	pitch movement, duration, intensity
Head syllable	pitch height, duration, intensity
Accented syllable	duration and intensity, F0 to a lesser degree
Unaccented primary stress (and accented secondary stress ?)	duration and intensity to a lesser degree
Unaccented secondary stress	duration and intensity to an even lesser degree
Unstressed unreduced syllables	vowel quality: full vowel
Unstressed semi-reduced syllables	vowel quality: tension
Unstressed reduced syllable	vowel and consonant quality: reductions

Tableau 3 : corrélats acoustiques de la proéminence : tendances

2.4. Conclusion : et l'emphase ?

J'ai tenté dans ce chapitre de hiérarchiser les proéminences prosodiques au vu des études que j'ai menées sur ce thème. J'ai distingué les proéminences sur les mots à l'échelle de l'UI et les proéminences sur les syllabes à l'échelle du mot. J'ai montré qu'il est difficile de faire correspondre des paramètres acoustiques précis à chaque niveau de proéminence, mais que des tendances se dégagent.

Qu'en est-il de l'emphase dans tout cela ? Comme je l'ai dit dans l'introduction à ce chapitre, l'emphase est une proéminence particulière, une façon non neutre, anormale de mettre un élément en relief, comme le dit Winkler (1984) : « 'emphatic' is the category for all sequences which seem to be non-neutral, non-normal, non-standard or non-factual/detached ». Mes travaux sur l'emphase ont montré que c'est une combinaison de plusieurs paramètres qui entre en jeu dans la perception de l'emphase, et que ces paramètres incluent le contexte et le sémantisme, mais aussi le style de parole (Herment à paraître b) et les locuteurs (Herment & Rouveyrol 2011). Les stratégies vont être différentes selon que l'on est en situation d'interaction plus ou moins agréable, ou bien en situation de monopole de la parole (Parlement ou journal), ou encore lorsque l'on commente un événement en direct comme les journalistes sportifs, lorsque l'on passe un entretien pour un travail, que l'on dispense un cours ou que l'on est en situation de débat politique. On rejoint ici les phonostyles de Léon (1993) (voir 1.1.2). Il existe en effet aussi une variation individuelle très forte, chacun utilisant des moyens différents de mettre un ou des élément(s) en relief. Mais il existe aussi des variations intra-locuteurs, chaque locuteur n'employant pas toujours les mêmes combinaisons pour rendre une emphase. C'est ainsi que nous affirmons dans Herment & Rouveyrol (2011) :

> [L]a variation est un phénomène essentiel. [...]Aucune formule de traits ne se retrouve deux fois, autrement dit : **l'énoncé emphatique est unique.**

Au terme de ce chapitre, je peux donc à nouveau distinguer proéminence et emphase. Les tendances données dans le tableau 3 concernent les proéminences, qui sont moins fortes que l'emphase. Il est plus aisé de trouver des traits communs aux proéminences qu'à l'emphase. L'emphase est une proéminence particulière, marquée soit par des traits de proéminence exacerbés tels que des mouvement mélodiques plus amples, un allongement beaucoup plus important, une intensité plus forte (ou beaucoup moins forte), soit par des traits autres comme la qualité de voix ou l'attaque des mots, soit par une combinaison de tous ces facteurs.

CHAPITRE 3

De l'emphase à l'interface prosodie/discours : d'une approche perceptuelle-acoustique à une approche discursive de la prosodie et des réalisations syntaxiques

La deuxième partie du volume des travaux, à laquelle ce chapitre va beaucoup faire référence, est consacrée à l'interface prosodie/discours. Après avoir étudié l'emphase au niveau acoustique dans ma thèse, c'est aux relations entre emphase et syntaxe que j'ai désiré m'intéresser ensuite, toujours dans le discours. Dans ce troisième chapitre de mon document de synthèse, je propose d'exposer la démarche qui m'a conduite à travailler dans ce domaine de recherche. Je souhaite montrer comment, en partant de l'emphase et d'une approche perceptuelle-acoustique, nous en sommes arrivées à élaborer une approche discursive de la prosodie et des réalisations syntaxiques. Tout comme la phrase précédente, ce chapitre commence à la première personne du singulier et se termine à la première personne du pluriel. En effet, j'ai débuté ce travail seule et ai ensuite étroitement collaboré avec ma collègue Laetitia Leonarduzzi. Les résultats exposés dans ce chapitre n'auraient pas vu le jour sans l'association de

mes compétences en prosodie et des compétences en syntaxe et pragmatique de L. Leonarduzzi.

3.1. Postulats de départ

3.1.1. La loi du moindre effort

La loi bien connue du moindre effort nous vient du linguiste américain Zipf (1949), qui a répertorié les mots les plus fréquents dans l'œuvre *Ulysses* de James Joyce et en a dérivé une loi, connue comme la loi de Zipf, qui prévoit que dans un texte donné, la fréquence d'occurrence d'un mot est liée à son rang dans l'ordre des fréquences. Cette loi est due selon lui au fait que l'homme a tendance à vouloir communiquer efficacement en faisant le moins d'efforts possible. Il semble s'agir d'un principe universel auquel Martinet (1955, 1960) renvoie lorsqu'il propose la notion d'économie linguistique, qui réside en « une synthèse de deux tendances contraires, à savoir celle à la réduction de l'effort et celle à la satisfaction des besoins communicatifs » (Peeters, 1992 : 45). La loi du moindre effort ne suppose pas le triomphe de la paresse, mais un équilibre entre l'effort et le résultat. Elle peut conduire à la réduction des entités du langage, mais peut aussi apporter son enrichissement.

C'est ce principe du moindre effort qui m'a amenée dans un premier temps à m'intéresser aux relations entre prosodie et syntaxe. Ayant étudié les corrélats acoustiques et prosodiques de l'emphase, je me suis demandée pourquoi une structure syntaxique était parfois associée à une emphase prosodique, et parfois non. J'entends par emphase prosodique une mise en relief « qui n'utilise aucun procédé syntaxique, mais uniquement des caractéristiques prosodiques » (Herment 2011). On a en effet en anglais la possibilité bien connue de réaliser une mise en relief uniquement grâce à la prosodie. Prenons la phrase

I went to the cinema |.

Si l'on prononce comme indiqué ci-dessus par le soulignage avec une mise en relief sur le pronom *I* en le mettant en position de noyau, alors la phrase prend le sens de

C'est moi qui suis allée au cinéma.

On note bien-sûr qu'en français, on est obligé de traduire en utilisant une clivée et que l'on ne pourrait vraisemblablement pas dire comme en anglais

?Je suis allée au cinéma |

avec un noyau sur le pronom « je ». Je nuance cependant cette affirmation car comme je l'expliquerai au chapitre 4, je pense que l'on peut entendre ces prosodies particulières en français, mais là n'est pas le propos ici, j'y reviendrai au chapitre 4. C'est donc aussi la comparaison entre le français et l'anglais qui m'a conduite à me tourner vers des structures comme les clivées. Si l'on prend une clivée en *it* basique comme

It's John who broke the vase,

on constate que malgré la structure syntaxique de mise en relief, la phrase hors contexte serait probablement prononcée avec une emphase prosodique sur *John.* Ceci est dû au fait que lorsque l'on emploie une clivée en *it*, la partie qui se trouve après le pronom relatif est normalement déjà connue, et donc désaccentuée en anglais. Mais pourquoi ne pas dire tout simplement

John broke the vase |

avec le noyau sur *John*, c'est-à-dire en plaçant *John* en position de focus grâce à la prosodie uniquement, puisque l'anglais en offre la possibilité ? D'autres questions viennent tout de suite à l'esprit : les deux énoncés

It's John who broke the vase |

et

John broke the vase |

ont-ils la même valeur ? En contexte isolé, on pourrait facilement penser que oui. Or le choix du locuteur se ferait en fonction du contexte de l'énonciation. En

réponse à la question *Who broke the vase?*, les deux énoncés sont possibles, mais ils véhiculeraient des nuances différentes.

On pourrait penser que toutes les mises en relief syntaxiques sont accompagnées d'une emphase prosodique. Ce n'est pas le cas. On trouve par exemple des occurrences de *do* appelé emphatique, c'est-à-dire inséré dans une phrase affirmative, qui ne portent pas d'emphase prosodique particulière, comme dans l'énoncé suivant, tiré du COEC, dont je donne le découpage en UI ci-dessous :

> *I think people who learn a language,* | *just by picking it up,* | *you know as adults,* | *then they probably do pick up all sorts of mistakes* | fs 29

Dans cet exemple *do* porte un accent de phrase mais n'est pas particulièrement emphatique prosodiquement. Il n'est même pas tonique de tête ici, comme on le voit sur la figure 31 ci-dessous. L'intonation pour cet énoncé est relativement plate et c'est le premier mot accentuable (premier mot à contenu), *probably*, dont la syllabe accentuée est tonique de tête.

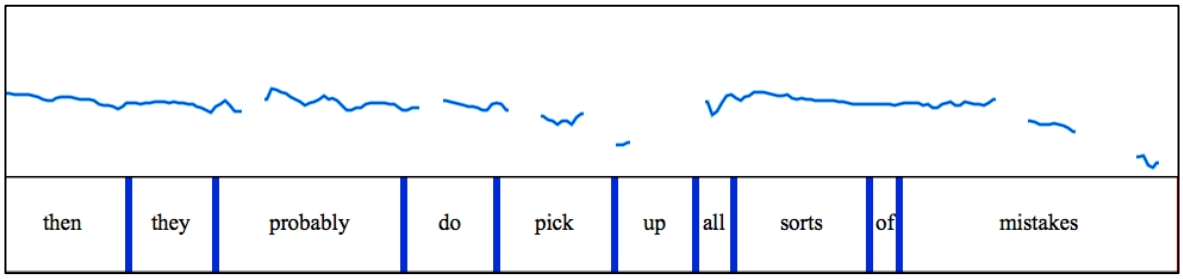


Figure 31 : courbe de F0 pour fs 29

Une rapide manipulation sous PRAAT permet de modifier les points de la courbe de F0 : j'ai remonté tous les points de *do* comme on le voit sur la figure 32, comme s'il était tonique de tête (les points gris clairs représentent les points d'origine). On entend très bien la différence : fs 30

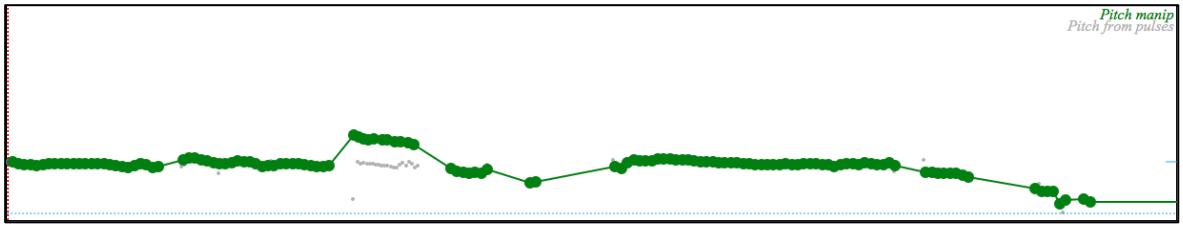


Figure 32 : fichier fs 29 manipulé sous PRAAT : fs 30

Cet énoncé montre que la réalisation authentique oblige à décorréler l'occurrence des marqueurs syntaxiques traditionnellement associés à la notion d'emphase de la réalisation prosodique proéminente interprétable comme une emphase. Ce *do* n'est pas particulièrement proéminent prosodiquement. Le fait de l'avoir inséré alors qu'il n'était pas nécessaire au niveau de la syntaxe n'est pourtant pas anodin, j'y reviendrai en 3.3.2.5. Une nuance de sens est donc exprimée et n'a aucunement besoin de l'aide de la prosodie.

A contrario dans l'exemple

I went to the cinema,

l'effet de contraste est présent grâce à la prosodie, sans le recours à la syntaxe.

Alors pourquoi utiliser parfois les deux procédés ? Suivant le principe du moindre effort, J'ai pensé que si deux marques, l'une syntaxique, et l'autre prosodique, étaient utilisées, il devait y avoir une bonne raison. La redondance en linguistique n'est pourtant pas rare. On a alors ici un paradoxe entre la loi du moindre effort et le principe de redondance.

3.1.2. La redondance

La redondance est une répétition d'information. Toutes les langues présentent des éléments redondants. Les accords de genre et de nombre par exemple sont redondants en français. L'énoncé suivant

Les petits garçons jouent

présente quatre marques de pluriel à l'écrit. Notons cependant qu'à l'oral on n'en a qu'une. C'est en effet plus particulièrement à l'écrit que les redondances sont présentes. Les redondances peuvent être également présentes à des niveaux

différents. Lorsque l'on pose une question comme dans cet exemple emprunté à Wit & Gillette (1999 : 6),

Do you like chocolate ?

on a deux marques d'interrogation : l'apparition de l'auxiliaire *do* et une intonation montante. On pourrait donner bien d'autres exemples.

Alors pourquoi ces redondances dans le langage ? Le terme redondance est en fait trompeur car il contient une connotation négative qui pourrait faire penser que puisque c'est redondant, c'est inutile. Or de nombreux linguistes pensent au contraire qu'il s'agit d'un attribut fondamental du langage qui facilite et assure même parfois la compréhension du message :

> Redundancy is a feature of an information source which insures that the communication receiver is able to reconstruct a message that has somehow suffered from transmission interference or deletion, and so interpret it satisfactorily. (Nubold & Turner, 1987 : 33-4)

Certains vont même plus loin en arguant que si la redondance vient à l'aide en cas de mauvaise compréhension, elle est aussi le seul moyen de transmettre un message complexe. Selon Wit & Gillette (1999), qui distinguent la redondance grammaticale et la redondance contextuelle,

> Besides containing grammatical redundancy languages also provide the possibility and freedom to the sender of a linguistic message to employ contextual redundancy, i.e., words or expressions that explicate, isolate, contrast, emphasize or even dramatize what was already contained in the message. (Wit & Gillette 1999 : 16)

C'est en cela que le terme « redondance » ne doit pas être mal interprété. La redondance linguistique n'a finalement rien de redondant, puisqu'elle est tout sauf inutile, comme le suggère Feller (1969 : 5) : « un mot de plus n'est pas forcément un mot de trop ». Car en effet, « Les mots supplémentaires sont complémentaires, et s'ils ne sont pas la source de lumière, ils sont l'ambiance » (Feller 1969 : 10). Feller poursuit en affirmant que « la façon de rédiger un message rajoute à l'information explicite toute une faune d'informations implicites qui en augmentent la valeur informatrice » (1969 : 11).

Ces dernières citations de Feller mettent en avant le caractère non redondant de la redondance, et de ce fait l'absence de contradiction entre la loi du moindre effort et la redondance.

Si j'applique cette idée à l'interface syntaxe/prosodie, l'opposition explicite/implicite est gênante. Est-ce la syntaxe qui est explicite et la prosodie qui vient rajouter une faune d'informations implicites ou bien est-ce plutôt l'inverse ? Cela me rappelle une série de dessins animés, « Pingu », mettant en scène un famille de manchots. Les dialogues sont dans un langage manchot (*Penguinese*) indéchiffrable. Grâce aux gestes et aux mimiques des personnages, mais surtout à l'intonation, les enfants du monde entier comprennent parfaitement bien les dialogues et suivent ce dessin animé sans aucun problème. On en revient aussi à la logique de l'acquisition : l'enfant dans le ventre de sa mère perçoit les contours mélodiques et non les mots. Je penche donc pour la deuxième solution, comme on l'aura compris : l'information n'est pas explicite sans l'intonation associée. Celle-ci fixe le sens. Quid de la syntaxe ? Est-ce elle qui apporterait une faune d'informations implicites ? Il semble difficile également d'affirmer que la syntaxe est implicite. La dichotomie explicite/implicite n'est certainement pas adaptée ici. Par contre je pense que c'est l'association d'une syntaxe particulière et d'une prosodie particulière, qui loin d'être tautologique, va véhiculer toute une faune d'informations supplémentaires qui augmentent la valeur informatrice de l'énoncé. C'est cette valeur informatrice supplémentaire que je cherche à identifier et caractériser dans mes travaux sur les relations entre prosodie et syntaxe.

3.1.3. Structures syntaxiques emphatiques

Suite à ma thèse et au postulat énoncé ci-dessus, à savoir qu'une structure syntaxique particulière associée à une prosodie particulière (qu'elle soit neutre ou non neutre, comme je le montrerai plus loin) apporte des nuances de sens supplémentaires, c'est donc logiquement aux structures syntaxiques traditionnellement décrites comme « emphatiques » dans les grammaires que j'ai

souhaité m'intéresser dans un premier temps. Arrivé *et al.* (1986 : 244) définissent ainsi l'emphase :

> Emphase : Le terme d'emphase - emprunté à la rhétorique - désigne les phénomènes d'insistance qui peuvent affecter les constituants de la phrase. L'emphase peut se manifester uniquement par des traits prosodiques. Elle peut entraîner en outre des transformations de la phrase. Ces transformations ont reçu, en grammaire générative, des descriptions diverses. Les phrases emphatiques sont caractérisées pas des phénomènes de détachement, de dislocation, de focalisation.

Ce sont ces transformations de la phrase pour mettre un élément en relief qui me préoccupent. Les structures de focalisation que sont les clivées ont attiré mon attention en tout premier lieu, et plus particulièrement les clivées en *it* (Herment à paraître) :

It's John who broke the vase.

Elles sont classées par les linguistes, comme les autres types de clivées, dans les structures focalisantes (Chomsky 1972, Prince 1978, Lambrecht 2001, ou Khalifa 2004), emphatiques (Larreya & Rivière 2005 ; Hewings 2005), ou de mise en relief (Lapaire & Rotgé 1991) et de *foregrounding* (Huddleston & Pullum 2002).

Les extrapositions (Herment 2009) m'ont également paru intéressantes car elles mettent en avant le jugement de l'énonciateur : le commentaire (*comment*) de l'énoncé est placé en début de phrase et le topique (*topic*) en fin (*cf.* Herment 2011) :

It was very impressive to see that.

Huddleston (1984 : 17) décrit l'extraposition comme une transformation qui déplace la proposition enchâssée à droite et met *it* à sa place.

Le *do* auxiliaire inséré dans un contexte affirmatif (voir Ranger 2004, Rivara 2004), communément appelé le *do* emphatique (Adamczewski & Delmas 1991 ou Souesme 1986), m'a bien entendu aussi interpellée (Herment 2009, 2010) :

I do like your tie.

La collaboration avec ma collègue Laetitia Leonarduzzi, qui est syntacticienne et pragmaticienne, nous a ensuite conduites à analyser en plus des

structures déjà mentionnées, que nous avons reprises, les dislocations à droite de groupe nominal (Leonarduzzi & Herment 2013b, Herment 2011) :

It looks swollen that foot.

Elles figurent parmi les phrases emphatiques données par Arrivé *et al.*. Ce sont des structures de thématisation (Miller & Weinert 1998) : le constituant qui se trouve en position initiale change.

Enfin, nous avons étudié les autres types de clivées, à savoir les clivées en *wh-* (Herment & Leonarduzzi à paraître) *:*

What we want is peace

et les clivées renversées (Herment & Leonarduzzi 2012) :

That's what we want.

Elles sont toutes deux classées de même que les clivées en *it* dans les structures de focalisation.

La caractéristique syntaxique commune à toutes ces structures est qu'elles se présentent comme une réorganisation de l'ordre canonique (voir Birner & Ward 2004 ou Huddleston & Pullum 2005 pour les extrapositions, Huddleston 1984 et Crystal 2003 pour les dislocations à droite de groupe nominal). Elles mettent toutes un élément en relief grâce à un procédé syntaxique. On peut donc considérer qu'elles sont marquées syntaxiquement. Vont-elles alors être également marquées prosodiquement ? Mais qu'entend-on par marqué ?

3.1.4. Théorie de la marque

Le concept de marque remonte aux linguistes de l'école de Prague et plus particulièrement Troubetzkoy et Jakobson :

> Enfin Trubetzkoy vint qui devait montrer (le 31 juillet 1930) que, de deux termes d'une opposition phonologique binaire, l'un était pourvu d'une certaine « marque » et l'autre dépourvu de la marque en question, et Jakobson, qui devait, le premier, comprendre la portée de cette découverte et l'étendre à la morphologie et au lexique. (Viel, 1984 : 17)

Cette première acception de la marque en termes binaires s'est étendue ensuite à la linguistique contemporaine pour signifier bien davantage et associer les formes non marquées à des valeurs par défaut :

> describing an observation as unmarked is often taken to mean that it is, for example, more frequent, natural, simple and predictable than the marked observation of the comparison set. The unmarked is also often referred to as the *default* member of a class : that is, it is the member to be assumed, the most basic member of the set. (Hume, 2011 : 80)

La controverse concerne les critères qui permettent de déterminer si un élément est « marqué » ou « non marqué ». Pour Chomsky & Halle (1968) et les phonologues générativistes dans leur sillage, les principes qui régissent la marque font partie des connaissances innées de l'être humain. La marque fait partie de la Grammaire Universelle, elle est prédéterminée chez les locuteurs et peut être prédite grâce au formalisme. Pour d'autres linguistes (*cf.* Hume 2004 ou Bybee 2011 par exemple), c'est une combinaison de facteurs tels que la fréquence d'usage, l'acquisition, la difficulté de production ou la stabilité diachronique qui permet de prédire la marque. Les relations de marque peuvent aussi être prédites grâce à des relations d'implication entre les langues, comme dans cette formulation de Zerbian (2012) :

> In phonology, markedness is used in the discussion of sounds and structures which occur less frequently than their unmarked counterpart, that are acquired later than unmarked sounds, that are more complex, unnatural or difficult to produce than unmarked sounds, and/or that are diachronically instable. Thus, evidence for markedness comes from typology, language acquisition, phonetics, and historical linguistics. Additionally, markedness can be determined by means of implicatures: a phenomenon A in some language is more marked than B if the presence of A implies the presence of B; but the presence of B does *not* imply the presence of A. (Zerbian 2012 : 447)

Si l'on en revient à la syntaxe et à la prosodie, on peut considérer sans trop prendre de risques qu'en anglais une structure syntaxique canonique serait non marquée. Dès lors que l'on aura une modification qui bousculerait l'ordre canonique, la structure sera considérée comme « marquée » (voir Kaltenböck 2004 et Smolka 2005 par exemple).

Au niveau de la prosodie, on peut faire les mêmes hypothèses et dire que la forme non marquée d'une UI pour l'anglais sera celle où l'on a le noyau placé sur le dernier mot à contenu sémantique de l'unité intonative avec un schéma descendant (*Last Lexical Item Rule*, voir par exemple Deschamps *et al.* 2004). En ce qui concerne la place du noyau, un consensus semble se dégager : les langues comme l'anglais, l'allemand ou le hollandais sont bien connues pour désaccentuer l'information déjà connue et donc faire remonter le noyau à gauche dans l'unité intonative (voir Ladd 2008 entre autres), comme dans l'exemple

I went to the cinema |

dont l'intonation présuppose que si l'accent nucléaire est sur *I*, *someone went to the cinema* est déjà connu. Or la désaccentuation de l'information ancienne n'est pas un universel du langage, comme l'a montré Cruttenden (2006). Ainsi le déplacement du noyau vers la gauche est marqué.

Nous avons cependant été confrontées au problème de l'association de la structure syntaxique marquée et de la forme prosodique : prenons toujours la même clivée en *it*

It's John who broke the vase.

On imagine aisément un contexte qui déclencherait l'utilisation de la forme clivée : quelqu'un dit *As I entered the room I saw that the vase was broken but I have no idea who has done that*, et la réponse serait notre clivée. Dans un tel contexte, il semble impossible de prononcer la phrase en mettant le noyau ailleurs que sur *John.* Cette prosodie constituerait-elle alors la forme non marquée de la clivée en *it* ? Si l'on prend une clivée en *wh-* classique telle que

What we want is peace

on s'attend à ce qu'elle soit prononcée avec le noyau sur *peace*, donc sur le dernier mot à contenu sémantique de l'UI, ce qui serait une forme non marquée. Quant à la clivée renversée comme

That's what we want

on aurait tendance à positionner le noyau sur *that*, en position marquée donc. Pour les extrapositions comme

It's clear who they are

l'élément extraposé (*who they are*) est censé être l'élément nouveau et sera donc accentué, avec un noyau en position non marquée. Enfin, les dislocations à droite de groupe nominal auront également un noyau en position marquée puisque l'élément disloqué est censé être connu, comme dans

It looks swollen that foot

où l'on s'attend à avoir le noyau sur *swollen.* On voit bien qu'un problème à la fois théorique et terminologique surgit ici. Comme pour la clivée en *it* ci-dessus, la forme prosodique non marquée pour toutes ces structures ne serait-elle pas au contraire la forme que l'on s'attend à entendre ?

La notion de marque est problématique aussi pour le contour mélodique : quand l'UI est terminale (en fin de phrase), on a un schéma descendant, mais quand l'UI est non terminale (en milieu de phrase), on a plus fréquemment un schéma ascendant. Prenons l'exemple

After some time, the bus stopped

Cet énoncé comporte deux UI, la première sera montante, car non terminale (c'est un contour continuatif, voir Du Bois *et al.* 1993), et la seconde sera descendante (contour terminal) :

After some <u>time</u>, R | the bus <u>stopped</u> F |.

Doit-on alors considérer le schéma ascendant comme marqué uniquement en position terminale ou bien dans toutes les positions ? Si on le considère marqué en position terminale, qu'advient-il des questions ? Ne doit-on pas plutôt considérer qu'un contour mélodique marqué est un contour non neutre ? Mais qu'est-ce qu'un contour non neutre ?

J'ai parlé jusqu'à présent de la place du noyau et des tons, je n'ai pas mentionné le découpage en unités intonatives. Je soulève la même problématique pour cette troisième dimension de l'intonation : reprenons l'exemple de la clivée en *wh-* :

What we want is peace

J'ai dit plus haut que l'on s'attend à avoir le noyau sur *peace*. Mais on pourrait avoir deux UI dans cet énoncé, avec deux noyaux donc, l'un sur *want* par exemple et l'autre sur *peace*, tous deux en chute :

What we <u>*want*</u> *F* | *is* <u>*peace*</u> *F* |.

Les noyaux sont en position non marquée ici, et les tons neutres. Qu'en est-il du découpage en UI ? Quelle est la prononciation marquée ? Celle avec une seule UI ou deux UI ? Intuitivement, on a envie de répondre celle avec deux UI. Cette forme prosodique pour une structure syntaxique donnée sera-t-elle alors marquée aussi ? Doit-on distinguer trois niveaux de marque en fonction des 3T ?

3.1.5. Une alternative : canonique/non canonique

Devant la complexité du phénomène de marque en relation avec nos structures syntaxiques, nous préférons parler dans nos travaux de canonique et non canonique, à la fois pour la syntaxe, à l'instar des linguistes travaillant sur le projet NOCANDO (*non canonical constructions in oral discourse: a cross-linguistic perspective*) (*cf.* Brunetti *et al.* 2011) et pour la prosodie. Ainsi une structure syntaxique violant l'ordre canonique, comme une clivée par exemple, sera considérée non canonique. Pour la prosodie les choses sont un peu plus complexes. Nous employons également les termes canonique/non canonique pour chaque structure que nous analysons et en référence à ces structures et non pas à la phrase simple qui serait prononcée dans un contexte neutre :

> What we shall call the canonical prosodic pattern is the pattern commonly acknowledged by linguists and which corresponds to the canonical use of the syntactic structure in terms of information structure. It slightly varies according to the structure under consideration, but in any case the pattern is made up of 1 IP and a neutral tone (F or R). What differs according to the syntactic structure is the place of the tonic (on the first or second part of the structure[)]. (Leonarduzzi & Herment 2013a)

Pour chaque structure, le patron prosodique canonique (nombre d'UI, place du noyau et mouvement mélodique) est donc défini en fonction de la structure informationnelle, comme expliqué en 3.1.4. C'est en fait sur la prononciation hors contexte et hors discours que nous fondons notre acception de départ de ce qui est

canonique ou non canonique, comme c'est aussi le cas dans la plupart des grammaires qui mentionnent une prosodie pour ces structures. Du point de vue des mouvements mélodiques, une prosodie canonique sera neutre, c'est-à-dire que l'on trouvera les tons neutres *fall* et *rise* par opposition aux tons non neutres *fall rise* et *high fall*[40].

C'est en analysant le langage en usage, dans le discours, que nous souhaitons confirmer, infirmer ou affiner le postulat de départ que deux formes différentes de mise en relief (syntaxique et prosodique) vont apporter « des valeurs informatrices supplémentaires » pour reprendre les termes de Feller (1969 : 11), *cf.* 3.1.2.

40. Je ne mentionne que ces quatre tons car le *high rise* n'a pas été trouvé dans nos études sur les structures syntaxiques mentionnées ici. Voir 1.3.2 et 3.3.3.3 pour davantage de détails sur les tons.

3.2. Le discours et la méthode perceptuelle

Ce sont donc les énoncés en situation de parole naturelle que nous avons voulu étudier. Il est reconnu que la parole de laboratoire a de grands avantages dans bien des domaines et il n'est nullement question ici de la remettre en cause. Bien au contraire la parole de laboratoire peut souvent apporter des éclairages sur des points qui restent flous en discours. Cependant, comme je l'explique au chapitre 1, dédié à une réflexion sur les corpus, chaque étude a des besoins particuliers et différents, et pour notre investigation de l'interface prosodie/syntaxe, c'est au discours que nous nous intéressons. A la suite de Harris (1952), les linguistes travaillant sur l'analyse du discours se sont penchés sur l'articulation du langage et du contexte, et sur les activités du locuteur. C'est en ce sens que nous voulons travailler, et par discours nous entendons le langage en usage (Halliday & Hasan 1976), résultant d'une combinaison d'informations linguistiques et situationnelles et revêtant une dimension communicationnelle. Ce positionnement ne nous empêche toutefois pas d'entrevoir les possibilités que pourraient nous offrir des expériences élaborées en laboratoire afin de tester les hypothèses qui ressortent de notre étude en discours (*cf.* chapitre 4).

3.2.1. Corpus : le parti-pris syntaxique et le primat de la structure

Mes premiers travaux dans le domaine de l'interface prosodie/syntaxe (Herment 2009, 2010b, à paraître a) se sont fondés sur l'analyse du corpus Aix-Marsec. Ce corpus regroupe de la parole dite naturelle, à savoir produite dans des situations de communication naturelles. La parole naturelle englobe la parole spontanée, mais on peut aussi trouver des monologues produits avec un script comme dans certaines émissions de radio (*cf.* chapitre 1 pour davantage de détails). Le corpus Aix-Marsec contient en fait très peu de parole vraiment spontanée, et très peu d'emphase aussi d'ailleurs, c'est la raison pour laquelle je l'avais écarté pour mon travail de doctorat. Dans le cadre d'une investigation de l'interface syntaxe/prosodie, il est évidemment bien plus simple de partir de la

structure syntaxique que de l'emphase prosodique. Il faut donc trouver des corpus transcrits et faire des recherches manuelles à partir de la transcription orthographique. Aix-Marsec semblait idéal car il est doté d'une annotation prosodique, les TSM (*tonetic stress marks, cf.* 1.3.2). Malheureusement, sur 5h30 de parole, je n'ai trouvé que 22 clivées en *it* et 22 *do* dits emphatiques. Le corpus est donc trop petit.

L. Leonarduzzi et moi nous sommes donc tournées vers le corpus ICE-GB (Aarts & Nelson 1999). ICE-GB contient à la fois de l'écrit et de l'oral, et sur le million de mots qu'il regroupe, 60% font partie de la composante orale, qui est bien entendu la seule que nous avons exploitée. Il s'agit cette fois d'un corpus de parole naturelle également, avec beaucoup plus de spontané. Les échantillons de parole incluent des dialogues et des monologues qui se répartissent comme indiqué dans le tableau 4 ci-dessous :

Dialogues (180)	**Private (100)**	face-to-face conversations (90) phonecalls (10)
	Public (80)	classroom lessons (20) broadcast discussions (20) broadcast interviews (10) parliamentary debates (10) legal cross-examinations (10) business transactions (10)
Monologues (100)	**Unscripted (70)**	spontaneous commentaries (20) unscripted speeches (30) demonstrations (10) legal presentations (10)
	Scripted (30)	broadcast talks (20) non-broadcast speeches (10)
Mixed (20)		broadcast news (20)

Tableau 4 : constitution du corpus ICE-GB[41]

ICE-GB ne contient pas d'annotation prosodique, par contre il a l'avantage d'être étiqueté syntaxiquement :

> The ICE-GB corpus was indeed designed primarily as a resource for syntactic studies [...]. Every text unit ('sentence') in ICE-GB has been syntactically parsed at function and category level, and each unit is presented in the form of a syntactic tree. The 83,394 trees in the corpus represent an invaluable resource for studies of the syntax of contemporary British English. (http://www.ucl.ac.uk/english-usage/projects/ice-gb/design.htm)

Il est possible de faire des recherches automatiques et de requêter des structures syntaxiques particulières, ce qui est d'un grand intérêt pour nous :

> ICE-GB comes complete with ICECUP. ICECUP allows you to perform a variety of different queries, including using the parse analysis *in* the corpus to construct Fuzzy Tree Fragments to search the corpus.
> (http://www.ucl.ac.uk/english-usage/projects/ice-gb/index.htm)

41. (http://www.ucl.ac.uk/english-usage/projects/ice-gb/design.htm)

Par contre les enregistrements ont eu lieu entre 1990 et 1993 sur des cassettes et ont ensuite été numérisés, ceci entraînant déjà une déperdition de la qualité acoustique. De plus, la parole ayant été enregistrée dans des conditions naturelles, elle est bien souvent de très mauvaise qualité à cause de bruits ambiants extrêmement perturbants parfois. Certains fichiers sont tout simplement inaudibles (fs 31 pour ne donner qu'un exemple). Nous avons malgré tout opté pour ce corpus, le parti-pris syntaxique et le primat de la structure prenant le dessus sur la qualité sonore.

3.2.2. Analyse aurale et approche croisée de la syntaxe et de la prosodie

En raison de la très mauvaise qualité sonore des fichiers son dans ICE-GB, aucune analyse acoustique n'a été possible et les analyses de la prosodie ont dû être conduites auralement.

Cela présente l'inconvénient de la subjectivité de l'annotation, dont j'ai discuté en 1.3.1. L'oreille est par définition subjective, on l'expérimente dans la vie de tous les jours lorsque l'on réalise que deux personnes écoutant la même personne au même moment n'entendent pas toujours la même chose.

3.2.2.1. Comment la mise en contexte peut changer la perception

S'il est connu que l'environnement a une influence dans la perception des phonèmes, peu d'études à ma connaissance traitent de ce phénomène au niveau suprasegmental. J'ai pourtant constaté à force d'écouter des segments de parole pour en définir la mélodie que le contexte peut influencer de façon significative la perception d'un contour. Il s'agit bien entendu toujours de contours plutôt neutres. Un contour ample sera perçu bien plus nettement.

L'exemple que je propose d'analyser est tout à fait révélateur de ce phénomène[42] :

42. Le journaliste explique qu'il y a une baisse du chômage depuis trois mois consécutifs et que les chiffres sont encourageants pour tout le monde. Voici le contexte avant et après : « and we could all be encouraged by them [the figures]. There's also encouragement for women workers. In recent months, it's men who've benefited from the slackening pace of unemployment. But now this seems to be helping women as well. »

In recent months it's men who've benefited from the slackening pace of unemployment. fs 8

J'ai déjà exploité cet exemple au chapitre 1 pour illustrer les discordances entre mon système d'annotation prosodique et les TSM. Je ne reviendrai donc pas sur ce point. Ce qui m'intéresse cette fois-ci est le contexte en relation avec la perception de la prosodie. Je prends mon annotation (mais comme expliqué en 1.3.2, même si les annotations diffèrent, la perception de la prosodie de l'énoncé reste la même) : à l'écoute, j'ai découpé relativement facilement cet énoncé en trois unités intonatives. Je perçois nettement les noyaux sur *months*, *benefited* et *unemployment*. Quand aux mouvements mélodiques, on a un *rise* sur *months*, un *fall rise* sur *benefited* et un *fall* sur *unemployment*. Ceci donnant[43] :

In recent months R | it's men who've benefited FR | from the slackening pace of unemployment F |.

La seule petite hésitation pourrait avoir lieu sur l'UI du milieu. Comme expliqué en 2.2.4, un auditeur non chevronné pourrait être tenté de confondre la tonique de tête avec le noyau. On pourrait croire que *men* est le noyau. Or il n'en est rien car on entend un mouvement mélodique descendant assez marqué sur la première syllabe de *benefited,* qui est donc le noyau. Le mouvement mélodique est aussi assez difficile à déterminer sur cette UI : j'ai d'emblée annoté ce mouvement comme un *fall rise*, avec chute sur *be-* et remontée sur *-ed*. C'est aussi l'avis des experts qui ont annoté le corpus avec les TSM puisqu'ils donnent le symbole '/ sur *benefited, qui* signifie *high fall rise*[44] :

it's men who've `/benefited |

Il n'y a donc aucun doute pour le mouvement mélodique. Notre perception est tout à fait en adéquation avec le contexte. La première UI est dite sur un ton montant simple, c'est un complément circonstanciel de temps en début de phrase.

43. Dans tout le chapitre, les frontières d'unités intonatives sont marquées par le symbole |, le noyau souligné et les mouvements mélodiques nucléaires notés F, R, HF et FR pour *fall*, *rise*, *high fall* et *fall rise* respectivement.

44. Rappelons que le *fall rise* simple n'existe pas dans leurs TSM, voir le tableau 1 au chapitre 1 (1.3.2).

Il est nécessairement prononcé comme une UI à part et souvent avec un ton montant, comme l'on fait lorsque l'on débute une phrase par *last week* ou *recently* par exemple. La même prosodie est produite pour un adverbe de phrase initial tel que *unfortunately* ou *apparently*. La deuxième UI est également non terminale, puisque la phrase n'est pas finie. On s'attend donc plutôt à un ton montant également. Une montée est d'autant plus attendue ici que l'on a une frontière d'unité intonative juste après le verbe, qui est pourtant prépositionnel d'après la structure de la phrase. La phrase ne pourrait pas s'arrêter après *benefited*, elle serait incorrecte. On attend donc le complément juste derrière. Comme on perçoit la chute sur la première syllabe du verbe, le schéma est forcément un *fall rise*.

Regardons maintenant attentivement la courbe de F0 et les tons Intsint pour le mot *benefited*, donnés en figure 33. On voit nettement sur la figure que la courbe de fréquence fondamentale est descendante sur *benefited* et ne remonte pas[45], tout comme les tons Intsint le montrent aussi, avec H(*igh*) D(*ownstep*) D(*ownstep*) et les valeurs de F0 qui les accompagnent : 195, 134 et 111 Hz, marquant bien la déclinaison. Si l'on réécoute plusieurs fois cet exemple en tentant de faire abstraction du contexte, du sémantisme des mots, on entend bien une chute.

45. La petite remontée que l'on voit à la fin n'est pas un mouvement mélodique, c'est une variation micro-mélodique (voir 1.4.3) : elle a lieu après l'explosion du /d/.

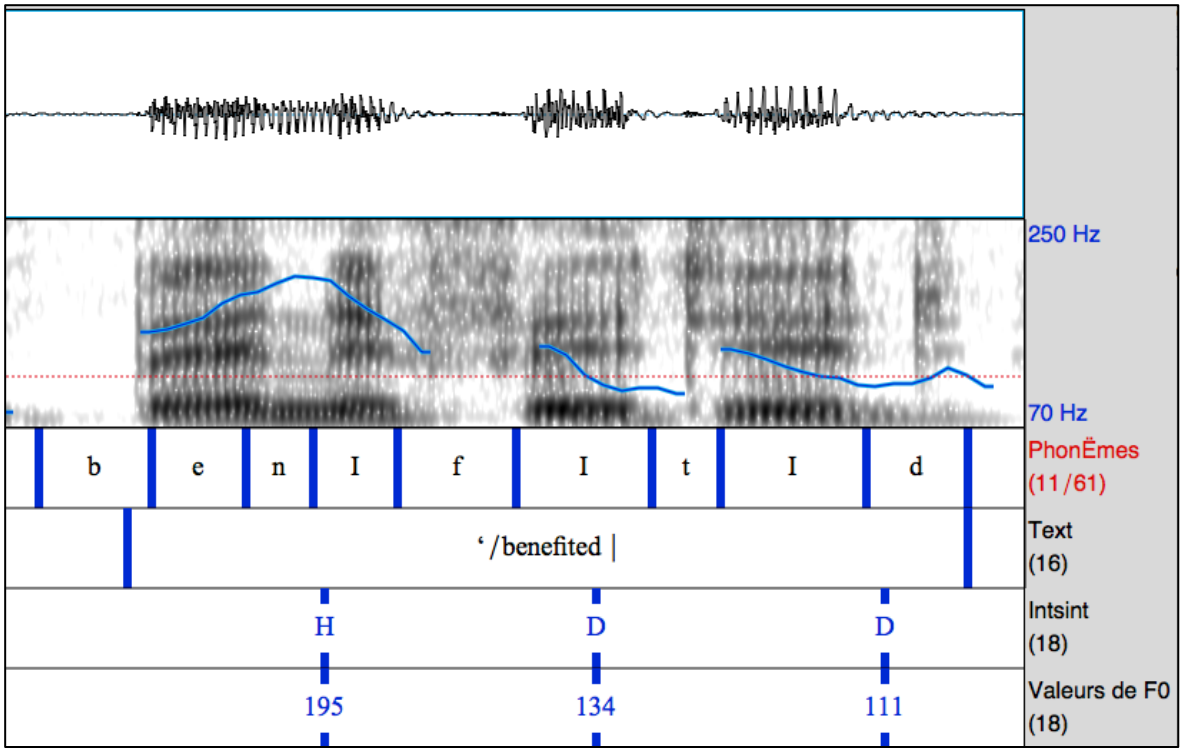


Figure 33 : courbe de F0 pour le mot *benefited* (fs 8)

Ainsi dans cet énoncé une chute est perçue comme un *fall rise* à cause du contexte. Si l'on regarde la courbe et l'annotation automatique, on voit clairement une chute. Si l'on écoute, on entend un *fall rise*. Alors comment faut-il étiqueter cette UI ? Doit-on se fier à son impression auditive ou bien doit-on tout vérifier sur la machine ? Cette question revient à se poser la suivante : travaille-t-on en perception ou en production, c'est-à-dire doit-on annoter ce que l'on perçoit ou ce que le locuteur produit effectivement ?

On sait que production et perception ne vont pas l'une sans l'autre. Nous avons montré dans Herment & Hirst (2002, 2003) et Herment (2006), que même si le côté « perception » est privilégié puisque les analyses sont fondées sur des tests de perception de l'emphase, on ne peut pas se passer du côté « production » étant donné que les analyses acoustiques sont réalisées sur la production des locuteurs.

Dans le cadre de nos études sur l'interface syntaxe/prosodie, nous ne regardons pour l'instant pas encore les critères acoustiques du fait de la qualité sonore médiocre de ICE-GB. Mais nous favorisons exactement la même approche,

à savoir une analyse perceptuelle, qui peut ensuite être suivie d'une analyse acoustique.

3.2.2.3. Comment la « mise en son » peut changer l'interprétation de la structure syntaxique

Ainsi la mise en contexte peut changer la perception auditive d'un énoncé. De la même façon, ce que j'appelle la « mise en son » (c'est-à-dire l'écoute) pour faire le parallèle avec la mise en contexte, peut aussi changer l'interprétation de la structure syntaxique. J'entends par là que lorsqu'on lit le script d'une conversation sans l'écouter, on peut donner certaines fonctions à des structures syntaxiques, que l'écoute peut éventuellement venir modifier. Les exemples bien connus de structures ambiguës désambiguïsées par la prosodie ne sont pas tous pertinents ici. En effet, bien souvent le contexte lève l'ambiguïté (voir Hirst 1977). L'énoncé

I thought he was married,

est ambigu à l'écrit hors contexte : il peut avoir les deux sens contradictoires de « je croyais qu'il était marié » ou bien « je le savais bien qu'il était marié » : à l'oral, on désambiguïsera en plaçant le noyau soit sur *married* pour la première interprétation, soit sur *thought* pour la deuxième. Mais on imagine mal que cet énoncé puisse être ambigu en contexte à l'écrit. Il existe pourtant des exemples qui peuvent l'être à l'écrit, même en contexte. Notre étude sur les limites entre extrapositions et dislocations à droite (Leonarduzzi & Herment 2013b) par exemple nous en fournit quelques-uns : des proposition en –ing telles que

It's wonderful meeting you

ou

Has it been a new experience working with disabled people

peuvent être soit extraposées, soit disloquées. D'après notre étude, seuls des critères prosodiques permettent de faire le départ entre les deux structures : si l'on a une seule unité intonative et le noyau sur la deuxième partie de la structure, alors il s'agit d'une extraposition, comme dans

It's wonderful meeting you F | just at this point after ten years is it when you were just leaving St Albans | fs 32

Si l'on a deux unités intonatives, le plus souvent séparées par une pause et ayant toutes deux un schéma globalement descendant, alors il s'agit d'une dislocation comme dans

Has it been a new experience F | working with disabled people F | fs 33

L'exemple plus classique des relatives appositives (voir les travaux de Auran & Loock 2010, 2011) ou restrictives fournit une autre illustration de ce phénomène. La phrase

They took the passengers who were injured to the hospital

pourrait tout à fait se trouver à l'écrit dans un contexte non désambiguïsant et seule la prosodie permettrait de distinguer la relative appositive

They took the passengers R | who were injured R | to the hospital F |

de la relative restrictive

They took the passengers who were injured R| to the hospital F |.

On pourrait aussi prendre l'exemple bien connu de Huddleston (1984) :

It was the vase that Agatha gave us.

Si l'on prononce en plaçant le noyau sur *vase,* alors on a une clivée en *it* :

It was the vase that Agatha gave us |.

Si au contraire on place le noyau en fin d'UI, sur *gave*, l'énoncé sera alors compris comme comprenant une proposition relative :

It was the vase that Agatha gave us |.

Tout comme pour l'analyse perceptuelle, l'analyse syntaxique et discursive reste sujette à la subjectivité : n'oublions pas qu'il est impossible de lire sans mettre une prosodie sur les phrases, même dans sa tête, et qu'il est donc impossible de lire sans interpréter.

3.2.2.4. Approche croisée de la syntaxe et de la prosodie

Fortes de ces réflexions, nous avons décidé de procéder à deux analyses séparées de la syntaxe et de la prosodie pour nos travaux et de croiser ensuite nos données. Ainsi L. Leonarduzzi a réalisé l'analyse syntaxique des structures. Elle les a extraites et a ensuite procédé à leur analyse discursive en tentant de dégager les fonctions discursives des structures observées et en en étudiant la structure informationnelle en fonction du contexte, sans écouter la bande son. Le contexte pouvait être très large et remonter assez loin dans le discours. Je me suis quant à moi concentrée sur le son : j'ai écouté les énoncés dans lesquels figurait la structure étudiée. Afin de réduire autant que possible la part inévitable de subjectivité, et ne pas tomber dans une démarche circulaire, j'ai écouté le moins de contexte possible. J'ai réalisé l'analyse prosodique des énoncés extraits par ma collègue en suivant la tradition britannique (*cf.* l'introduction), à savoir en m'appuyant sur les 3T : j'en ai déterminé le découpage en UI, la place du noyau et le mouvement mélodique.

Nous avons ensuite comparé nos données, favorisant ainsi une approche croisée de la syntaxe et de la prosodie.

3.3. Réinterprétation en discours et approche de la prosodie discursive

3.3.1. Remise en cause de la grammaire de l'écrit

Nous avons étudié les trois types de clivées : les clivées en *it* (*it clefts* : *it's John who broke the vase*), les clivées en *wh-* (*wh- clefts* : *what we want is peace*) et les clivées renversées (*rwh- clefts* : *that's what we want*). Nos résultats remettent clairement en cause ce que nous avions défini au départ (en nous fondant sur la littérature) comme la prosodie canonique des structures étudiées. Nous montrons qu'en discours ces prosodies canoniques sont en fait largement minoritaires, comme le tableau 5 le récapitule :

Structure (total nb of tokens)	**canonical prosody**	**% of canonical prosody**
it-clefts (154)	(F-)	11%
wh-clefts (-152)	(-F)	8%
rwh-clefts (151)	(F-)	2%

Tableau 5 : pourcentage de clivées prononcées avec une prosodie canonique

Pour les extrapositions (*It was very impressive to see that*) et les dislocations (*It looks swollen that foot*), la littérature n'est pas diserte sur le patron prosodique et le peu d'auteurs ayant commenté ce point se contredisent (*cf.* Leonarduzzi & Herment 2013b). Nous avons donc déterminé la structure canonique prosodique en fonction de la structure informationnelle la plus fréquente. Dans les extrapositions l'élément extraposé correspond normalement à une information nouvelle. Il est donc accentué. A l'inverse, dans une dislocation à droite, l'élément détaché représente une information ancienne, par conséquent désaccentuée. Les résultats montrent que la prosodie canonique de ces deux structures est aussi la plus fréquente à l'oral.

Si nous partons du principe que la forme canonique est la forme la plus fréquente, appliquant en quelque sorte la théorie de la marque (puisque l'on peut dire que de plusieurs formes, la forme non marquée est celle qui est la plus fréquente, *cf.* 3.1.4), alors nous remettons en cause la valeur canonique donnée au départ pour les clivées. Pour les extrapositions et les dislocations à droite, la forme attendue est la plus courante. Nous pouvons donc refaire un tableau avec cette fois-ci les prosodies les plus fréquentes dans le discours pour chaque structure :

Structure (total nb of tokens)	**most frequent prosody**	**% of most frequent prosody**
extrapositions (86)	-F	43%
dislocations (66)	F-	44%
it-clefts (154)	F/F	47%
wh-clefts (-152)	F/F	49%
rwh-clefts (151)	-F	80%

Tableau 6 : prosodies les plus fréquentes en discours pour chaque structure

Ce sont ces prosodies-là que nous pourrions considérer canoniques. Il s'agit cependant d'un problème délicat :

> La forme prototypique n'est pas nécessairement celle qui est statistiquement la plus fréquente. Elle entretient avec la forme à plus fort effectif un rapport complexe qui me paraît très proche de que Bourdieu appelle "un essentialisme probabiliste". Elle a "de fortes chances d'être telle ou telle chose", mais ceci ne réduit pas le prototype à une identité avec le plus fort effectif. (Ballier 2004 : 66)

Peut-être vaudrait-il mieux dire alors que ces prosodies-là sont canoniques à l'oral, mais que cela n'en fait pas pour autant des prototypes. Les prototypes seraient-ils alors dérivés de l'écrit ? Ces réflexions me conduisent à penser que la grammaire de l'écrit a ses limites et qu'il faudrait envisager une grammaire de l'oral.

La question suivante consiste à se demander pourquoi, dans le discours spontané, les réalisations prosodiques des clivées ne sont pas celles que l'on attend, celles qui sont décrites dans la littérature, celles qui seraient prototypiques. Un élément de réponse vient tout de suite à l'esprit : le contraste, qui est donné comme la fonction principale de la clivée en *it,* peut être exprimé par la prosodie uniquement (*cf.* 3.1.1). La clivée n'est pas vraiment nécessaire à l'oral. C'est d'ailleurs la raison pour laquelle on en trouve si peu (22 clivées en *it* dans tout le corpus Aix-Marsec). Il en va de même pour les clivées en *wh-* : les plus courantes à l'oral ne sont pas celles où l'on met en relief l'élément qui est rejeté en fin de phrase, mais celles dans lesquelles les deux éléments de la

structure ont une importance informative, comme dans l'exemple donné dans Herment & Leonarduzzi (à paraître) :

what I'm saying F / is that they're not going to be covering the windows F | fs 34

où la locutrice a été mal comprise de son interlocuteur et insiste. Cet exemple est tout à fait révélateur de ce qui peut se passer en interaction, alors que cela n'arrive pas à l'écrit. Il s'agit d'ailleurs d'une structure typique de l'oral.

Ces constatations s'appliquent également aux clivées renversées : on en trouve très peu qui sont contrastives à l'oral avec l'accent nucléaire sur *that* et la plupart marque en fait l'importance du présupposé. L'emploi de la clivée relève d'un procédé discursif de récapitulation, transition ou marquage du topique, comme dans l'exemple suivant où un étudiant a fait un exposé sur *Nature vs. Nurture* et termine la partie où il vient de définir ces notions avant de passer à une autre :

Right so that 's what nature and nurture is F | fs 35

Le tableau 7 montre la répartition pour chaque structure étudiée, entre la forme prosodique décrite dans les grammaires traditionnelles (canonique, au sens donné en 3.1.5) et les formes prosodiques effectivement réalisées (non canoniques). Lorsque le schéma prosodique est en une UI, le ton est noté suivi ou précédé d'un tiret. Lorsque le ton est avant le tiret cela signifie qu'il a lieu dans la première partie de la structure et inversement (F- ou -F par exemple). S'il est précisé « marked tonic » quand le noyau a lieu dans la deuxième partie de la structure, c'est qu'il ne tombe pas sur le dernier mot accentuable de l'UI. Lorsqu'il y a deux UI, les deux tons nucléaires sont donnés, séparés par une barre oblique (F/F par exemple). A nouveau s'il est précisé « marked tonic » après le ton, cela signifie que le noyau est en position marquée. On pourra se reporter à Herment & Leonarduzzi (2012) et Leonarduzzi & Herment (2013) pour des exemples pour chaque structure et chaque patron prosodique.

synt \ proso	Canonical		Non canonical	
	pattern	%	pattern	%
extrapositions (86 tokens)	- F	43	Marked tonic (- F)	10,5
			- HF	1
			F/F	18,5
			FR/F (+/- marked tonic)	13
			HF/(H)F	14
dislocations (66 tokens)	F -	44	Marked tone FR -	16.5
			HF -	1.5
			F -	3
			F/F	23
			F/F marked tonic	6
			HF/HF +/- marked tonic	4.5
			FR/R	1.5
it clefts (154 tokens)	F -	11	- F	19.5
			F/F	46.7
			F marked tonic /F	11
			HF/F or FR/(H)F	11.7
wh- clefts (152 tokens)	- F	8	F/F	49.3
			FR/F	28.3
			FR/F marked tonic	9
			R/F	5.5
rwh- clefts (151 tokens)	F -	2	- F	80
			- HF or -FR	16
			FR - or HF-	2

Tableau 7 : tableau récapitulatif des patrons prosodiques canoniques et non canoniques et de leur répartition pour chaque structure.

On se rend bien compte à la lecture de ce tableau que les catégories de l'écrit ne rendent pas compte des phénomènes effectivement observés à l'oral en condition d'interaction.

3.3.2. Réinterprétation en discours

Je propose dans cette sous-section l'analyse de quelques exemples afin d'examiner les paramètres de la réinterprétation en discours. Je vais donc me pencher sur les cas dans lesquels la prosodie ne correspond pas à la prosodie canonique de la structure.

3.3.2.1. Mise en relief non canonique

Dans certains cas, la prosodie met en relief un élément différent de celui que la syntaxe souligne. Dans certaines clivées un élément du présupposé est saillant en plus ou à la place de l'élément qui se trouve en position de focus syntaxique. Cet élément est contrastif et porte un ton non neutre. Ces cas sont donc loin d'être isolés puisque l'on trouve ce patron prosodique dans 16% des clivées renversées analysées et 9% des clivées en *wh-*. L'exemple qui suit en est une illustration :

That is what unites us HF,| it is not what divides us F | fs 36

Dans cette clivée renversée, le focus syntaxique *that* n'est pas particulièrement marqué prosodiquement, c'est le verbe du présupposé qui l'est et qui est d'ailleurs très emphatique : il est prononcé sur un ton ample. Le contraste est souligné par la présence de deux antonymes, dont l'un est négativé. Il est à noter que dans cette antithèse, la clivée renversée est suivie par une autre structure qui ressemble à une clivée renversée, mais qui commence par *it* au lieu de *that* ou *this* et dans laquelle le verbe du présupposé est également accentué, mais n'est pas emphatique (ton neutre F). La structure négative parallèle à la première renforce le contraste et le ton neutre sur *divides* est un autre moyen de renforcer l'importance de *unites*.

Nous avons pu repérer également des dislocations dans lesquelles un élément du détachement, pourtant normalement déjà connu, porte une prosodie particulièrement marquée, un *high fall* dans l'exemple donné ci-après. Il n'y a pas

contraste dans l'exemple ci-dessous mais une simple mise en relief, une emphase, qui est amplifiée par l'emphase déjà marquée par le même ton *high fall* dans la première partie de la structure. Il s'agit d'un enchaînement de tons identiques (*tone concord*, Wells 2006 : 85-86) : le ton utilisé dans la première partie de l'énoncé est répété dans la deuxième.

It's just amazing HF| the way she's so quick HF | at picking up the music F | fs 37

Dans l'extraposition qui suit, le découpage est extrêmement intéressant. Il est question dans cet exemple de personnes handicapées, dont la locutrice fait partie :

Uhm it 's nice for me to do something FR/ where I 'm moving F | because I teach now F | fs 38

La première frontière d'UI n'apparaît pas entre les deux parties de la structure syntaxique mais au milieu de l'élément extraposé, qui est *for me do something where I'm moving*. Le fait de mettre *me* en position nucléaire dans la première UI l'oppose non pas elle à une ou d'autre(s) personne(s) mais elle en tant qu'handicapée aux personnes n'ayant pas de handicap. Le début de la phrase (*it's nice*) et la partie *do something* de l'élément extraposé ne constituent pas une information très importante, par contre le deuxième morceau de l'élément extraposé (*I'm moving*) l'est, et cela est souligné par le découpage en UI, qui permet de marquer deux informations comme étant importantes, ainsi que la sous-section suivante va le montrer.

3.3.2.2. Informativité

J'ai explicité en 2.2.2. le concept d'informativité (*informativeness*) que nous avons développé dans Herment & Leonarduzzi (2012) et Leonarduzzi & Herment (2013a), inspiré des travaux de Chafe (1979, 1980, 1993), qui établit la distinction entre deux unités, l'unité intonative (*intonation unit)* et l'unité accentuelle (*accent unit*). L'unité intonative regroupe selon lui l'information qui occupe l'attention de la conscience du locuteur à un moment donné (1993 : 39). L'unité accentuelle est selon lui le domaine d'activation des informations nouvelles et/ou anciennes. Ainsi la structure prosodique d'un énoncé permet de

marquer ce qui est important, pertinent pour le locuteur au moment de l'énonciation. Cette notion recoupe celle d'information nouvelle/ancienne sans pour autant y être équivalente : une information ancienne peut se révéler pertinente au moment où le locuteur parle. C'est grâce au découpage en UI que l'on va pouvoir rendre compte de cela (voir 3.3.3.1). Les structures que nous avons étudiées peuvent être prononcées en une seule unité intonative, mais dans de nombreux cas, une frontière d'unité intonative est réalisée au milieu de la structure. Cette frontière peut être due à une incise ou à la longueur des constituants, mais bien souvent c'est l'informativité des constituants qui est dévoilée par le découpage en UI.

Parmi les clivées en *it* étudiées, celles qui sont prononcées avec deux UI (ou davantage) sont les plus nombreuses dans notre corpus : elles représentent 46.7% des occurrences analysées (voit tableau 7). Ce sont les clivées que nous avons nommées *all-informative* (Herment & Leonarduzzi 2012) :

It was in nineteen hundred and six F | that the Queen's great-grandfather King Edward the Seventh FR | decreed … fs 39

Cet exemple illustre la catégorie des clivées historiques. Certains les définissent comme stylistiques (Larreya & Rivière 2005 entre autres) ou de rappel de fait (Dubois-Charlier & Vautherin 1997). Souvent cette clivée est utilisée comme un moyen stylistique de rappeler une date à quelqu'un en faisant croire que l'on pense qu'il la connaît, alors que ce n'est pas forcément le cas, d'où l'importance informative de la date.

Les clivées en *wh-* sont aussi largement prononcées avec deux UI au moins, comme dans l'exemple donné plus haut et déjà commenté :

So what one wonders is FR / what went on in his mind HF | fs 40

Un effet de suspens semble présent dans cet énoncé, probablement provoqué par le ton FR, j'y reviendrai en 3.3.3.2. Seules 8% des clivées en wh- étudiées comportent une seule UI (voir tableau 7) : dans ces cas-là, la première partie de la structure n'apporte pas d'information particulière. On y trouve d'ailleurs souvent

des verbes neutres (*do, happen, mean, say*). L'essentiel de l'information se trouve dans l'élément clivé :

What you have to do is maybe check. F | fs 41

On a vu en 3.3.1. que pour les extrapositions et les dislocations, la prosodie canonique consiste à prononcer la structure en une seule UI. Lorsque la structure est divisée en deux UI, les deux parties de l'énoncé sont informatives :

> *It is up to the people of Kuwait F* | *to decide what kind of government they want HF* | fs 42

On entend un noyau sur *Kuwait*, qui est opposé à *us* plus tôt dans la discussion et se trouve donc contrastif, donc informativement essentiel, et une emphase sur *want* rendue par le ton ample *high fall.*

3.3.2.3. Récapitulation (clivées)

Certaines structures présentant une prosodie non canonique sont utilisées dans l'organisation du discours et plus particulièrement pour récapituler et offrir une transition. En toute logique avec la section précédente, elles ne sont prononcées qu'en une seule UI.

La fonction de récapitulation est connue pour les clivées en *wh-*, ou pour les clivées renversées comme nous l'avons vu en 3.3.1 avec l'exemple très clair

Right so that 's what nature and nurture is F | fs 35

La place du noyau n'est pas canonique, celui-ci se trouvant en fin d'UI sur la partie de la structure censée être présupposée.

On n'attribue généralement pas la fonction de récapitulation aux clivées en *it*. On la trouve pourtant dans le discours :

So So you mean that it's the nerves that feed them F |. fs 43

Dans cet exemple tout est déjà connu. L'information *the nerves feed them* vient d'être donnée et le locuteur reprend avec une clivée dont la fonction n'est pas du tout de mettre *nerves* en relief, mais bien d'articuler le discours.

3.3.2.4. Clarification (dislocations)

La dislocation à droite de groupe nominal est normalement prononcée en une seule UI, l'élément disloqué étant déjà connu. Le tableau 7 montre que 35% des dislocations étudiées sont découpées en deux UI. Lorsque l'on a le schéma F/F, la dislocation endosse une fonction de clarification dans le discours. Dans l'exemple que j'ai choisi pour illustrer ce cas, on imagine bien le locuteur se dire que peut-être son énoncé n'est pas très clair, on ne sait plus de qui il parle, il rajoute donc la dislocation, accompagnée du marqueur de discours *you know* (voir Schiffrin 2001 et les travaux du groupe OSliA[46] : Slamowicz 2011, Dahak 2013 entre autres), pour clarifier son propos :

He looks like him F/ you know the father F | fs 44

Dans les énoncés de ce type, une pause est souvent marquée avant l'élément disloqué, comme si le locuteur attendait de voir si son interlocuteur suit, et qu'il ajoutait une information après-coup :

That's a Tefl or teaching in English in Japan F / [uhm] the JET scheme F | fs 45

3.3.2.5. Renforcement de la relation prédicative ou de l'adverbe (do)

Je n'ai pas encore mentionné le *do* dit emphatique. On s'attend en général à ce qu'il soit accentué en position nucléaire. On pourrait dire qu'il s'agit de sa prosodie canonique. Or il peut être accentué en tonique de tête et non en noyau, renforçant ainsi la relation prédicative p au lieu de nier p', comme c'est le cas lorsqu'il porte un accent nucléaire (*cf.* Herment 2009). Dans l'exemple ci-dessous tiré de Aix-Marsec, la locutrice parle de la joueuse de tennis Martina Navratilova et explique qu'elle est presque imbattable, presque car elle a quand-même perdu 3 fois en 1986 :

I say almost | *because she did lose FR* | *3 times in 1986* | fs 46

Ainsi le *do* renforce le verbe au lieu de nier <she /not lose>.

46. Oral Spontané et Linguistique Anglaise, coordonné par Isabelle Gaudy-Campbell.

L'auxiliaire *do* inséré en contexte affirmatif peut aussi renforcer le sémantisme de l'adverbe qui le précède : ceci est marqué par le fait qu'il ne porte pas d'accent fort comme la tonique de tête ou le noyau (*cf.* Herment 2011) :

We certainly did have dire matches in Mexico fs 47

Le locuteur est un consultant sportif pour le football à qui l'on demande ce qu'il a pensé de la coupe du monde de Mexico. On pourrait traduire ici en français par « certes, nous avons vu des matchs médiocres […], mais nous avons aussi assisté à d'excellentes rencontres… ».

3.3.2.6. Résumé

En évoquant les paramètres de réinterprétation en discours, j'ai tenté de montrer que les catégories traditionnellement décrites pour les structures que nous avons étudiées ne recouvrent pas toutes les nuances de sens que l'on trouve à l'oral. Il semble que les écarts entre l'intonation et le message ou entre l'intonation et le contexte véhiculent des attitudes :

> I assume that at least *some* intonationally conveyed attitudes are also conveyed by some kind of mismatch, for example between intonation and the message, or between intonation and the context. (Wichmann 2000 : 145-6)

J'ai montré dans cette sous-section que des écarts entre la prosodie « attendue », canonique, pour une certaine structure syntaxique et celle effectivement réalisée sont porteurs de sens et endossent certaines fonctions dans le discours.

Tout comme pour l'emphase, réalisée grâce à une combinaison de paramètres variables, l'interaction de l'intonation avec d'autres paramètres permet de véhiculer du sens :

> It is extremely unlikely that there are any attitudes which are conveyed uniquely by intonation… It is possible that intonation patterns that are regarded as attitudinally marked use the intonation system in an unexpected way, and possibly in conjunction with other linguistic patterns. (Knowles 1987 : 205-6)

Nous sommes de plain pied dans une approche pragmatique de l'intonation (*cf.* Herment & Leonarduzzi 2012), entièrement dépendante du contexte.

3.3.3. Vers une approche discursive de la prosodie

L'interaction de la prosodie et de la syntaxe permet de donner des nuances à certaines structures syntaxiques que l'écrit autorise nettement moins. C'est la raison pour laquelle les grammaires traditionnelles décrivant les structures syntaxiques que nous avons étudiées pêchent selon moi à en décrire toutes les finesses. Pour la même raison, les manuels, ouvrages ou articles traitant d'intonation et de prosodie ont bien du mal à décrire des invariants prosodiques en lien avec des fonctions pragmatiques puisque ces dernières sont nombreuses et varient en fonction des structures. En ce qui concerne les structures syntaxiques non canoniques que nous avons regardées de près, on peut cependant dégager des tendances, en discours. Je propose dans cette sous-section de formaliser la réflexion précédente en l'appréhendant sous l'angle des 3T (*tonality*, *tonicity* et *tone*), afin de montrer, en discours, quelles fonctions recouvre chacun des trois types de décision que le locuteur doit prendre.

3.3.3.1. Le découpage en UI (tonality)

Le découpage en unités intonatives est révélateur de l'informativité de chaque constituant. J'ai développé ce point en 3.3.2.2, je ne m'attarderai donc pas sur cette section. Dans les structures étudiées, si le locuteur choisit de découper en deux UI et donc de réaliser deux noyaux, cela signifie que les deux parties de la structure sont informatives pour lui, même si elles viennent d'être mentionnées comme dans l'exemple ci-dessous[47] où la locutrice vient juste de dire *it is not just one person that is responsible* :

It is the system F | that is responsible F | fs 48

La pause marquée entre les deux parties de la structure vient ajouter un effet encore plus dramatique que la locutrice veut créer. Elle insiste sur l'idée de responsabilité collective, qui lui semble absolument cruciale.

47. Voir aussi tous les autres exemples donnés en 3.3.2.2.

3.3.3.2. La place du noyau (tonicity)

Contrairement à la tonalité, la tonicité est liée à la notion d'information nouvelle et ancienne, au niveau de l'UI. Comme je l'ai dit en 2.2.3 un noyau déplacé vers la gauche signale que ce qui se trouve à droite du noyau est de l'information ancienne, déjà connue, qui n'a pas besoin d'être réaccentuée au moment précis de l'énonciation. Des illustrations de ce phénomène sont observées dans toutes les structures étudiées, comme dans cette extraposition où *someone is involved* est déjà connu :

Then it doesn't matter who's involved F | fs 49

La position du noyau permet dans certains cas de désambiguïser une structure syntaxique, comme nous l'avons montré dans Leonarduzzi et Herment (2013b) : un énoncé comme

It's wonderful meeting you

peut être soit une extraposition si le noyau est placé à droite :

It's wonderful meeting you F |

soit une dislocation si le noyau est déplacé sur la gauche (voir 3.2.2.3) :

It's wonderful meeting you F |.

La place du noyau peut aussi impliquer un contraste, en particulier lorsque le noyau tombe sur un mot outil (*cf.* 2.2.3) comme dans cette clivée en *wh-* :

What I want FR | *is some new people FR* | *telling me some new lies F* | fs 50

Dans cet énoncé, la place du noyau montre que *want* est déjà connu (en fait on le reconstitue grâce au contexte), mais le fait d'accentuer un pronom qui est normalement non porteur d'un accent de phrase contribue à le mettre en relief et surtout à le rendre contrastif : « ce que moi je veux… ».

Enfin, la place du noyau peut aussi contribuer à mettre en relief une attitude émotionnelle de la part du locuteur. Dans l'exemple ci-après, la locutrice vient de dire qu'une guêpe était rentrée dans le métro. La dislocation de *poor wasp* permet à la locutrice d'exprimer son sentiment pour cette malheureuse guêpe :

And it got out | *this poor wasp F* | fs 51

3.3.3.3. Les tons (tones)

Je distingue les tons neutres, *fall* et *rise*, des tons non neutres, *high fall* et *fall rise* (voir 1.3.2). Le ton *high rise* n'a jamais été perçu dans les structures étudiées.

Lorsque l'UI n'est pas terminale, un *low fall* est la marque d'une information majeure (Wells 2006 : 72 ; Tench 1996 : 81), contrairement au *low rise* qui est plutôt utilisé pour une information mineure (Wells 2006 : 72). Ainsi dans nos structures, la prosodie R/F se distingue de la prosodie F/F. Dans le cas de R/F, la coupure d'UI est due à la longueur de la phrase et on marque que le contenu de la première UI n'est pas important grâce au ton montant, comme dans cette clivée en *wh-* où l'incise au milieu de la structure impose la frontière d'UI :

what we will be seeing very shortly R / on April 1st to be precise R / is the advent of all sorts of changes within the NHS F / fs 52

Dans les cas où les deux parties de la structure sont importantes, on aura donc plutôt F/F :

It is very odd F/ this whole aspect of code switching F fs 53

Les tons non neutres représentent les formes prosodiques marquées. Les formes canoniques sont toujours prononcées avec un *fall* ou un *rise* selon qu'elles sont importantes ou pas. Les tons non neutres revêtent différentes fonctions selon la structure dans laquelle on les trouve, comme je vais le montrer ci-après.

Le *high fall* va servir à mettre une emphase sur un élément : dans l'exemple suivant, le *high fall* sur *mind* rend ce mot emphatique. Cela n'aurait pas été le cas avec un *low fall*.

So what one wonders is FR / what went on in his mind HF fs 40

Si cet élément est déjà en position de focus syntaxique comme dans une clivée en *it* par exemple, alors il sera souvent contrastif. Le *high fall* renforce alors la focalisation déjà présente grâce à la syntaxe.

Indeed it may be because they are ideals HF/ and not truths F/that they are so deeply attached to them F fs 54

Le *fall rise*, quant à lui, indique toujours une attitude particulière de la part du locuteur, qui peut prendre des formes variées :

- une plus grande implication (*implicational fall rise* chez Wells 2006), comme dans les deux exemples ci-après :

It looks swollen that foot FR fs 55

La locutrice signifie avec le ton *fall rise* qu'elle aimerait en savoir plus et qu'elle voudrait que son interlocuteur lui raconte comment il s'est fait mal.

It's not the nasal field FR / that decussates F fs 56

Ici le contexte est un oral de faculté et le *fall rise* trahit l'empathie de l'enseignant pour l'étudiante. Il la prévient gentiment en quelque sorte qu'il faut faire attention de ne pas confondre *nasal field* et *retina*.

- le désir d'attirer l'attention de l'interlocuteur sur ce qui suit (voir entre autres Brazil 1997 ou Gussenhoven 2004) dans le clivées en *wh-* et dans les extrapositions. Le *fall rise* sur la première UI de la structure indique que l'on apporte peu d'information dans cette première partie et que l'on accorde davantage d'importance à la suite :

So it 's important FR/ that every time a product leaves one of these stages F/that it works properly F fs 57

Ce ton peut aussi être réalisé pour s'assurer de ne pas perdre l'attention de son interlocuteur, et donc de ne pas se faire interrompre et perdre son tour de parole, en lui signifiant que la suite vaut la peine d'être entendue. C'est le cas dans l'énoncé qui suit, où le locuteur hésite un peu, il doit chercher ses mots, et le *fall rise* est un procédé lui permettant de mettre un peu de suspens pour que l'on ne lui coupe pas la parole pendant qu'il cherche ses mots.

what happened is uh FR / they caught her without a licence F fs 58

- l'expression d'un contraste (Cruttenden 1986, Tench 1996) : cette fonction du *fall rise* est plus connue. L'énoncé ci-dessous oppose *saw* et *assumed must be there*. Le premier FR n'annonce pas l'importance de la dislocation qui suit, mais le contraste. La dislocation est d'ailleurs prononcée sur un ton montant, ce qui marque qu'elle n'est pas particulièrement informative. Elle est porteuse d'un

accent car elle se trouve dans une UI séparée à cause de la structure qui fait le parallèle entre *something you saw* et *something you assumed must be there* :

is that something you saw FR / this piece of jagged metal R/ or something you assumed must be there FR fs 59

3.4. Conclusion : l'interface prosodie/discours

J'ai souhaité nuancer mes propos en disant que pour les structures étudiées, des tendances se dégageaient. Pour donner un exemple, il est vrai qu'un *fall rise* sur le focus syntaxique d'une clivée en *it* réalisée avec deux UI ne peut pas avoir le même sens qu'un *fall rise* sur l'élément présupposé d'une clivée en *wh-* réalisée aussi en deux UI puisque le présupposé et le focus sont inversés au niveau de l'ordre de la phrase. Mais si l'on prend ces considérations en compte, je pense que les tendances que j'ai données sont valables pour le discours en général. Le développement qui constitue ce chapitre sur l'interface prosodie/syntaxe m'amène à parler d'interface prosodie/discours. J'ai montré que les prosodies décrites pour l'écrit volent en éclat en discours spontané ou semi-spontané et je prône une approche discursive de la prosodie et des réalisations syntaxiques.

Chapitre 4

Perspectives

Les pages précédentes m'ont permis de poser le cadre méthodologique et théorique dans lequel je travaille, d'exposer les résultats de mes recherches et de tenter de défendre certains points de vue. Dans le présent chapitre, c'est vers l'avenir que je souhaite me tourner en développant des perspectives de recherche future que mes travaux actuels entrouvrent. Je détaillerai trois axes principaux. Deux d'entre eux sont en relation directe avec les thèmes de recherche exposés plus haut, à savoir les proéminences prosodiques et l'interface syntaxe/prosodie. J'essaierai de montrer quels prolongements mes travaux actuels laissent supposer et vers quels nouveaux horizons ils peuvent mener. Je développerai ensuite un troisième axe qui est moins présent dans les chapitres précédents mais sur lequel j'ai déjà commencé à travailler, celui des interphonologies, à la fois inter-langues avec les études sur les apprenants, et intra-langues avec l'étude des variations dialectales. Le cadre méthodologique et théorique reste le même : mes analyses se fondent sur des études de corpus et la méthode perceptuelle est privilégiée même si l'utilisation d'outils d'annotation et de traitement de la parole font partie intégrante de mes recherches.

4.1. Proéminences prosodiques

Le chapitre 2 de ce document de synthèse traite des proéminences prosodiques, j'ai tenté d'en modéliser la hiérarchie et ai mis au jour le problème de l'accent secondaire dans une conception phonologique qui propose une interface entre hiérarchie accentuelle et hiérarchie de proéminences. Je suggère dans un premier temps une expérience qui permettrait de tester les hypothèses émises en 2.2.5. J'ai également mis en avant l'importance de la tonique de tête dans le chapitre 2. Je montrerai dans un deuxième temps quels genres d'études on pourrait mener sur cette thématique. Toujours à propos des proéminences, mais au chapitre 3 cette fois, j'ai expliqué que l'anglais pouvait mettre un élément en relief grâce à des moyens prosodiques uniquement. Je pense qu'une forme de calque de la prosodie de l'anglais sur le français est en train de se produire, j'y consacrerai le troisième point de cette sous-section.

4.1.1. Accents secondaires

Une brèche a été ouverte au chapitre 2 s'agissant du degré de proéminence des accents secondaires.

4.1.1.1. Hypothèses

J'ai distingué deux niveaux de proéminence, l'un sur le mot à l'échelle de l'UI (figure 22) et l'autre sur la syllabe à l'échelle du mot (figure 23). J'explique qu'à l'échelle de l'UI, certains mots sont plus saillants que d'autres : il s'agit des mots porteurs d'accents de phrase (*accent*). Dans ces mots accentués (*accented words*), l'accent primaire reçoit plusieurs degrés de proéminence, selon qu'il est simple accent de phrase, tonique de tête, ou noyau. Dans les mots non accentués au niveau de la phrase, l'accent primaire est à un niveau encore inférieur au simple accent de phrase. Qu'en est-il de l'accent secondaire de ces mots-là ? J'ai émis l'hypothèse en 2.2.5 que l'accent secondaire dans un mot accentué au niveau de la phrase était sur le même plan qu'un accent de mot primaire non accentué au

niveau de la phrase et qu'un accent secondaire dans un mot non accentué au niveau de la phrase était moins proéminent et se trouvait donc à un niveau inférieur à *unaccented stressed syllable.*

Cela est très difficile à entendre, et c'est la raison pour laquelle il est si difficile de mettre en relation les deux niveaux de proéminence. Les proéminences sont relatives. Au vu de ce que j'ai conclu sur les corrélats acoustiques des proéminences, à savoir qu'une combinaison de paramètres souvent variables entrait en jeu dans la perception des proéminences, il semble également difficile de procéder à des mesures. Des tendances ayant quand-même été dégagées, on pourrait envisager de faire des mesures de durée et d'intensité sur les syllabes (voir tableau 3). Il serait aussi intéressant de mesurer le tilt spectral des voyelles, à la suite de Sluijter (1995) et Sluijter & van Heuven (1993, 1996), Campbell (1995) et Campbell & Beckman (1997) (*cf.* 2.2.5) :

> for dialogue speech, spectral information can be very helpful in discriminating prominences. (Campbell, 1995 : 678)

4.1.1.2. Protocole

Il faudrait mettre en place un protocole expérimental et une expérience en laboratoire. Il serait en effet impossible pour cette étude de travailler sur de la parole non contrôlée pour deux raisons.

La première est que l'anglais est plus riche en mots courts qu'en mots longs, et qu'il n'est donc pas facile de trouver des mots porteurs d'accent secondaire dans le discours, surtout spontané non contrôlé. Un rapide coup d'œil sur le COEC confirme cette tendance. On en trouve davantage dans Aix-Marsec, qui est du spontané semi-contrôlé et une pré-étude pourrait éventuellement utiliser ce corpus pour voir si les mesures d'intensité et de durée semblent pertinentes.

La deuxième raison vient du fait qu'il est impossible de faire des mesures précises en parole spontanée à cause des chevauchements et aussi parce qu'il est très délicat de mesurer l'intensité :

> [...] intensity is too susceptible to noise. If the speaker accidentally turns his head or passes a hand before his mouth, intensity drops of greater magnitude than those caused by the difference between stressed and unstressed syllables

> will easily occur. For this reason, manipulating intensity in stress perception experiments seems ill-advised. (Sluijter & van Heuven, 1996 : 2372).

Ces propos sont tout de même à nuancer puisque plusieurs études ont prouvé l'importance de l'intensité, même en discours spontané (Kochanski *et al.* 2005, Herment à paraître b). Le tilt spectral est par contre certainement très difficile à mesurer et mieux vaut-il donc sûrement équiper les locuteurs de micro casques et les enregistrer en chambre sourde.

Le protocole pourrait consister en une lecture de petits textes (fabriqués ou recopiés) dans lesquels on s'assurerait qu'il y ait des mots longs contenant des accents secondaires et que ces mots possèdent différents degrés d'accentuation au niveau de la phrase. Des mesures de durée, intensité et tilt spectral des voyelles dans des syllabes présentant différents degrés de proéminence permettraient de classer les accents secondaires dans la hiérarchie des proéminences et de confirmer ou infirmer les hypothèses émises.

4.1.2. Toniques de tête

J'ai expliqué en 2.2.4 que les toniques de tête avaient souvent une importance particulière au niveau de l'UI. Elles tombent sur des mots importants dans l'énoncé : ce sont soit des mots à contenu sémantique important, informatifs au moment précis de l'énonciation, soit des mots à contenu vague ou des mots grammaticaux comme des pronoms que l'on veut contraster sans pour autant les placer en position de noyau. L'exemple fs 8 illustre bien cela avec *men* qui est tonique de tête et s'oppose à *women* plus haut dans le discours.

4.1.2.1. Niveau de l'UI

J'ai donné comme corrélats acoustiques la hauteur mélodique, la durée et l'intensité. Ces corrélats pourraient être vérifiés sur les données de l'expérience décrite juste avant ou pourquoi pas sur le COEC, qui est d'excellente qualité acoustique mais présente l'inconvénient des nombreux chevauchements et rires dus au nombre de locuteurs.

C'est aussi en parole spontanée qu'il serait intéressant de tester l'hypothèse de l'informativité du premier accent d'une UI. Cela pourrait se faire sur le COEC également, avec l'inconvénient qu'il s'agit d'un petit corpus. Selon Couper-Kuhlen (1986), chaque locuteur a un niveau tonal propre sur lequel il préfère débuter les unités intonatives. Ceci est valable pour les syllabes inaccentuées de début d'UI (les anacruses), mais aussi pour le premier accent. Brazil *et al.* (1980) distinguent trois niveaux tonals : *high*, *mid* et *low* pour les toniques de tête, contrairement à Couper-Kuhlen (2001 : 18), qui suggère que deux niveaux *high* vs. *non high* seraient suffisants. L'analyse du discours de plusieurs locuteurs permettrait de vérifier les hypothèses énoncées plus haut :

- les variétés inter-locuteurs : chaque locuteur a-t-il bien un niveau tonal propre pour débuter les énoncés ?
- les variation intra-locuteurs : deux niveaux suffisent-ils pour décrire les débuts d'UI, ou bien en faut-il plutôt trois, ou davantage ?

4.1.2.2. Au delà de l'UI

L'importance de la tonique de tête semble encore plus grande au niveau du paraton (*paratone*) (Yule 1980), qui est défini comme une unité prosodique qui s'étend sur un thème (*topic*) ou sous-thème (*sub-topic*) du discours d'une façon en quelque sorte équivalente au paragraphe à l'écrit. Déjà Lehiste (1975) avait montré qu'un des traits prosodiques récurrents du « paragraphe conceptuel » (*conceptual paragraph*) à l'oral était le réajustement extra aigu (*extra high reset*) qui en caractérisait le début. Brazil *et al.* (1980), Brown *et al.* (1980) et Yule (1980) ont tous fait des observations similaires sur la parole conversationnelle.

Les travaux menés par Chafe (1988), Schuetze-Coburn *et al.* (1991) et Du Bois *et al.* (1993) ont continué à s'intéresser à la structure intonative au-délà de l'UI. En particulier l'étude de t'Hart *et al.* (1990) sur l'effet de déclinaison suggère que les unités de déclinaison (DU, *declination units*) structurent le discours. Elles sont plus larges que les UI. Quand plusieurs UI forment une DU, la position d'une UI à l'intérieur de l'unité plus large est détectable grâce à la hauteur mélodique de

fin d'UI, mais aussi et surtout grâce à celle de début d'UI. Cette propriété est désormais bien établie et le terme *topic-reset* est maintenant largement employé.

Ainsi Brown *et al.* (1980 : 71) distinguent « major paratones, which follow topic pauses and are characterised by very high peaks […] and minor paratones which follow contour pauses and begin with lower peaks ». Cette affirmation est confirmée par Wichmann (2000 : 42) : elle a réalisé des mesures pour un fichier du corpus SEC[48] (Knowles, Williams and Taylor 1996) : « sentences beginning a new topic (news item) start at a pitch in the region of 240Hz, while the first accented syllable in other sentences occurs mainly around 140-150 Hz ». Ses observations sur ce locuteur l'amènent à aller plus loin et à proposer 4 niveaux de tête :

- Low (autour de 135 Hz)
- Default (autour de 150 Hz)
- Raised (autour de 180 Hz)
- Extra high (autour de 240Hz).

Le peu de données empêche de faire des statistiques mais

> the data suggest at least that topic initial onsets belong to a category of their own. The 'high' pitch used to signal topic shifts is high both syntagmatically (in relation to the sequence in which it occurs) and paradigmatically (in relation to the set of choices available to the speaker at any one time). (Wichmann 2000 : 43).

Couper-Kuhlen (2001) a aussi réalisé une étude à petite échelle en s'intéressant, quant à elle, à l'interaction. Elle affirme que les toniques de tête jouent un rôle à deux niveaux dans la conversation : chez un même locuteur à l'intérieur de son tour de parole, et entre deux locuteurs au niveau du changement des tours de parole. Il s'agit seulement d'une étude pilote, mais dont elle peut tirer les conclusions suivantes :

> participants use high onset and its absence as a strategic resource for cueing new topics. This does not mean that on other occasions high onset or its

48. *Spoken English Corpus*. Ce corpus est ensuite devenu le MARSEC sous la houlette de Peter Roach, puis Aix-Marsec au Laboratoire Parole et Langage (voir note de bas de page n°2).

> absence might not signal something different. The inferencing which the deployment of onset level cues must be expected to be sensitive to the sequential location of the verbal content of the TCU [turn constructional unit] in question. (Couper-Kuhlen 2001 : 21)

Ces conclusions me semblent très prometteuses dans le sens qu'il reste encore beaucoup à faire sur l'analyse des débuts d'unités intonatives, et au delà, des débuts d'unités plus larges que l'UI, que ce fût le paraton, l'unité de déclinaison ou le tour de parole, en parole lue, spontanée contrôlée ou non contrôlée : « It is the way units begin which forms one of the new territories for exploration beyond the intonation phrase » (Couper-Kuhlen 2001 : 17).

4.1.3. Saillance comparée du français et de l'anglais

La troisième thématique que l'on pourrait développer en lien avec l'axe « proéminences prosodiques » est celui d'une approche comparée entre l'anglais et le français. J'ai rappelé en 3.1.1 que l'anglais possède cette capacité de mettre en relief un mot dans un énoncé grâce à des procédés uniquement prosodiques, sans user de la syntaxe. C'est ce que j'appelle la focalisation prosodique. Le français n'est pas censé avoir cette possibilité. Ainsi pour traduire

I went to cinema

dont le noyau serait sur *I,* on emploiera un clivée en français :

C'est moi qui suis allée au cinéma

On peut toutefois arguer que l'énoncé

Je suis allée au cinéma

est tout à fait possible en français dans un contexte contrastif et plutôt de polémique, comme par exemple :

Non je parle | et tu te tais.

Cela irait à l'encontre de ce que Cruttenden (2008 : 328) affirme lorsqu'il dit que les langues romanes ne déplacent pas le noyau vers la gauche. Il est vrai que ce type d'intonation est probablement davantage employé dans un registre familier.

Voici un autre exemple de mise en relief d'un pronom : une petite fille va à l'école avec sa mère. Cette dernière veut se garer et accompagner sa fille mais la petite fille lui dit :

Non, dépose moi devant parce que l'autre fois on était en retard. Enfin, j'étais en retard.

L'anglais aurait mis en relief le pronom *I*. C'est exactement pareil en français, à cela près que le *je* étant tronqué à cause de la voyelle qui le suit, l'accent s'est déplacé sur la première syllabe du verbe *étais*. Mais c'est bien le pronom personnel que la petite fille a voulu mettre en relief, par opposition à *nous*. Ce déplacement de l'accent peut aussi avoir lieu sur le pronom réfléchi en discours spontané toujours, comme dans l'exemple suivant :

Ca y est on s'est lavées. Enfin, j'me suis lavée.

Au lieu de prononcer « je me suis lavée » la locutrice garde la forme tronquée de *je* typique de l'oral et déplace l'accent sur le pronom réfléchi.

Ainsi la focalisation prosodique d'un pronom personnel en français est possible. Il semble aussi possible de faire porter une focalisation prosodique sur un verbe, un peu comme en anglais. La conversation de tous les jours en est truffée :

- *On dirait qu'elles sont fermées.*
- *ah mais, elles sont fermées.*

La personne aurait pu répondre « ah mais elles le sont ». Cet exemple me rappelle étrangement l'anglais « they are closed ».

Ce genre de prononciation me semble calqué sur l'anglais et je me demande s'il ne s'agit pas d'un changement linguistique en cours. Il serait intéressant d'écouter le Corpus d'Orléans (Bergounioux 1996) par exemple, pour une perspective diachronique, ou bien d'interroger des personnes d'âges divers. On pourrait penser que ces énoncés ne sont produits que dans des contextes familiers, où la parole n'est pas très formelle. Mais j'ai entendu exactement le même genre de focalisation prosodique chez des collègues anglicistes lors de colloques, loin d'un contexte familier donc. Serait-ce une influence de l'anglais chez ce type de

locuteurs ? J'ai entendu la même chose dans des films anglais ou américains doublés. S'agit-il de traductions littérales, qui finiraient par nous influencer ? Il serait intéressant de voir si cela apparaît dans tous les styles, chez tous les locuteurs. Ces exemples étant produits uniquement en spontané complètement non contrôlé, il est très difficile de les analyser. Ce type de propriété n'est pas immédiatement interrogeable dans les corpus oraux spontanés, type CID, faute d'annotation multi-niveaux poussées, ce qui invite à considérer d'autre stratégies. On pourrait mettre en place un test de perception : interroger des locuteurs unilingues, bilingues et anglicistes non bilingues et voir s'ils acceptent ce genre d'énoncés ou non.

Le prochain exemple est différent et très intéressant : un garçon part au collège début septembre alors qu'il fait encore chaud avec un maillot de bain qui ressemble à un short. Sa mère lui dit

on dirait que tu vas à la plage.

et le garçon répond avec un sourire espiègle :

je vais à la plage

L'humour est pour beaucoup ici je pense. Il aurait pu répondre « ben justement, j'y vais à la plage » ou « ben justement c'est là que je vais ». La façon très dénuée de répondre sans ajouter « mais » par exemple ou un marqueur de discours comme « ben » mais uniquement une marque prosodique me fait à nouveau penser à l'anglais. Si on voulait traduire en anglais, on dirait quelque chose comme « I am off to the beach ».

Le dernier exemple que je souhaiterais donner est à nouveau tiré de conversations informelles entendues dans la vie de tous les jours et qui n'ont donc pas été enregistrées (mais soigneusement notées). La scène se passe pendant le repas. Le papa fait réviser les tables de multiplication à sa fille :

- 6 x 8 ?

- Attends, je mâche.

- Ça t'empêche pas d'réfléchir.

- Je réfléchis.

A nouveau, une traduction en anglais ferait ressortir un accent sur l'auxiliaire (*I am thinking*) et non sur le pronom. Le français n'utilisant pas d'auxiliaire pour cette forme de présent, la mise en relief a lieu sur le pronom, et non sur le verbe lui-même. On n'aurait pas pu avoir :

**je réflechis.*

Ces questions de proéminence prosodique en français et en anglais m'interpellent et mériteraient qu'on y prête une attention particulière, tout spécialement en lien avec la traduction, sur laquelle je vais m'arrêter dans la section suivante.

J'ai détaché trois grandes thématiques de l'axe proéminences prosodiques mais beaucoup de domaines restent encore à explorer. Ces pistes de recherche invitent à des collaborations avec des spécialistes d'autres disciplines : spécialistes de l'analyse de discours, pragmaticiens, francisants, diachroniciens, traducteurs et traductologues.

4.2. Prosodie et syntaxe

Le deuxième grand axe qui me préoccupe est l'interface syntaxe/prosodie (voir chapitre 3). J'ai travaillé principalement sur le discours spontané et sur l'interface syntaxe/prosodie en discours, un domaine où de nombreux prolongements sont possibles. Ceci constituera mon premier point. Certains énoncés peuvent être ambigus syntaxiquement et la prosodie est souvent le seul moyen de les départir (Hirst 1977). Le deuxième point de cette sous-section traitera des homotaxies. Enfin, la traduction de certains énoncés associant une structure syntaxique et une prosodie particulière fera l'objet d'une troisième piste de recherche possible pour cet axe.

4.2.1. Prosodie et syntaxe en discours

Nous avons étudié l'interface prosodie/syntaxe sous l'angle de certaines structures syntaxiques que nous avons appelées non canoniques. Nous avons plus

particulièrement regardé les clivées, les extrapositions et les dislocations à droite de groupe nominal. Nos résultats montrent que des fonctions pragmatiques de l'intonation peuvent être dégagées en regard de l'organisation de la structure informationnelle, de l'organisation du discours, du contraste et de la focalisation (Herment & Leonarduzzi 2012), et nous avons défini des fonctions particulières aux 3T (*tonality, tonicity* et *tones*) de l'intonation (Leonarduzzi & Herment 2013). Il serait intéressant de prolonger ces travaux avec l'étude d'autres structures syntaxiques. Tout comme nous avons comparé les extrapositions et les dislocations (Leonarduzzi & Herment 2013b), il serait intéressant de faire le même type d'étude avec les dislocations à gauche et les antépositions, qui possèdent des caractéristiques communes et sont décrites comme des structures en distribution complémentaire par certains auteurs (Prince 1997, Rodman 1997, Van Riemsdjik 1997) :

John I like : antéposition

John I like him : dislocation à gauche

Leonarduzzi (2008) affirme que « la DG [dislocation à gauche] renvoie le plus souvent à une entité nouvelle alors que l'antéposition reprend la plupart du temps une entité discursivement ancienne ». Une différence prosodique devrait donc être trouvée, mettant à l'épreuve la notion de *deaccenting* (Ladd 1980). Il serait intéressant de voir dans quelle mesure la prosodie peut remettre en question certaines affirmations fondées uniquement sur l'écrit.

Les *question-tags* pourraient aussi faire l'objet d'études dans le cadre de l'interface syntaxe/prosodie dans le discours. On pourrait comparer les catégories énonciatives décrites par Guillaume (2006) à l'étude prosodique de Dehé & Braun (2013). La prosodie pour ces structures serait sans aucun doute révélatrice de l'attitude que le locuteur veut exprimer.

La prosodie des exclamatives est un autre thème que je souhaiterais aborder. Plusieurs études se sont focalisées sur ces structures (voir Kerfelec 2009 pour une synthèse), mais très peu traitent de l'oral. Il s'agit d'un sujet complexe : que considère-t-on une exclamation ? Sur quels critères ? Il serait tentant de parler de

critères prosodiques. Or un rapide coup d'œil aux exclamatives directes (en *what* et *how*) dans le corpus ICE-GB, avec comme seul critère la syntaxe, montre deux choses. Tout d'abord il y en a très peu, même en discours spontané :

- 12 exclamatives en *how* (avec des adjectifs comme *clever, lovely, easy, different*),
- et 20 exclamatives en *what* (suivi entre autres de *shame, surprise, nuisance, mess*).

Ensuite il est frappant à l'écoute de constater que ces énoncés ne sont pas tous prononcés avec une prosodie emphatique. Certains le sont :

How lovely *fs 60 &* fs *61*

What a mess fs 62

d'autres beaucoup moins (on entend quand-même un allongement important de *shame* dans l'énoncé ci-dessous mais le mouvement mélodique n'est pas ample) :

What a shame fs 63

ou même pas du tout :

How very democratic you are fs 64

Ce dernier exemple est très intéressant d'un point de vue sémantique et discursif. L'association des trois mots *how, very* et *democratic* n'est déjà pas neutre. La locutrice ne ressent donc peut-être pas le besoin d'utiliser une prosodie très marquée.

Les exclamatives en *so* sont dignes d'intérêt également. Contrairement à ce que l'on pourrait penser, l'adverbe *so* n'est en général pas accentué. C'est l'adjectif qui le suit qui l'est. *So* peut même posséder une forme réduite dans certains cas et être prononcé /sə/ et non /səʊ/. L'exemple qui suit comprend deux occurrences de *so*. La première, suivie d'un adjectif, est accentuée en noyau avec *so* prononcé /səʊ/ et assez long (150 ms.) ; la deuxième, suivie d'un adverbe, est inaccentuée, courte (90 ms.) et prononcée /sə/ :

The divisions are so fundamental FR | and they're arguing so badly over Europe FR | fs 65

Là encore, une étude poussée serait certainement révélatrice de facteurs intéressants et l'analyse des trois types d'exclamatives directes (en *how*, *what* et avec *so*) fournirait probablement des pistes alléchantes quant aux exclamatives indirectes, sur lesquelles je vais revenir dans la sous-section suivante.

Il y a sans aucun doute matière à continuer à creuser l'interface syntaxe/prosodie au travers de toutes ces structures et même à aller plus loin en débordant sur la sémantique et la pragmatique pour aborder pleinement le discours, à l'instar de Dehé (2009) et Dehé & Wichmann (2010) et traiter de l'interface prosodie/syntaxe/discours.

4.2.2. Ambiguïté des structures

L'ambiguïté des structures est un des points que j'ai abordés dans le chapitre 3 et sur lequel nous avons plus particulièrement travaillé en analysant les propositions en *–ing* qui se trouvent à la limite entre extrapositions et dislocations à droite (voir 3.2.2.3) :

It's wonderful meeting you.

Dans la même lignée, une analyse prosodique comparée des interrogatives indirectes et des exclamatives indirectes pourrait à mon avis donner des résultats éclairants. Grimshaw (1977) et Elliott (1971) considèrent que des exemples comme

It's amazing how long he can stay under water

sont des exclamatives indirectes (*embedded exclamations*), alors des énoncés tels que

I wonder how long he can stay under water

seraient des interrogatives indirectes. A la suite de Ohlander (1985) et Huddleston (1993), Leonarduzzi (2004) soutient que les propositions en *wh-* introduites par des adjectifs non neutres sémantiquement (voir 2.2.3) tels que *suprising, incredible, amazing* fonctionnent syntaxiquement comme des interrogatives indirectes. Qu'en est-il d'un point de vue prosodique ? Il serait intéressant d'étudier la prosodie des subordonnées en *wh-* afin de voir si elle s'accorde avec le sémantisme exclamatif

(la notion de haut degré, voir Culioli 1974, Rivara 1979) ou la syntaxe. Dans le premier cas, on peut faire l'hypothèse que la prosodie serait plutôt emphatique : deux UI et des tons non neutres, par exemple :

It's amazing HF | how long he can stay under water HF |

ou encore une seule UI avec un ton non neutre toujours et la tonique de tête sur *amazing*, probablement en ton statique très haut ou très bas :

It's amazing how long he can stay under water HF |

Dans le cas où la prosodie s'accorderait avec la syntaxe, elle serait neutre :

It's amazing how long he can stay under water F |

Il faudra cependant garder à l'esprit que toutes les exclamatives directes ne sont pas produites avec une prosodie emphatique. Les résultats d'une telle étude pourraient alors être corrélés ou comparés à ceux de l'analyse de structures encore plus ambiguës car ne comportant pas d'adjectifs sémantiquement non neutres comme *amazing* ou *surprising*, telles que

I know how tall John is

Cet énoncé peut être soit une interrogative indirecte, soit une exclamative indirecte. Les deux interprétations possibles dépendent de la présupposition :

- dans le cas d'une interrogative indirecte, la présupposition est *John has a certain height.* La phrase signifie « je connais la taille de John » (je sais la taille qu'il fait).
- dans le cas d'une exclamative indirecte, le présupposé est *John is tall* (voire *very tall*). La phrase signifie « je sais combien John est grand ».

Est-ce que seul le contexte permet de désambiguïser ou bien va-t-on systématiquement trouver une prosodie différente ?

L'ambiguïté dans tous ses états est un sujet porteur, qui peut rassembler plusieurs disciplines et favoriser des collaborations, pour preuve le projet porté par l'Université de Tübingen, financé par la « German Research Foundation » à hauteur de trois millions d'euros pour quatre années et demi, intitulé « Ambiguity

– Production and Perception[49] », qui regroupera dès octobre 2013 des chercheurs et des doctorants de disciplines aussi diverses que Linguistique, Critique Littéraire, Rhétorique, Droit, Théologie, Psychologie, Etude des Médias et Philosophie/Ethique.

4.2.3. Questions de traduction

Lorsque l'on travaille sur une langue étrangère, on a souvent tendance à chercher à traduire pour tenter de clarifier ses propos, ou à paraphraser les énoncés que l'on étudie. Cette « manie », si tant est que cela en fût une, m'a souvent conduite à me poser des questions intéressantes. Comment traduire l'emphase prosodique par exemple ? J'ai expliqué en 4.1.3 que le français pouvait faire usage de focalisation prosodique à l'oral. A l'écrit, l'emploi d'une structure syntaxique sera souvent nécessaire, des exemples classiques ont déjà été commentés plus haut. Ce sont des cas moins répandus qui attirent mon attention.

4.2.3.1. Do inséré dans un contexte affirmatif

J'ai mentionné l'auxiliaire *do* inséré dans un contexte affirmatif à plusieurs reprises dans ce document de synthèse. Il peut recevoir trois degrés d'accentuation différents selon le contexte dans lequel on le trouve et les traductions vont être toutes différentes. J'en propose quelques unes ci-dessous :

- noyau (do nie p') :

But if you do have spare money to invest fs 66

« Mais si **au contraire** vous avez de l'argent disponible à investir »

I'm beginning to wonder what I do want fs 67

« Je commence à me demander ce que je veux **vraiment** ».

- tonique de tête (do renforce la relation S/P) :

they did have trains *fs 17*

49. http://www.uni-tuebingen.de/en/research/forschungsschwerpunkte/research-training-groups/grk-1808-ambiguity-production-and-perception.html

« Ils avaient **bien** des trains, mais ils pouvaient venir comme ils pouvaient ne pas venir, avoir un jour de retard… »

I say almost because she did lose fs 46

« je dis presque parce qu'elle a **tout de même** perdu 3 fois en 1986 »

- inaccentué au niveau de la phrase (*unaccented*) :

I really do think fs 18

« Je pense vraiment qu'une organisation comme VSO … ». Dans cet exemple, il est difficile de traduire *do* car il porte sur l'adverbe et vient en renforcer le sémantisme (*cf.* Herment 2010 et 2.2.4). Je pense qu'une emphase prosodique sur « vraiment » rendrait l'insertion du *do.*

We certainly did have dire matches in Mexico fs 47

« **Certes**, nous avons vu des matchs médiocres […], mais nous avons aussi assisté à d'excellentes rencontres… ».

4.2.3.2. The *prononcé /ði/ devant consonne*

La traduction de l'article indéfini *the* prononcé /ði/ devant consonne (Féron 2013) serait probablement aussi un thème porteur pour les traducteurs. Dans son mémoire de Master 2, A. Féron a dégagé en plus de l'emphase des fonctions pragmatiques liées à cette prononciation particulière de l'article, tout comme elle l'avait fait pour son mémoire de M1 présenté au colloque PAC 2013 (Féron 2012) à propos de l'article défini *a* prononcé /eɪ/ devant consonne. Comme pour le *do,* plusieurs degrés de proéminence sont décelables et la traduction ne sera pas la même en fonction de ces degrés.

Je ne donne qu'un seul exemple de traduction possible, pour nourrir la réflexion et montrer qu'il y a beaucoup à étudier dans le domaine que l'on pourrait appeler interface prosodie/traduction :

Jeannie? The /ði/ [pause] head of East Coast operations, Jeannie? fs 68

« tu fais allusion à **la (fameuse)** Jeannie qui est à la tête des opérations de la Côte Est ? ».

La diversité des moyens de traduire *do* ou les articles ayant une prononciation pleine représente très certainement une manne pour les traducteurs et traductologues, avec qui les prosodistes gagneraient à s'associer.

4.3. Interphonologies

Mon implication dans le projet PAC est le signe de mon intérêt pour l'étude des variétés d'anglais et plus largement pour l'étude des langues en contact et des interphonologies. Je participe en effet à l'édition d'un volume chez Springer Verlag (Delais-Roussarie *et al.* à paraître), intitulé *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*. La table des matières provisoire donnée en annexe 6 reflète la diversité des études qui peuvent être menées sur ce thème ainsi que les deux grands volets que je propose de développer ci-après, en lien plus étroit avec l'anglais et la prosodie : les interphonologies dialectales et les interphonologies L1-L2. Les deux domaines sont liés. Les variétés dialectales reflètent les traces d'un substrat. Les langues secondes (au sens de ESL, *English as a Second Language*) telles que l'anglais de Jamaïque ou de certains pays d'Afrique par exemple, et l'anglais parlé par des apprenants (EFL, *English as a Foreign Language*) sont marqués par deux systèmes phonologiques en interaction. Comment les systèmes phonologiques des deux langues s'influencent-ils ? Certains éléments linguistiques sont-ils plus faciles à acquérir que d'autres ? Certaines caractéristiques sont-elles plus difficiles à faire disparaître que d'autres ? La comparaison entre les caractéristiques des ces différents parlers devrait permettre de proposer de nouvelles perspectives sur ces questions.

4.3.1. Interphonologies dialectales

J'ai commencé à travailler sur l'anglais du Nord de l'Irlande avec Gabor Turcsan (Turcsan & Herment à paraître) et nous montrons que la notion d'interphonologie est présente au niveau segmental mais aussi au niveau suprasegmental des contours mélodiques de par l'influence du substrat gaélique

(voir Dalton & Ní Chasaide 1997). Calabrese (2011) avait fait des constatations similaires pour l'anglais parlé au Pays de Galles. Les intonations terminales montantes typiques des grandes villes urbaines du Nord de la Grande Bretagne (UNBI, *Urban Northern British Intonation*) (*cf.* Grabe 2004 par exemple et Wales 2006) pourraient venir d'une intonation celtique (Wilhelm 2011) et plus précisément de l'intonation nordique des Vikings qui ont envahi le Nord des îles britanniques (Hirst 2008, 2009). Des prolongements à ces hypothèses pourraient être envisagés, en lien avec des diachroniciens et pourquoi pas des historiens.

Les intonations montantes terminales UNBI sont les tons les plus fréquemment utilisés pour les contours terminaux dans les variétés parlées à Belfast (Jarman & Cruttenden 1976), Manchester (Cruttenden 2001) ou encore Glasgow (Cruttenden 2007). Elles se distinguent de ce que l'on appelle HRT (*high rising terminal*) ou encore *upspeak*, c'est-à-dire les intonations montantes non systématiques typiques de certaines variétés telles que l'anglais américain (Ching 1982), l'australien (Guy *et al.* 1986), le néo-zélandais (Britain 1992, 1998), et qui s'étend également à la Grande-Bretagne (*cf.* Bradford, 1997 ; Cruttenden, 1994 & 1997 ; Ladd, 2008). A l'instar de Barry (2008), qui a comparé les intonations HRT du Sud de la Californie et du Sud de l'Angleterre, on pourrait tenter de comparer les contours mélodiques entre d'autres variétés.

Très peu d'études ont comparé UNBI et HRT (voir le *Language Log* de Liberman et sa discussion avec S. Wilhelm[50]). Il y aurait là des analyses très prometteuses à entreprendre. Le contour est-il différent ? Les queues d'énoncés sont-elles similaires ? Qu'en est-il des débuts d'énoncés ?

De la même façon on pourrait comparer les intonations HRT des assertions et les intonations des questions en montée. Les contours pré-nucléaires et la hauteur des toniques de tête sont à mon avis différentes. J'ai comparé le contour d'une question en montée et le contour d'une assertion en montée, chez deux locutrices canadiennes. Les énoncés sont tirés du corpus d'anglais canadien de

50. Liberman, M. "Uptalk vs. UNBI again". *Language Log*, 23/11/2008 <*http://languagelog.ldc.upenn.edu/nll/?p=863*>

Rodrigues Da Mota (2010), qui est du spontané non contrôlé : les figures ci-dessous montrent les contours intonatifs. Les deux premiers énoncés sont produits par la même locutrice et les deux derniers par une autre locutrice. Le premier énoncé de chaque paire est l'assertion en montée (figures 34 et 36), le deuxième énoncé de chaque paire (figures 35 et 37) est la question. Le noyau est souligné dans les exemples et il est sélectionné en bleu sur les figures 36 et 37.

And then we went to Scat. fs 69

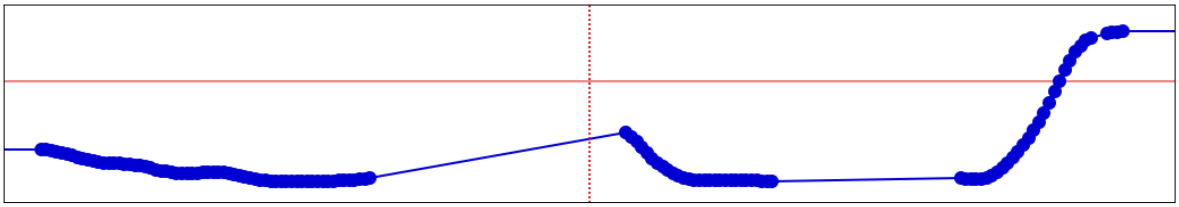


Figure 34 : courbe de F0 pour fs 69 : assertion en montée (HRT)

Have you been here since September? fs 70

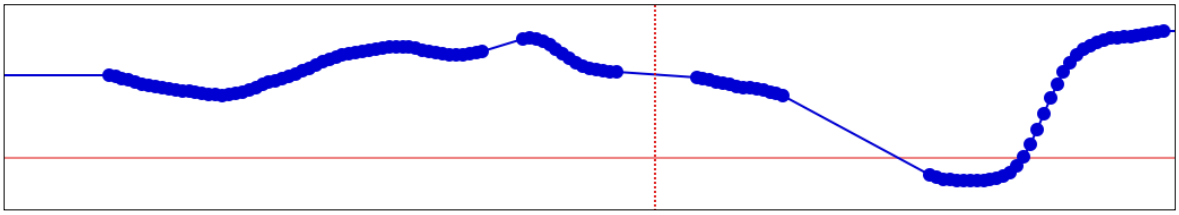


Figure 35 : courbe de F0 pour fs 70 : question ascendante

And they wouldn't let me into class first semester fs 71

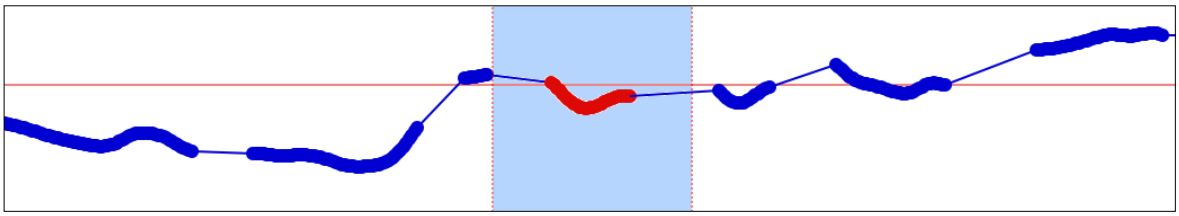


Figure 36 : courbe de F0 pour fs 71 : assertion en montée (HRT)

Do you have any plans for that yet ? fs 72

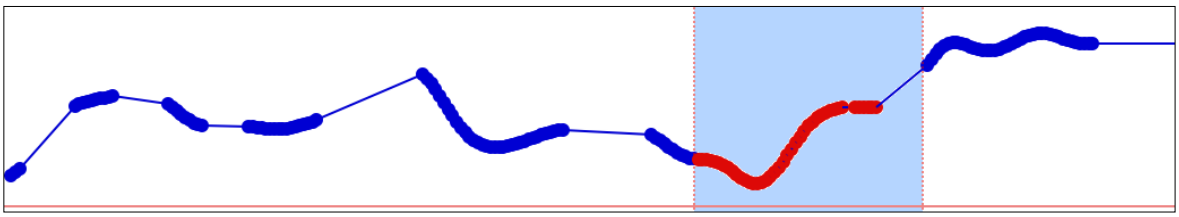


Figure 37 : courbe de F0 pour fs 72 : question ascendante

Il ne s'agit ici nullement d'une étude mais la simple visualisation des courbes de F0 est encourageante : dans les deux assertions montantes, l'UI part de beaucoup de plus bas que dans les questions et le contour prénucléaire est beaucoup plus plat. Une étude plus poussée mérite d'être envisagée. On pourrait comparer des contours prénucléaires comportant le même nombre de syllabes, mesurer la hauteur de la tonique de tête et des syllabes qui se trouvent avant la tonique de tête. Des mesures de tempo pourraient être réalisées, etc. Les queues d'énoncé sont de longueur variable dans les quatre énoncés ci-dessus. De la même façon que pour les syllabes prénucléaires, l'étude des syllabes post-nucléaires se montrerait peut-être révélatrice. Enfin, on pourrait regarder comment le ton nucléaire est réalisé dans les deux contours montants : dans les deux questions ci-dessus (fs 70 et fs 72), les noyaux sont prononcés sur un ton statique grave alors que dans fs 71 le noyau est en statique beaucoup plus haut. Dans fs 69 le noyau est en montée car il constitue la dernière syllabe et il part de bas, comme dans les questions.

On voit à la simple observation de quatre énoncés que l'on a une manne pour des recherches futures. Les interphonologies dialectales sont une bonne occasion d'interroger le phonologique au sein de la diversité phonétique du dialectal. La confrontation des montées selon les variétés questionne la catégorisation et la sémantique des tons. Da Mota (2010) a commencé à explorer les fonctions pragmatiques de HRT en anglais canadien. Il serait intéressant de reprendre les différentes fonctions qu'elle a définies, de les comparer à d'autres variétés où HRT existe et d'entreprendre une analyse phonétique des contours correspondants. Trouverait-on les mêmes caractéristiques dans tous les dialectes

où l'on entend HRT ? Nous avons comparé HRT au marqueur de discours *eh* typique de l'anglais canadien (Rodrigues Da Mota & Herment 2011) : les deux phénomènes remplissent des fonctions pragmatiques similaires. Or on entend très peu d'occurrences de *eh* dans le corpus de parole spontanée étudié, et beaucoup de cas de HRT. L'hypothèse d'une diffusion de HRT est envisageable. Peut-on parler de la grammaticalisation d'une mode, d'un phonostyle, dans le cas de HRT ? Cela évoque le phénomène de *levelling* (Trudgill 1986) qui aurait lieu au niveau des caractéristiques suprasegmentales. Cette tendance à la réduction, et même à la disparition de certaines caractéristiques dialectales marquées conduit à l'émergence d'accents supra-locaux : *supralocalisation* (voir notamment Britain 2010) ou *regional dialect levelling* (Kerswill 2003). Dans le cas de l'anglais canadien, la disparition du *eh* au profit d'une intonation montante chez les jeunes locuteurs participe-t-elle peut-être du même phénomène mais à une échelle non plus régionale mais internationale. Les caractéristiques segmentales du phénomène de supralocalisation ont été largement décrites (voir Britain 2011 pour une synthèse). Le niveau suprasegmental mériterait que l'on s'y intéresse aussi.

4.3.2. Interphonologies L1-L2

La question des interphonologies et des langues en contact m'intéresse aussi en tant qu'enseignante d'anglais pour des étudiants francophones. La question du transfert de la L1 vers la L2 (voir notamment Rasier & Hiligsmann 2007 ou Gut 2009 pour une synthèse) ainsi que celle de l'application de la notion marque à l'acquisition de la L2 (voir Eckmann 2004 pour un panorama) sont des problématiques actuelles. Avec la constitution du corpus AixOx (voir Herment *et al.* à paraître), de nombreuses perspectives s'ouvrent. En effet, ce corpus regroupe des passages en anglais lus par des natifs anglophones et des apprenants francophones et les passages équivalents en français lus par des natifs français et des apprenants anglophones. Les apprenants sont divisés en 2 catégories : niveau B1/B2 et niveau C1/C2.

On peut faire un tas d'investigations au niveau segmental (deux étudiants ont fait leur mémoire de M1 en se fondant sur AixOx cette année : Boyer (2013) a étudié la prononciation du <th> et Gerboud (2013) a travaillé sur les digraphes <aw>, <ew> et <ow>). Le niveau suprasegmental peut aussi être analysé. Santiago-Vargas & Delais-Roussarie (2012a et 2012b) ont étudié l'intonation des questions chez des mexicains apprenant le français. Ils montrent que certains cas de productions erronées peuvent être attribués au transfert des contours de la L1 vers la L2. Ce transfert est bien souvent à l'origine des productions entendues chez les apprenants. Cependant, toutes les erreurs ne peuvent pas être expliquées ainsi : Santiago-Vargas & Delais-Roussarie proposent une piste qui irait dans le même sens que Trouvain & Gut (2007), à savoir que c'est le processus d'acquisition lui-même qui pourrait être à l'origine de certains des contours produits de façon erronée par les apprenants. Il serait donc intéressant de comparer les questions en anglais L2 et en français L1 ou bien en français L2 et en anglais L1 pour voir si les hypothèses décrites ci-dessus se retrouvent dans AixOx. Une telle étude n'est pas simple à mettre en œuvre car si AixOx fournit des données quasiment similaires dans les deux langues, elles ne le sont pas exactement. Par exemple la question anglaise

Do you take reservations by telephone?

correspond en français à une question sans inversion sujet-verbe, qui déclenchera donc nécessairement un contour ascendant :

Vous prenez les réservations par téléphone ?

De surcroît, la question de l'acceptabilité de tel ou tel contour est délicate :

> [M]ore than one phrasing can be associated with any given sentence. This may unjustly reject acceptable learner's productions, unless the exact imitation of a model speaker is required. (Ballier & Martin à paraître b)

J'ai pris l'exemple des questions mais d'autres structures pourraient être étudiées et AixOx, de par son annotation, représente un corpus intéressant, dont l'investigation peut par exemple servir de base pour émettre des hypothèses à tester ensuite sur de la parole semi-spontanée ou spontanée.

De nombreux domaines restent encore à explorer en prosodie et des liens avec d'autres disciplines plus ou moins proches peuvent être envisagés. Cette problématique pluridisciplinaire à l'œuvre dans les contacts de langue est au cœur du projet ANR CLAPOTY[51] (Contacts de Langues : Analyses Plurifactorielles assistées par Ordinateur et conséquences Typologiques).

51. http://clapoty.vjf.cnrs.fr/

Conclusion

Au-delà d'un bilan de mes recherches, ce document de synthèse constitue une prise de position théorique découlant des travaux que je mène au sein de cette vaste discipline qu'est la prosodie anglaise.

J'ai posé dans le chapitre 1 les bases méthodologiques et théoriques qui sous-tendent mes recherches. Je travaille dans le cadre de la tradition prosodiste britannique et mes travaux s'appuient sur des études de cas à partir d'analyses d'occurrences tirées de corpus oraux. J'utilise des outils de traitement de la parole et d'annotation qui sont aujourd'hui à la disposition de la communauté scientifique, mais j'ai souhaité mettre en avant dans cette synthèse la méthode perceptuelle. L'apparition de grands corpus et d'outils d'annotation automatique est un immense atout qui va permettre d'accélérer les recherches dans le domaine de la prosodie, mais je crois qu'il ne faut pas pour autant abandonner les analyses perceptuelles. J'ai montré l'intérêt des outils, mais ai voulu mettre en garde contre leur utilisation parfois hâtive si l'on ne possède pas une certaine expérience de l'écoute et un certain bagage théorique nécessaire. A l'instar de Wichmann (2000), je crois que loin de s'exclure, les deux méthodes, auditive et instrumentale, se complètent au contraire :

> Both methods of analysis have their advantages and their limitation. There is in my view no question of preferring one to another. Both have a unique contribution to make. (Wichmann 2000 : 13)

J'ai consacré le chapitre 2 de ce document aux proéminences prosodiques : j'ai tenté de montrer pourquoi il me semble plus logique de partir du suprasegmental pour descendre vers le segmental. Les proéminences se trouvent à la base de mon travail de prosodiste et les différentes études que j'ai menées sur cette thématique m'ont conduite à proposer dans le présent document un modèle de hiérarchies des proéminences prosodiques. J'en discute les corrélats acoustiques et prosodiques pour en donner finalement des tendances : en effet, en discours, les réalisations des proéminences ne sont pas toujours celles qui sont attendues.

Le chapitre 3 présente une réflexion sur les travaux que j'ai entrepris sur l'interface syntaxe/prosodie. J'y expose le cheminement qui m'a amenée à m'intéresser à ce domaine, avec comme fil d'Ariane l'emphase et les proéminences. Dans ce chapitre, la méthode perceptuelle est à nouveau mise en exergue puisqu'en raison de sa mauvaise qualité sonore, le corpus sur lequel j'ai travaillé ne permettait pas le recours aux outils d'annotation automatique. Contrairement à la description de l'écrit, la description de l'oral est encore loin de cette phase de caractérisation prescriptive canonique : une structure donnée, une prosodie particulière. La grammaire de l'écrit ne peut donc pas s'appliquer à l'oral. Je tente de montrer en m'appuyant sur les 3T de la tradition britannique qu'une approche discursive de la prosodie permet une réinterprétation des structures syntaxiques. La combinaison attendue entre la prosodie et la structure n'est pas toujours celle qui est réalisée en discours et des réinterprétations sont donc induites. Le découpage en unités intonatives, la réalisation des tons, mais aussi le lieu des proéminences et leur réalisation sont autant d'indices qui permettent de comprendre les subtilités qu'exprime le locuteur. Le chapitre 3 met ainsi en relation ma conviction théorique exposée au chapitre 1, les différentes proéminences décrites au chapitre 2 et une approche interactionnelle de la prosodie.

Le chapitre 4 ouvre ce travail de synthèse et de réflexion sur des projets de recherche future dans le prolongement de mes travaux actuels, mais aussi vers de nouveaux horizons comme le questionnement des invariants prosodiques au-delà

des variations observables en contact de langues. L'analyse de variétés dialectales et de parlers d'apprenants devrait permettre d'offrir des perspectives sur un ensemble de questions relatives à la prosodie et d'en comprendre mieux le fonctionnement.

Au terme de de ce parcours, je souhaiterais nuancer les frontières posées par Couper-Kuhlen (2001) dans le champ de la prosodie en discours. Elle distingue trois approches :

- *Intonation-as-grammar*, une approche qui a tenté d'établir des correspondances entre les différents types de phrases et les contours intonatifs (Pierrehumbert & Hirschberg 1990 ou Couper-Kuhlen & Selting 1996 par exemple).
- *Intonation-and-information-flow*, une approche initiée par Chafe (1979, 1980, 1993) grâce à laquelle le découpage en unités intonatives et le caractère continuatif ou terminal des unités (Du Bois *et al.* 1993) nous permet d'émettre des hypothèses relatives à la conscience du locuteur au moment précis de son élocution. Cette approche s'occupe moins des mouvements mélodiques à proprement parler et s'intéresse davantage à la segmentation en unités et à la continuité entre les unités. Contrairement à la première approche, celle-ci se fonde plus spécifiquement sur l'observation du discours.
- *Prosody-as-contextualization cue* : cette approche délibérément interactionnelle est fondée sur l'idée principale que les signes linguistiques ne sont pleinement interprétables qu'en contexte (Cook-Gumperz & Gumperz 1976). Ce sont des indices de contextualisation (*contextualization cues*) verbaux et non verbaux (Auer & Di Luzio 1992) choisis par les locuteurs, qui rendent les contextes pertinents. L'intonation et au-délà la prosodie avec le rythme et l'intensité, sont autant d'indices.

Selon Couper-Kuhlen, la dernière approche est complémentaire de la deuxième mais s'oppose à la première en ce que celle-ci utilise principalement de la parole de laboratoire.

Je pense quant à moi que les trois approches sont complémentaires. La première constitue le point de départ de la réflexion. On cherche à établir une grammaire de l'oral, entendue comme un répertoire de configurations intonatives prototypiques pour une construction canonique donnée. Les expériences réalisées en laboratoire pour des besoins scientifiques précis ne sont pas à rejeter. Elles permettent au contraire de faire affleurer des régularités réalisationnelles associées à des intentions de sens inférables de manière suffisamment contrainte (*cf.* les « situated phonologies » de Couper-Kuhlen 2007). La deuxième et la troisième approche représentent finalement deux variantes du même principe interactionnel : c'est en discours et en contexte que les réalisations prennent leur sens. A travers la réalisation prosodique d'une structure donnée associée à des mots particuliers comme les mots à sémantisme non neutre par exemple, le locuteur donne des indices quant à sa conscience au moment de l'élocution, mais il permet aussi à son interlocuteur de réinterpréter son message grâce aux indices de contextualisation qu'il lui donne. La variation individuelle, à laquelle l'oreille s'ajuste très rapidement, et la variation dialectale sont à prendre en compte pour la pertinence des contextes.

J'espère avoir prouvé à travers ce document de synthèse que l'on peut donc réconcilier les trois approches et en tirer le meilleur parti. C'est l'interaction en discours entre la configuration prosodique, la structure syntaxique, le choix des mots, la variation individuelle et la variation dialectale qui nous permet d'interpréter le sens de l'énoncé en contexte. L'analyse de cette interaction en corpus de discours, fondée sur des régularités réalisationnelles extraites de corpus de parole, devrait nous permettre de poser les bases de cette renégociation en contexte, située, des configurations attendues.

Bibliographie

Aarts, B. & Nelson, G., 1999, Investigating English around the world: The International Corpus of English, in R. Wheeler (ed.), *The workings of language: from prescriptions to perspectives*, New York : Praeger, 107-115.

Aarts, B., Nelson, G. & Wallis, S., 2002, *Exploring natural language: the British component of the International Corpus of English,* Amsterdam : John Benjamins. http://www.ucl.ac.uk/english-usage/projects/ice-gb/

Abercrombie, D., 1964, Syllable quantity and enclitics in English, in Abercrombie, D., Fry, D.B., McCarthy, P.A.D., Scott, N.C. & Trim, G.L.M. (eds.), 1964, *In Honour of Daniel Jones,* London : Longman, 216-222.

Adamczewski, H. & Delmas, C., 1991, *Grammaire linguistique de l'anglais* (3ème édition), Paris : Armand Colin.

Arvaniti, A. & Baltazani, M., 2000, Greek ToBI: A System For The Annotation Of Greek Speech Corpora, *Proceedings Second International Conference on Language Resources and Evaluation (LREC 2000),* vol. II, 555-562.

Arrivé, M., Gadet, F. & et Galmiche, M., 1986, *La grammaire d'aujourd'hui*, Paris : Flammarion.

Auer, P. & Di Luzio, A. (eds.), 1992, *The contextualization of language,* Amsterdam : Benjamins.

Auran, C., 2004, *Prosodie et anaphore dans le discours en anglais et en français : cohésion et attribution référentielle*, Thèse de doctorat, Laboratoire Parole et Langage, Université de Provence, Aix-en-Provence, France.

Auran, C., 2012, *Approches prosodiques du discours en corpus : méthodes, outils et analyses,* Habilitation à diriger des recherches, Université Paris 7 Diderot.

Auran C., Bouzon C. & Hirst, D., 2004, The Aix-MARSEC Project: An Evolutive Database of Spoken British English, *Proceedings of the 2nd International Conference on Speech Prosody*, Nara, Japon.

Auran C. & Loock R., 2010, Appositive Relative Clauses and their Prosodic Realization in Spoken Discourse: a Corpus Study of Phonetic Aspects in British English, *in* Kühnlein, P., Benz, A. & Sidner, C. L. (eds.), *Constraints in Discourse 2*, Amsterdam, John Benjamins, 163–178.

Auran, C. & Loock, R., 2011, The prosody of discourse functions: the case of Appositive Relative Clauses in spoken British English, *Corpus Linguistics and Linguistic Theory 7(2)*, 181-201.

Ballier, N., 2004, *Praxis métalinguistiques et ontologie des catégories*, Habilitation à Diriger des Recherches, Université de Paris X, Nanterre, France.

Ballier, N. & Martin, P., à paraître (a), Developing corpus interoperability for phonetic investigation of learner corpora, in Díaz-Negrillo, A., Ballier, N. & Thompson, P. (eds.) *Automatic Treatment and Analysis of Learner Corpora*, Amsterdam : Benjamins, Studies in Corpus Linguistics Series.

Ballier, N. & Martin, P., à paraître (b), Pronunciation/Prosody, in Granger, S., Gilquin, G. & Meunier F. (eds.), *The Cambridge Handbook of Learner Corpus Research,* Cambridge : Cambridge University Press.

Barry, A., 2008, The *Form, Function And Distribution Of High Rising Intonation: A Comparative study of HRT in Southern Californian and Southern British English, Saarbrücken : Verlag Dr Müller.*

Baude, O. (coord.), Blanche-Benveniste, C., Calas, M-C., Cordereix, P., Goury, L., Jacobson, M., de Lamberterie, I., Marchello-Nizia, C. & Mandada, L., 2006, *Corpus oraux : guide des bonnes pratiques 2006*, Presses Universitaires d'Orléans, CNRS Editions.

Beckman, M. E. & Ayers, G. M., 1994, Guidelines for ToBI labelling (version 2.0) Electronic document /opt/tobi/TOBI-TRAINING/labelling_guide-V2.ASCII on OUPLSun.

Beckman, M. E., Hirschberg, J. & Shattuck-Hufnagel, S., 2005, The original ToBI system and the evolution of the ToBI framework. In Jun (ed.), *Prosodic*

typology: The phonology of intonation and phrasing. Oxford : Oxford University Press, 9–54.

Beckman, M. & Jun, S-A., 1996, K-ToBI (Korean ToBI) Labelling Convention, Version 2, Manuscript, Ohio State University and UCLA. http://www.linguistics.ucla.edu/people/jun/sunah.htm.

Beckman, M.E. & Pierrehumbert, J.B., 1986, Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255–309.

Bell, B. & Bénard, F. 2013, Digital curation: the SLDR experience, *Proceedings of TRASP (Tools and Ressources for the Analysis of Speech Prosody) 2013*, 30 août 2013, Aix-en-Provence, France.

Bergounioux, G., 1996, Etude Socio-Linguistique sur Orléans (1966-1970), *Revue Française de Linguistique Appliquée,* I-2, 87-88.

http://bach.arts.kuleuven.be/elicop/

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*, vol. 49, no. 3. http://sldr.org/sldr00027/fr

Bigi, B., 2012, SPPAS: a tool for the phonetic segmentation of speech*, Proceedings of the Language Resource and Evaluation Conference*, Istambul, Turquie, 1748-1755.

Bigi, B. & Hirst, D., 2012, SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody, *Proceedings of the 6th International Conference on Speech Prosody,* Shanghai, Chine, 19-22. http://www.lpl-aix.fr/~bigi/sppas

Bigi, B., Péri, P. & Bertrand, R., 2012, Orthographic Transcription : which Enrichment is required for phonetization ? *Proceedings of the Language Resource and Evaluation Conference*, Istambul, Turquie, 1756-1763.

Birner, B.J. & Ward, G., 2004, Information Structure and Non-canonical Syntax, in Horn, L.R. & Ward, G. (eds.), *The Handbook of Pragmatics*, Oxford: Basil Blackwell, 153-174.

Blanche-Benveniste, C. & Jeanjean, C., 1987, *Le français parlé : transcription et éditions,* Paris : Didier Erudition.

Boersma, P. & Weenink, D., 2001, PRAAT, a system for doing phonetics by computer, *Glot International 5*(9/10): 341-345. http://www.praat.org

Bolinger, D., 1958, A theory of pitch accent in English, *Word* 14, 109-149, reprinted in Abe, I. & Kanekiyo, T., (eds.), 1965, *Forms of English : Accent, Morpheme, Order,* Harvard : Harvard University Press.

Bouzon, C., Auran, C. & Hirst, D.J., 2004, Comparative approaches to prosody across varieties of English, *Tribune Internationale des Langues Vivantes 36*, 123-138.

Bradford, B., 1997, 'Upspeak in British English', in *English Today* 51, 13.3 : 33-36, reprinted in Collins, B.S. and *Mees*, I.M. (eds.), *2003, Practical Phonetics and Phonology*, *Londres : Routledge,* 229-234.

Brazil, D., 1997, *The communicative value of intonation in English,* Birmingham Bleak House and ELR. 1985, Republished Cambridge: Cambridge University Press.

Brazil, D. Coulthard, M. & Johns C., 1980, *Discourse Intonation and Language Teaching*, London : Longman.

Britain, D., 1992, Linguistic Change in Intonation: the use of High Rising Terminals in New Zealand English, *Language Variation and Change* 4, 77-104.

Britain, D., 1998, High Rising Terminals in New Zealand English: Who uses them, when and why?, in *Essex Research reports in Linguistics*, 21/09/98, 33-58.

Britain, D., 2010, Supralocal regional dialect levelling , in Llamas, C. & Watt, D. (eds.), *Language and identities*, Edimbourg : Edinburgh University Press, 193-204.

Britain, D., 2011, The heterogeneous homogenisation of dialects in England, *Taal & Tongval* 63 : 43-60.

Brown, G., Currie, K.L., & Kenworthy, J., 1980, *Questions of intonation*, Londres : Croom Helm.

Bruce, G., 1977, *Swedish word accents in sentence perspective,* Lund : Gleerup.

Brunetti, L., Bott, S., Costa, J., Valduví, E., 2011, A multilingual annotated corpus for the study of information structure, in Konopka, M., Kubczak, J., Mair, C., Štícha, F. & Wassner, U. (eds.), *Grammar & Corpora/ Grammatik und Korpora 2009, Third international Conference/ Dritte Internationale Konferenz*, Tübingen : Narr Verlag, 305-326.

Burnard, L. & Aston, G., 1998, *The BNC handbook: exploring the British National Corpus*. Edimbourg : Edinburgh University Press.

Bybee, J., 2011, Markedness : iconicity, economy and frequency, in Song, J.J., *The Oxford handbook of language typology,* Oxford : Oxford University Press, 131-147.

Calabrese, L., 2011, *Caractéristiques prosodiques spécifiques de l'anglais parlé au Pays de Galles : l'influence de la langue Galloise,* Thèse de doctorat,, Laboratoire Parole et Langage, Aix-en-Provence : France.

Campbell, N., 1995, Loudness, spectral tilt, and perceived prominence in dialogues, *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Stockholm 1995), 3 : 676-679.

Campbell, N., & Beckman, M., 1997, Stress, prominence, and spectral tilt, *Proceedings of Intonation: Theory, models and applications*, 18-20 septembre 1997, Athènes : Grèce, 67-70.

Campbell, N. & Venditti, J., 1995, J-ToBI: an intonational labeling system for Japanese, Paper presented at the Autumn meeting of the Acoustical Society of Japan.

Cappeau, P. & Hanote, S. (eds.), 2012, *Focalisations*, Rennes : Presses Universitaires de Rennes.

Catford, J.C., 2001, *A Practical Introduction to Phonetics*, 2ème édition (1ère édition 1988), Oxford : Oxford University Press.

Carr, P., 2013, *English phonetics and phonology, an introduction*, 2ème édition (1ère édition 1999), Chichester : Wiley-Blackwell.

Carr, P., Durand, J. & Pukli, M. (2004). The PAC project: Principles and Methods. *Tribune des Langues Vivantes* 36: 24-35. http://www.projet-pac.net

Chafe, W., 1979, The flow of thought and the flow of language, in Givón, T. (ed.), *Syntax and Semantics* Vol. 12: *Discourse and Syntax*, New York : Academic Press, 159-81.

Chafe, W., 1980, The deployment of consciousness in the production of a narrative, in W. Chafe (ed.), *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*, Norwood, NJ : Ablex, 9–50.

Chafe, W., 1988, Linking intonation units in spoken English, in Haiman J. & Thompson S.A. (eds), *Clause Combining in Grammar and Discourse*, Amsterdam and Philadelphia : Benjamins, 1-27.

Chafe, W., 1993, Prosodic and functional units of language, in Edwards, J. & Lampert (eds.), *Talking Data : Transcription and Coding Methods for Language Research,* Hillsdale, NL : Lawrence Erlbaum.

Ching, M., 1982, The Question Intonation In Assertions, *American Speech* 57, 95-107.

Chomsky, N., 1972, Deep Structure, Surface Structure and Semantic Interpretation, IUCL 1969 reprinted in Chomsky, N., *Studies on Semantics in Generative Grammar*, The Hague: Mouton.

Chomsky, N. & Halle, M., 1969, *The Sound Pattern of English*, New-York : Harper & Row.

Coadou, M., 2007, *Qualité de voix et accents régionaux en anglais britannique*, Thèse de doctorat, Laboratoire Parole et Langage, Université de Provence, Aix-en-Provence, France.

Coleman, J., Liberman, M., Kochanski, G., Burnard, L. & Yuan, J., 2011, Mining a Year of Speech, *VLSP 2011: New Tools and Methods for Very Large Scale Phonetic Research.* University of Pennsylvania. 16-19.

http://www.phon.ox.ac.uk/SpokenBNC

Collins, B. & I. M. Mees, 2013, *Practical phonetics and phonology*, 3ème édition (1ère édition 2003), London and New York : Routledge.

Cook-Gumperz, J. & Gumperz, J.J., 1976, Context in children's speech, *Papers on Language and Context*, Working paper n°46, Language behaviour research laboratory, Berkeley.

Council of Europe, E.C., 2001, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge : Cambridge University Press.

Couper-Kuhlen, E., 1986, *An Introduction to English Prosody*, Tübingen : Niemeyer.

Couper-Kuhlen, E., 2001, Intonation in discourse, in Schiffrin, D., Tannen, D. & Hamilton, H.E., The Handbook of Discourse Analysis, Malden, Oxford : Blackwell, 13-34.

Couper-Kuhlen, E., 2007, Situated phonologies: patterns of phonology in discourse contexts, in Pennington, M. (ed.), *Phonology in context*, Basingstoke : Palgrave, 186-218.

Couper-Kuhlen, E. & Selting, M., 1996, Towards an interactional perspective on prosody and a prosodic perspective on interaction, in Couper-Kuhlen E. & Selting, M. (eds.), *Prosody in Conversation*, Cambridge: Cambridge University Press, 11–56.

Cox, F. M., 2006, Australian English pronunciation into the 21st Century, *Prospect: Australian Journal of TESOL*, *21*, 3-21.

Cruttenden, A., 1994, Rises in English, in Windsor Lewis, J. (ed.), *Studies in General and English Phonetics: Essays in Honour of Professor J. D O'Connor*, Londres : Routledge, 155–173.

Cruttenden, A., 1997, *Intonation,* 2ème édition (1ère édition 1986), Cambridge : Cambridge University Press.

Cruttenden, A., 2006, The de-accenting of given information: A cognitive universal?, in Bernini, G. & Schwartz, M.L. (eds.), *Pragmatic organization of discourse in the languages of Europe,* Berlin: Mouton de Gruyter, 311-356.

Cruttenden, A., 2008, *Gimson's Pronunciation of English*, 7ème édition (1ère édition Gimson 1962), Londres : Hodder Education.

Crystal, D., 1969, *Prosodic systems and intonation in English*, Cambridge : Cambridge University Press.

Crystal, D., 2003, *A Dictionary of Linguistics and Phonetics*, 5ème edition (1ère édition 1980), Oxford : Blackwell Publishing.

Culioli, A., 1974, A propos des énoncés exclamatifs, *Langue Française*, 6-15.

Dahan, D. & Bernard, J-M., 1996, Interspeaker variability in emphatic accent production in French, *Language and Speech* 39/4 : 341-374.

Dahak, A., 2013, It's sort of you know another discourse marker, if you like, *Journée d'études OSLiA*, 5 avril 2013, Paris, France.

Dalton, M. & Ní Chasaide, A., 2007, Nuclear accents in four Irish dialects, *Proceedings of ICPhS XVI,* Saarbrücken, 6-10 August 2007, 965-968.

Dehé, N., 2009, Clausal parentheticals, intonational phrasing, and prosodic theory. *Journal of Linguistics* 45(3) : 569-615.

Dehé, N. & Braun, B., 2013, The prosody of question tags in English, *English Language and Linguistics* 17(1) : 129-156.

Dehé, N. & Wichmann, A., 2010, The multifunctionality of epistemic parentheticals in discourse: prosodic cues to the semantic-pragmatic boundary, *Functions of Language* 17(1) : 1-28.

Delais-Roussarie, E., 2003, Constitution et annotation de corpus : Méthode et Recommendations, in Delais-Roussarie, E. & Durand, J. (eds.), *Corpus et Variation en Phonologie du Français : Méthodes et Analyses,* Toulouse : Presses Universitaires du Mirail.

Delais-Roussarie, E., Avanzi, M. & Herment, S. (eds.), (à paraître), *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*, New York : Springer Verlag.

Delais-Roussarie, E. & Post, B., à paraître, Corpus annotation and transcription systems, in Durand, J., Gut, U. & Kristoffersen, G. (eds.), *Handbook of Corpus Phonology*, Oxford : Oxford University Press.

Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S-A., Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R. & Yoo, H-Y., à paraître, Developing a ToBI system for French, in Frota, S. & Prieto, P. (eds.), *Intonational Variation in Romance*, Oxford : Oxford University Press.

De Looze, C., Zhi, N., Auran, C., Cho, H-S., Herment, S., Nesterenko, I., 2012, Providing linguists with better tools: Daniel Hirst's contribution to prosodic annotation, *Proceedings of the 6th International Conference on Speech Prosody,* Shanghai, Chine, 615-620.

Deschamps, A., Duchet, J-L., Fournier, J-M., O'Neil, M., 2004, *English Phonology and Graphophonemics*, Paris : Ophrys.

Di Cristo, A., 2013, *La prosodie de la parole*, Bruxelles : Boeck-Solal

Diana, A., 2010, La phonétique dans l'enseignement de l'anglais aux spécialistes d'autres disciplines : enjeux et priorités, *Cahiers de l'APLIUT vol. XXIX n°3 : Phonétique, phonologie et enseignement des langues de spécialité* – Volume 2. http://apliut.revues.org/

Du Bois, J.W., Schuetze-Coburn, S., Cumming, S. & Paolino, D., 1993, *Outline of discourse transcription,* in Edwards, J. & Lampert (eds.), *Talking Data : Transcription and Coding Methods for Language Research,* Hillsdale, NL : Lawrence Erlbaum.

Dubois-Charlier, F. & Vautherin, B., 1997, *Syntaxe Anglaise*, Paris : Vuibert.

Duchet, J-L, 1994, *Code de l'anglais oral*, 2ème édition, Paris : Ophrys.

Durand, J., Laks, B. & Lyche, C., 2002, La phonologie du français contemporain : usages, variétés et structure, In Pusch, C. & Raible, W. (eds.), *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*. Tübingen : Gunter Narr Verlag, 93-106. http://www.projet-pfc.net

Eckman, F.R., 2004, From Phonemic Differences to Constraint Rankings, Research on Second Language Phonology, *Studies in Second Language Acquisition* 26, 513-549.

Edwards, J.E., 2001, The Transcription of Discourse, in Schiffrin, D., Tannen, D. & Hamilton, H.E., *The Handbook of Discourse Analysis*, Oxford : Blackwell, 321-343.

Elliott, D.E., 1971, *The Grammar of Emotive and Exclamatory Sentences,* Doctoral Dissertation, The Ohio State University, Columbus, Ohio, reproduced in Working Papers in Linguistics 8, Computer and Information Center, The Ohio State University.

Feller, J., 1969, Les mots en réserve d'emploi : éloge de la redondance, *Communication et langages* N°2, 5-13.

Féron, A., 2012, A study of the Pronunciation of the Indefinite Article 'A' in Spoken English, Mémoire de Master 1, sous la direction de S. Herment, Aix-Marseille Université.

Féron, A., 2013, *L'article indéfini the : prononciation et pragmatique*, Mémoire de Master 2, sous la direction de S. Herment, Aix-Marseille Université.

Fournier, J-M., 2010, *Manuel d'anglais oral,* Paris : Ophrys.

Fraser, H., 2001, *Teaching Pronunciation: A handbook for teachers and trainers. Three Frameworks for an Integrated Approach*, Australie : Department of Education Training and Youth Affairs (DETYA).

Gibbon, D., Moore, R. & Winski, R., (eds.), 1997, *Handbook of Standards and Resources for Spoken Language Systems, vol. 1 : Spoken Language Systems and Corpus Design*, Berlin : Mouton De Gruyter.

Gimson, A.C., 1962, *An Introduction to the Pronunciation of English*, Londres : Edward Arnold.

Ginésy, M., 1995, *Mémento de phonétique anglaise*, Paris : Nathan Université.

Giraudo, H., Przewozny, A. & Tarrier, J-M., 2013, LVTI / Language, Urban Life, Work and Identity : Ways and Means, *Conférence international PAC 20131,* 30-31 mai 2013, Aix-en-Provence, France.

Goldman-Eisler, F., 1968, *Psycholinguistics : Experiments in Spontaneous Speech,* New York : Academic Press.

Goldsmith, J., 1976, *Autosegmental phonology*, PhD Thesis, MIT, Distributed by IUCL, and published 1979 by Garland Press : NewYork.

Goudet, L-G., 2013a, *CATSSAA, Corpus of the Accent Tag : Spoken Scots and African American*, SLDR000830. http://sldr.org/sldr000830

Goudet, L-G., 2013b, Phonologie et lexicologie « sauvages » : l'Accent Tag sur Youtube, *53ème Congrès de la SAES*, 17-19 mai 2013, Dijon, France.

Grabe, E., 2004, Intonational variation in urban dialects of English spoken in the British Isles. In Gilles, P. and Peters, J. (eds.), *Regional Variation in Intonation.* Linguistische Arbeiten, Tübingen : Niemeyer, 9-31.

http://www.phon.ox.ac.uk/files/apps/IViE/

Grabe, E., Nolan, F. & Farrar, K.J., 1998, IViE, a comparative transcription system for intonational variation in English, *Proceedings of the 5th international conference on spoken language processing (ICSLP'98)*, 30 novembre-4 décembre 1998, Sydney, Australie, vol. 4 : 1259-62.

http://www.phon.ox.ac.uk/files/apps/old_IViE/guide.html

Grabe, E. & Post, B., 2004, Intonational variation in the British Isles, in Sampson G. & McCarthy, D. (eds.), *Corpus Linguistics: Readings in a Widening Discipline*, Continuum International.

Grabe, E., Post, B. & Nolan, F., 2001, Modelling Intonational Variation in English. The IViE system, in S. Puppel & G. Demenko (eds.), *Proceedings of Prosody 2000*, Adam Mickiewitz University, Poznan, Poland.

Greenbaum, S., 1996, Comparing English Worldwide: The International Corpus of English. Oxford : Clarendon Press.

Grice, M. & Baumann, S., 2002, Deutsche Intonation und GToBI. *Linguistische Berichte* 191, 267-298.

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J., Batliner, A., 1996, Consistency in Transcription and Labeling of German Intonation with GToBI, in *Proceedings of the International Conference On Spoken Language Processing*, Philadelphia, PA, USA, vol. 3, 1716-1719.

Grimshaw J., 1977, *English Wh- Constructions and the theory of Grammar*, Doctoral dissertation, University of Massachusetts, Amherst, Massachusetts.

Guillaume, B., 2006, *Approche énonciative des question tags en anglais contemporain.* Numéro spécial des Cahiers de Recherche, Paris : Ophrys.

Gussenhoven, C. 1983, A three-dimensional scaling of nine English tones, *Journal of Semantics* 2, 186-204.

Gussenhoven, C., 1984, *On the grammar and semantics of sentence accent*, Dordrecht : Foris.

Gussenhoven, C., 2004, *The phonology of tone and intonation,* Cambridge : Cambridge University Press.

Gut, U., 2009, *Non-native Speech. A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German,* Francfort : Peter Lang.

Guy, G., Horvath, B., Vonwiller, J., Daisley, E. & Rogers, I., 1986, An intonation change in progress in Australian English, *Language in Society* 15, 23-51.

Habert, B., Nazarenko, A. & Salem, A., 1997, *Les linguistiques de corpus*. Collection U, série « Linguistique », Paris : Armand Colin.

Halliday, M.A.K., 1967, *Intonation and Grammar in British English,* La Hague : Mouton.

Halliday, M.A.K. & Greaves, W.S., 2008, *Intonation in the Grammar of English*, London-Oakville: Equinox.

Halliday, M.A.K. & Hasan, R., 1976, *Cohesion in English,* London : Longman.

Harrington, J., Cox, F., and Evans, Z., 1997, An acoustic phonetic study of broad, general, and cultivated Australian English vowels, *Australian Journal of Linguistics* 17, 155-184.

Harris, Z., 1952, Discourse Analysis, *Language* 28, 1-30.

Herment-Dujardin, S., 1997, *Experimental validation of the intonation transcription system INTSINT in English*, Mémoire de DEA, Université de Provence, Aix-en-Provence, France.

Herment-Dujardin, S., 2001, *L'emphase dans le discours spontané anglais : corrélats acoustiques et prosodiques*, Thèse de doctorat, Laboratoire Parole et Langage, Université de Provence, Aix-en-Provence, France.

Herment, S., 2006, Etude perceptive sur la focalisation en anglais, in Wlodarczyk A. & H. (ed.), *La focalisation dans les langues*, Paris : L'Harmattan, 309-318.

Herment, S., 2009, Prédication et emphase en anglais : le rôle de la prosodie, in J-M. Merle (ed.), *La Prédication*, Bibliothèque de faits de langue, vol. 30 & 31, 243-252.

Herment, S., 2010a, COEC: a Corpus of English Conversation. SLDR. http://sldr.org/sldr000014/fr

Herment, S., 2010b, Emphase prosodique et emphase syntaxique : le cas de « do » dans un corpus de parole naturelle, *Revue CORELA* – Numéro thématique : *Paroles*. URL : http://corela.edel.univ-poitiers.fr/index.php?id=1059

Herment, S., 2010c, The pedagogical implications of variability in transcription, the case of [i] and [u], in Henderson, A. (ed.), *English Pronunciation: Issues and*

Practices (EPIP): Proceedings of the First International Conference (3-5 juin 2009, Université de Savoie, Chambéry, France), 177-188.

Herment, S., 2011, Relations entre prosodie et syntaxe : le cas de quelques structures syntaxiques non neutres, *Anglophonia 30*, Toulouse : Presses Universitaires du Mirail, 101-117.

Herment, S., 2013, Point d'enquête Marseille pour Prieto, P., Borràs-Comes, J., & Roseano, P. (coords.) (2010-2014). *Interactive Atlas of Romance Intonation*. Web page: http://prosodia.upf.edu/iari/ Mis en ligne juin 2013.

Herment, S., à paraître (a), Interdépendances entre prosodie et syntaxe dans les clivées en IT, *Actes du XIV colloque de Villetaneuse sur l'anglais oral* (3-4 avril 2008, Villetaneuse, France).

Herment, S., à paraître (b), Is intensity a relevant criterion in the perception of spontaneous speech? The case of emphasis in English, *Revue CORELA*, ISSN 1638-5748.

Herment, S. & Hirst, D.J., 2002, Emphasis in English : a perceptual study based on manipulated synthetic speech, in Bel, B. & Marlien, I. (eds.), *Proceedings of the First International Conference on Speech Prosody* (11-13 avril 2002 : Aix-en-Provence, France), 379-382.

Herment, S. & Hirst, D.J., 2003, L'emphase en anglais : critères acoustiques et analyse statistique. Tentative de définition. *Actes d'Interfaces Prosodiques* (27-29 mars 2003 : Nantes, France), 51-56.

Herment, S., Kerfelec, V., Leonarduzzi, L. & Turcsan, G., 2010, *A Learners' Corpus of Reading Texts*. SLDR. http://sldr.org/sldr000015/fr

Herment, S., Loukina, A., Tortel, A., 2012, *AixOx*. SLDR. http://sldr.org/sldr000784/fr

Herment, S. & Leonarduzzi, L., 2012, The Pragmatic functions of prosody in English Cleft Sentences, *Proceedings of the Sixth International Conference on Speech Prosody*, (22-25 mai 2012 : Shanghai, Chine), 713-716.

Herment, S. & Leonarduzzi, L., à paraître, L'expression de la saillance dans les clivées en *wh-* en anglais contemporain : prosodie, syntaxe, discours, in Boisseau, M. & A. Hamm (eds.), *La Saillance en Langue et en Discours*, Presses Universitaires de Franche Comté.

Herment, S., Tortel, A., Bigi, B., Hirst, D., Loukina, A., à paraître, AixOx, a multi-layered learners corpus: automatic annotation, in Díaz Pérez Javier & Díaz Negrillo Ana (eds.), *Specialisation and variation in language corpora*, Bern : Peter Lang.

Herment-Dujardin, S. & Rouveyrol, L., 2011, Emphase et variation individuelle dans le débat télévisé : paramètres acoustiques et pragmatico-énonciatifs, *Actes du XII° colloque de Villetaneuse sur l'anglais oral* (2-3 avril 2004, Villetaneuse, France), 59-70.

Herment, S. & Turcsan, G., 2013, A heuristic corpus for English word prosody: disyllabic nonce words, *Proceedings of the TRASP workshop* (30 août 2013, Aix-en-Provence, France). http://aune.lpl.univ-aix.fr/~bigi/TRASP/19894.pdf

Herment, S. & Turcsan, G., à paraître, Revisiting contrasts for high vowels in English: Allophony, Neutralisation and Phonological corpora, in Ballier, N. & C. Migette, *Corpora in English Phonology*, Newcastle : Cambridge Scholars Publishing.

Herry-Bénit, N., 2011, *Didactique de la phonétique anglaise*, Rennes : Presses Universitaires de Rennes.

Hewings, M., 2005, *Advanced Grammar in Use*, Cambridge : Cambridge University Press.

Hewings, M., 2007, *Pronunciation practice activities*, Cambridge : Cambridge University Press.

Hilton, H., 2008, Corus PAROLE – Parallèle Oral en Langue Etrangère, http://www.umr7023.cnrs.fr/sites/sfl/IMG/pdf/PAROLE_manual.pdf

Hirst, D.J., 1977, *Intonative Features : a Syntactic Approach to Prosody*, The Hague : Mouton.

Hirst, D.J., 1998, Intonation in British English, in Hirst, D.J. & Di Cristo, A. (eds.) : 56-77.

Hirst, D.J., 1999, The symbolic coding of duration and timing : an extension to the INTSINT system, *Proceedings of Eurospeech '99* (septembre 1999, Budapest, Hongrie), 1639-1642.

Hirst, D.J., 2005, Prosodic aspects of speech and language, in Brown, K. (ed.), *Encyclopedia of Language and Linguistics*, Elsevier Science, vol X, 167-178.

Hirst, D.J., 2008, Could the rising intonation patterns of Urban Northern Britain be Nordic Prosody? *Nordic Prosody* 10 (4-6 août 2008), Helsinki, Finlande.

Hirst, D.J., 2009, Declarative High Rising Tones and the frequency code, *PROSICO, International Conference on Prosody and Iconicity* (9-10 avril 2009, Rouen, France.

Hirst, D., Bigi, B., Cho, H., Ding, H., Herment, S. & Wang, T., 2013 Building OMProDat: an open multilingual prosodic database, *Proceedings of the TRASP workshop* (30 août 2013, Aix-en-Provence, France), http://aune.lpl.univ-aix.fr/~bigi/TRASP/19767-tag.pdf

Hirst, D.J. & Di Cristo, A., 1998, *Intonation Systems : a Survey of Twenty Languages,* Cambridge : Cambridge University Press.

Hirst, D.J., Di Cristo, A., Espesser, R., 2000, Levels of representation and levels of analysis for intonation, in Horne, M. (ed.), *Prosody, theory and experiment,* Dordrecht : Kluwer Academic Publishers, 51-88.

Huart, R., 2002, *Grammaire de l'anglais oral*, Paris : Ophrys.

Huart, R., 2010, *Nouvelle grammaire de l'anglais oral*, Paris : Ophrys.

Huddleston, R., 1984, *Introduction to the Grammar of English*, Cambridge Textbooks in Linguistics, Cambridge : Cambridge University Press.

Huddleston, R., 1993, Remarks on the construction *you won't believe who Ed has married*, *Lingua* 91, 175-184.

Huddleston, R., Pullum, G.K., 2002, *The Cambridge Grammar of the English Language*, Cambridge : Cambridge University Press.

Hume, E., 2004, Deconstructiong markedness : a predictability-based approach, *Proceedings of the annual meeting, Berkeley Linguistics Society* 13, 182-198.

Hume, E., 2011, Markedness, in Oostendorp van, M., Ewen, C.J., Hume, E. & Rice, K. (eds.), *The Blackwell Companion to Phonology*, Chichester : Wiley-Blackwell, Vol. 1, 79-106.

Ide, N. & Véronis, J., 1995, *Text Encoding Initiative. Background and Context*, Dordrecht : Academic Press.

Jassem, W., 1952*, Intonation of Conversational English (Educated Southern British)*, Wroclaw : Wroclawskie Towarszystwo Naukow.

Jones, D., 1909a, *The pronunciation of English*, second edition 1914, third edition 1950, 4ème édition 1956, Cambridge : Cambridge University Press.

Jones, D., 1909b, *Intonation curves,* Berlin : Teubner.

Jones, D., 1918, *An outline of English phonetics*, 1ère édition (9ème édition 1960), Leipzig : Teubner.

Jones, D., 2011, *English Pronouncing Dictionary*, 18th edition edited by P. Roach, J. Setter, & J. Esling, Cambridge : Cambridge University Press.

Kaltenböck, G., 2004, Using non-extraposition in spoken and written texts: A functional perspective, in Aijmer, K. & Stenström, A-B. (eds.), *Discourse Patterns in Spoken and Written Corpora*, Amsterdam- Philadelphia: John Benjamins, 219-242.

Kerfelec, V., 2009, *L'exclamation en français et en anglais. Formes, sens, effets,* Collection Langue & Langage, Aix-en-Provence : Publications de l'Université de Provence.

Kerswill, P., 2003, Dialect levelling and geographical diffusion in British English, in Britain, D. & Cheshire, J. (eds.), *Social dialectology: In honour of Peter Trudgill*, Amsterdam, Benjamins, 223-243.

Khalifa, J-C., 1999, *La syntaxe anglaise aux concours, théorie et pratique de l'énoncé complexe*, Paris : Armand Colin.

Kingdon, R., 1958, *The Groundwork of English Intonation,* Londres : Longman.

Kirk, J.M., 2003, The Irish component of the International Corpus of English, *7th International Conference on the Languages of Scotland and Ulster*, Dumfries.

Knowles, G., 1987, *Patterns of Spoken English,* London : Longman.

Knowles, G., Wichmann, A. & Alderson, P., 1996, *Working with Speech*, Londres : Longman.

Knowles, G., Williams, B. & Taylor, L., 1996, *A corpus of formal British Speech : The Lancaster /IBM Spoken English Corpus*, Londres : Longman.

Kochanski, G., Grabe, E., Coleman, J. & Rosner, B., 2005, Loudness predicts prominence: fundamental frequency lends little, *Journal of the Acoustical Society of America* 118, 1038-1054.

Labov, W., 1972, *Sociolinguistic patterns*, Pennsylvanie : University of Pennsylvania Press.

Ladd, R., 1980, *The structure of intonational meaning: evidence from English,* Bloomington : Indiana University Press.

Ladd, R., 2008, *Intonational Phonology*, 2ème édition, Cambridge : Cambridge University Press.

Ladefoged, P., 2005, *Vowels and Consonants,* 2ème édition (1ère edition 2001), Oxford : Blackwell.

Lambrecht, K., 2001, A Framework for the Analysis of Cleft Constructions, *Linguistics* 3(3): 463-516.

Lapaire, J.-R. & Rotgé, W., 1991, *Linguistique et grammaire de l'anglais,* Toulouse : Presses Universitaires du Mirail.

Larreya, P. & Rivière, C, 2005, *Grammaire explicative de l'anglais*, 3ème édition, (1ère édition 1991, Paris : Longman France), Paris : Pearson Education France.

Laver, J., 1980, *The Phonetic description of voice quality*, Cambridge : Cambridge University Press.

Leben, W., 1976, *Suprasegmental phonology*, PhD Thesis, MIT, publiée en 1980 par Garland Press : NewYork.

Lehiste, I., 1975, The phonetic structure of paragraphs, in Cohen, A. & Nooteboom, S.G. (eds.), *Structure and Process in Speech Perception*, New York : Springer Verlag.

Léon, P., 1993, *Précis de phonostylistique, Parole et expressivité,* Paris : Nathan Université.

Leonarduzzi, L., 2004, La subordonnée interrogative en anglais contemporain, Aix-en-Provence, Presses universitaires de Provence.

Leonarduzzi, L., 2008, Dislocations à gauches et antépositions : des variantes en distribution complémentaire ? *Actes du 46ème Congrès de la SAES* (12-14 mai

2006, Nantes, France), Saint-Etienne: Publications de l'Université de Saint-Etienne, 121-142.

Leonarduzzi, L. & Herment, S., 2013a, Non-canonical Syntactic Structures in Discourse: Tonality, Tonicity and Tones in English (Semi-)spontaneous Speech, *Proceedings of Interspeech 2013* (25-29 août 2013, Lyon, France).

Leonarduzzi, L. & Herment, S., 2013b, The limits between extraposition and right-dislocation, *Revue CERCLES*, numéro 29. http://www.cercles.com

Liberman, M., 1975, *The intonational system of English,* PhD Thesis, MIT. Distributed 1978 by IUCL.

Lilly, R. & Viel, M., 1998a, *Initiation raisonnée à la phonétique de l'anglais,* édition révisée (1ère édition 1975), Paris : Hachette Supérieur.

Lilly R. & Viel M., 1998b, *La prononciation de l'anglais, Règles phonologiques et exercices de transcription*, Edition corrigée (1ère édition 1977), Paris : Hachette Université.

MacWhinney, B., 2000, *The CHILDES Project : Tools for Analysing Talk*. 3ème édition, Mahwah, NJ : Lawrence Arlbaum Associates.

Martin, P., 2005, Winpitch TLT, un logiciel multimédia d'enseignement de la prosodie, *Revue Alsic* vol. 8 n°2, 95-108.

Martin, P., 2009, *Intonation du français*, Paris : Armand Colin.

Martinet, A., 1955, *Économie des changements phonétiques*, Berne, Francke.

Martinet, A., 1960, *Eléments de linguistique générale,* Paris: Armand Colin.

Mertens, P., 2004a, Un outil pour la transcription de la prosodie dans les corpus oraux, *Traitement Automatique des langues (TAL)* 45 (2) : 109-130.

Mertens, P., 2004b, The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model, *Proceedings of the 2nd International Conference on Speech Prosody* (23-26 Mars 2014), Nara, Japon.

Mertens, P., 2011, A system for automatic tonal annotation for speech corpora, *IDP (Interface Discours Prosodie) Conference*, University of Salford, September 2011.

Milde, J.T. & Gut, U., 2002, A prosodic corpus of non-native speech, in Bel B. & Marlien, I. (eds.), *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence : France, 503-506.

Miller, J. & R. Weinert, 1998, *Spontaneous Spoken Language, Syntax and Discourse*, Oxford : Oxford Clarenden Press.

Milroy, L., 1980, *Language and social networks*, Oxford : Blackwell.

Nguyen, N. & Adda-Decker, M. (eds.), 2013, *Méthodes et outils pour l'analyse des grands corpus oraux*, Lavoisier.

Ní Chasaide, A. & Gobl, C., 2000, Testing affective correlates of voice quality through analysis and resynthesis, *Proceedings of the ISCA workshop on Speech and Emotion* : 178-183.

Nicaise, A. & Gray, M., 1998, *L'intonation de l'anglais*, Paris : Nathan Université, Collection 128.

Nolan, F. and Grabe, E., 1997, Can ToBI transcribe intonational variation in the British Isles?, *Proceedings of the ESCA Tutorial and Research Workshop on Intonation,* Athènes, Grèce, 259-262.

Nubold, P. & Turner, M., 1987, *Death is the mother of beauty : mind, metaphor, criticism,* Chicago: University of Chicago Press.

O'Connor, J.D., Arnold, G.F., 1961, Intonation of Colloquial English, 2ème édition 1973, Londres : Longman.

Ochs, E., 1979, Transcription as theory, in Ochs, E. & Schieffellin, B.B. (eds.), *Developmental Pragmatics*, New-York : Academic Press, 43-72.

Ohlander, S., 1985, That's when your heartaches begin, *Papers on Language and Literature presented at Alvar…,* Gottenburg Studies in English, n°60, 280-302.

Oliviéro, C., 2012, *Acquisition de la prosodie : anglais langue maternelle*, Mémoire de Master 2 sous la direction de S. Herment, Aix-Marseille Université.

Palmer, H.E., 1922, *English intonation, with systematic exercises,* Cambridge : Heffer.

Peeters, B. 1992, *Diachronie, phonologie, et linguistique fonctionnelle*, Louvain-la-Neuve : Peeters.

Pierrehumbert, J. & Hirschberg, J., 1990, The meaning of intonational contours in the interpretation of discourse, in Cohen, P.R., Morgan, J., & Pollack, M.E. (eds.), *Intentions in communication*, Cambridge, MA: MIT Press, 271–311.

Pierrehumbert, J., 1980, *The phonology and phonetics of English intonation,* PhD Thesis, MIT. Published 1988 by IUCL, Bloomington, IN.

Pike, K.L., 1945, *The intonation of American English*, Ann Arbor : University of Michigan Press.

Pitrelli, J., Beckman, M., Hirschberg, J., 1994, Evaluation of Prosodic Transcription Labeling Reliability in the ToBI framework, in *Proceedings of the International Conference On Spoken Language Processing*, Yokohama, Japon, vol. 2, 123-126.

Prieto, P., Borràs-Comes, J., & Roseano, P. (Coords.), 2010-2014, *Interactive Atlas of Romance Intonation*. Web page: http://prosodia.upf.edu/iari/.

Prince, E., 1978, A Comparison of WH-clefts and IT-clefts in Discourse, *Language, Journal of the Linguistic Society of America 54(4)*: 883-906.

Rahilly, J., 1997, Aspects of prosody in Hiberno-English: the case of Belfast. In Kallen, J. (ed.), *Focus on Ireland* (Varieties of English Around the World G21), Amsterdam : John Benjamins,109-32.

Ranger, G., 2004, The auxiliary DO in affirmative contextes, *Anglophonia* 16, Toulouse : Presses Universitaires du Mirail, 99-120.

Rasier, L. & Hiligsmann, P., 2007, Prosodic transfer from L1 to L2. Theoretical and methodological issues, *Nouveaux cahiers de linguistique française* 28, 41-66.

Repp, B., 1992, Probing the cognitive representation of musical time : structural constraints on the perception of timing perturbations, *Haskins Laboratories Status Report on Speech research,* 111/112, New Haven, Connecticut, 293-320.

Rivara, R., 1979, La quantification corrélative, *Sigma* n°4, Publications du CELAM, Montpellier, 65-107.

Rivara, R., 2004, Marqueurs intonatifs et marqueurs textuels. Le cas du *do* anglais, in Rivara, R., *Pragmatique et Enonciation*, Aix-en-Provence : Publications de l'Université de Provence, 261-294.

Roach, P., 2009, *English Phonetics and Phonology, A Practical Course*, 4ème édition (1ère édition 1983), Cambridge : Cambridge University Press.

Roach, P., Knowles, G., Varadi, T., Arnfield, S., 1993, MARSEC : A Machine-Readable Spoken English Corpus, *Journal of the International Phonetic Association,* 23/2 : 47-53.

Rodrigues Da Mota, C., 2010, *Eh, marqueur de discours au Canada,* Mémoire de Master 2, sous la direction de S. Herment, Aix-Marseille Université.

Rodrigues Da Mota, C. & Herment, S., 2011, *Le marqueur eh et l'intonation montante terminale en anglais canadien : similitudes et différences*, Journée d'études « combinatoire de marqueurs en anglais oral contemporain », Paris 7 Diderot, 1er avril 2011.

Sabio, F., 2011, *Syntaxe et organisation des énoncés - Observations sur la grammaire du français parlé,* Habilitation à diriger des recherches, Laboratoire Parole et Langage, Aix-en-Provence, France.

Santiago-Vargas F. & Delais-Roussarie, E., 2012a, La prosodie des énoncés interrogatifs en français L2, *Actes des Journées d'études sur la Parole (JEP-TALN 2012)*, Grenoble, France, .

Santiago-Vargas F. & Delais-Roussarie, E., 2012b, Acquiring phrasing and intonation in French as a second Language: the case of Yes-No questions produced by Mexican Spanish Learners, *Proceedings of the Sixth International Conference on Speech Prosody*, (22-25 mai 2012 : Shanghai, Chine), 338-341.

Scheer, T., 2004, Le corpus Heuristique : un outil qui montre mais ne démontre pas, *Corpus 3 : Usage des corpus en phonologie,* 153-192. Mis en ligne le 02 décembre 2005, consulté le 17 janvier 2013. URL : http://corpus.revues.org/210.

Schiffrin, D., 2001, Discourse markers : Language, Meaning and Context, in Schiffrin, D., Tannen , D. & Hamilton, H.E., *The Handbook of Discourse Analysis*, Oxford : Blackwell, 54-75.

Schottman, W., 2012, *Advanced English Pronunciation,* Paris : Ophrys.

Schuetze-Coburn, S., Shapley, M., & Weber, E.G., 1991, Units of intonation in discourse: a comparison of acoustic and auditory analyses, *Language and Speech*, 34.3, 207-34.

Selting, M., 1994, Emphatic speech style - with special focus on the prosodic signalling of heightened emotive involvement in conversation, *Journal of Pragmatics,* 22 : 375-408.

Shattuck-Hufnagel, S. & A. E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25, 193–247.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M. Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J., 1992, ToBI: A standard for labelling English prosody, In Ohala, J. J., Nearey, T.B., Derwing, B.L., Hodge M.M. & Wiebe, G. E. (eds.), *ICSLP 92 Proceedings: 1992 International Conference on Spoken Language Processing*, Volume 2, Department of Linguistics, University of Alberta, 867-870.

Simon, A-C., Auchlin, A., Avanzi, M., Goldman, J.-P., 2010, Les phonostyles. Une description prosodique des styles de parole en français, in Abecassis, M. & Ledegen G. (eds.), *Les voix des Français, vol. 2 : en parlant, en écrivant,* Berne: Peter Lang, 71-88.

Sinclair, J., 1996, *Preliminary recommandations on corpus typology*, Technical report, EAGLES (Expert Advisory Group on Language Engineering Standards).

Sluijter, A., 1995, *Phonetic correlates of stress and accent*, Ph.D. dissertation, University of Leiden.

Sluijter, A. & van Heuven, V., 1993, Perceptual clues of linguistic stress : intensity revisited, *Working Papers 41, Proceedings of the ESCA Workshop on Prosody,* Lund University, Sweden, 246-249.

Sluijter, A. & van Heuven, V., 1996, Spectral balance as an acoustic correlate of linguistic stress, *Journal of the Acoustical Society of America* 100/4 : 2471-2485.

Smolka, V., 2005, Non-extraposed Subject Clauses, in *Theory and Practice in English Studies 3: Proceedings from the Eighth Conference of British, American and Canadian Studies*, Brno, MasarykovaUniverzita, 137-143.

Souesme, J.-C., 1986, DO modalité de rang 1, *SIGMA* 10, Aix-en-Provence et Montpellier, 51-81.

Stuart-Smith, J., à paraître, *Mediating the local: The role of television for a changing urban vernacular*, Oxford : Oxford Studies in Sociolinguistics, Oxford University Press.

Syrdal, A., McGorg, J., 2000, Inter-Transcriber Reliability of ToBI Prosodic Labeling, in *Proceedings of the International Conference On Spoken Language Processing*, Beijing, Chine, vol. 3, 235-238.

Szlamowicz, J., 2011, *Oh yes* ! *well*, *you know*, *I mean* : combinatoire de marqueurs et co-énonciation à l'oral, *Journée d'études OSLiA*, 1er avril 2011, Paris, France.

Tannen, D., 1984, *Conversational Style,* Norwood, NJ : Ablex.

Tench, P., 1996, *The Intonation Systems of English,* London: Cassell.

't Hart, J., Collier, R., & Collier, A., 1990, *A Perceptual Study of Intonation*, Cambridge : Cambridge University Press.

Toivanen, J., 2005, ToBI or not ToBI? Testing two models in teaching English intonation to Finns, *Proceedings of the Phonetics Teaching and Learning Conference 2005*.

Tortel, A., 2008, ANGLISH : base de données comparatives L1 et L2 de l'anglais lu, répété et parlé, *Travaux Interdisciplinaires du Laboratoire Parole et Langage (TIPA)*, 27, 111-122. http://sldr.org/sldr000731/fr

Trager, G.L. & Smith, H.L., 1951, *An Outline of English Structure,* Norman, Oklahoma : Battenburg Press.

Trouvain, J. & Gut, U. (eds.), 2007, *Non-Native Prosody. Phonetic Description and Teaching Pactice,* Berlin : Mouton de Gruyter.

Trudgill, P., 1986, *Dialects in contact,* Oxford : Blackwell.

Turcsan, G. & Herment, S., 2013, Making sense of nonce word stress in English, *Proceedings of EPIP 3 (3rd International conference on English Pronunciation : Issues and Practices),* 8-10 mai 2013, Murcie, Espagne.

Turcsan, G. & Herment, S., (en attente de mise en ligne), The Aix-Di-Nonce Corpus, SLDR.

Turcsan, G. & Herment, S., à paraître, L'anglais du Nord de l'Irlande, in Brulard, I., Durand, J. and Carr, P. (eds.), *La prononciation de l'anglais contemporain*, Toulouse : Presses Universitaires du Mirail.

Venditti, J., 1995, *Japanese ToBI Labeling Guidelines*, Manuscript with examples, Ohio State University.

Viel, M., 1984, *La notion de marque chez Trubetzkoy et Jakobson, un épisode de l'histoire de la pensée structurale*, Paris : Didier Erudition.

Viel, M., 2003, *Manuel de phonologie anglaise*, Paris : Armand Colin CNED.

Wales, K., 2006, *Northern English, a cultural and social history*, Cambridge: Cambridge University Press.

Walker, J., 1781, *Elements of elocution*, two volumes, Reprinted by the Scholar Press Menston, Yorkshire, 1969, *English Linguistics 1500-1800* n° 148.

Walker, J., 1787, *The melody of speaking delineated,* London : Printed for the Author, Reprinted by the Scolar Press, Menston, Yorkshire, 1970.

Wallis, S., Aarts, B & Nelson, G., 1999, Parsing in reverse – Exploring ICE-GB with Fuzzy Tree Fragments and ICECUP, in Kirk, J. (ed.), *Corpora Galore: Analyses and Techniques in Describing English*, Amsterdam: Rodopi, 335-344.

Watbled, J-P., 1996, *La prononciation de l'anglais,* Paris: Nathan Université, collection 128.

Wells, W.H.G., 1986, An experimental approach to the interpretation of focus in spoken English, in Johns-Lewis, C., (ed.), *Intonation in Discourse,* London & Sydney : Crook Helm : 53-75.

Wells, J.C., 1990, Syllabification and allophony, in Ramsaran, S. (ed.), *Studies in the pronunciation of English, A commemorative volume in honour of A.C. Gimson,* London and New York: Routledge, 76-86.

Wells, J.C., 2006, *English Intonation, An Introduction*, Cambridge : Cambridge University Press.

Wells, J.C., 2008, *Longman Pronunciation Dictionary*, 3rd edition, London : Longman.

Wells, J., Barry, W., Grice, M., Fourcin, A., & Gibbon, D., 1992, Standard Computer-Compatible Transcription. *Esprit project 2589 (SAM),* Doc. no. SAM-UCL-037. Londres : Phonetics and Linguistics Dept., UCL.

Wichmann, A., 2000, *Intonation in Text and Discourse, beginnings, middles and ends*, Harlow : Pearson.

Wightman, C., 2002, ToBI or not ToBI?, in Bel, B. & Marlien, I. (eds.), *Proceedings of the First International Conference on Speech Prosody* (11-13 avril 2002 : Aix-en-Provence, France), 25-29.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P., 1991 Segmental durations in the vicinity of prosodic phrase boundaries*, Journal of the Acoustical Society of America* 91, 3 : 1707-1717.

Wilhelm, S., 2011, Innovations segmentales et suprasegmentales dans le NW Yorkshire : Implications pour l'étude du changement accentuel dans l'anglais des îles Britanniques. Thèse de Doctorat, Université de Bourgogne, France. <http://ressources-numeriques.u-bourgogne.fr/theses/27009150/these_WILHELM_STE-PHAN_00_00.pdf>

Winkler, P., 1984, Interrelations between Fundamental Frequency and Other Acoustic Parameters of Emphatic Segments, in Gibbon, D. & Richter, H. (eds.), *Intonation, Accent and Rythm, Studies in Discourse Phonology : Research in Text Theory,* Berlin, New York : de Gruyter, 327-338.

Wit, E-J.C. & Gillette, M., 1999, What is linguistic redundancy ? *Technical report*, november 1999, University of Chicago.

Yule, G., 1980, Speaker's topics and major paratones, *Lingua 52, North Holland, 33-47.*

Zerbian, S., 2012, Markedness in the prosody of contact varieties of South African English, *Proceedings of the Sixth International Conference on Speech Prosody*, (22-25 mai 2012 : Shanghai, Chine), 446-449.

Zipf, G. K., 1949, *Human behavior and the principle of least effort : an introduction to human ecology*, Addison-Wesley Press.

Dictionnaires en ligne consultés en juin 2013 :

Collins English Dictionary – Complete and Unabridged, http://www.collinsdictionary.com/dictionary/english

Longman Dictionary of Contemporary English, http://www.ldoceonline.com/

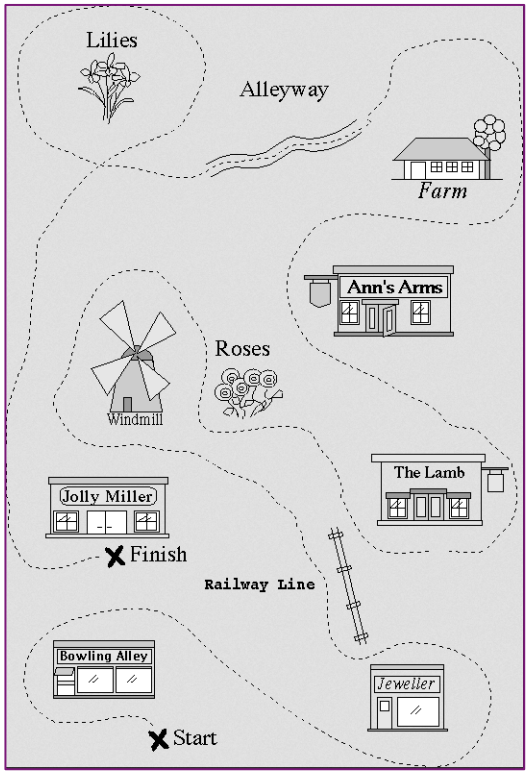
Oxford Dictionary of English, http://oxforddictionaries.com/

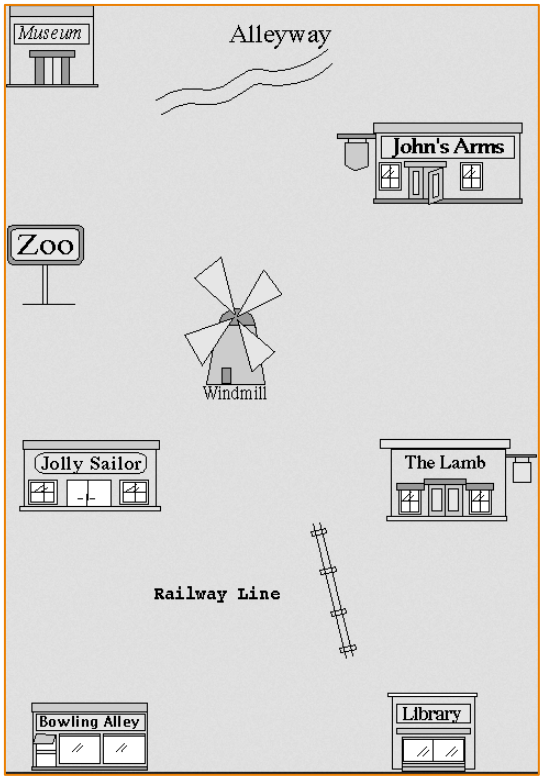
The American Heritage Dictionary of the English Language, http://www.ahdictionary.com/

Annexes

Annexe 1 : Maptask

Maptask du corpus IViE Bradford (Grabe *et al.* 1998, Grabe & Post 2004) fs 73





Annexe 2 : Fiche de renseignements / Information sheet

Date de l'enregistrement / Date of recording : ……

Lieu et conditions de l'enregistrement / Place and setting of the recording:……

Prénom / First name: ……

Nom / Name:……

Nationalité / nationality : ……..

Age au moment de l'enregistrement / Age at date of recording : ……

Lieu de naissance / Place of birth: ……

Lieu de résidence au moment de l'enregistrement / Current place of residence : ……

Lieux de résidence antérieurs / Previous places of residence:

Lieu/ place nombre d'années / number of years age (de_ à_) / age (from_ to_)

……..

……..

……..

……..

Activité / Occupation:……

Activités antérieures / Previous occupations:……

……

……..

Education (specifier jusqu'à quel âge et quel type d'éducation)/ Education (specify until what age and what type of education):……

……

……

……

Langues parlées / Languages spoken:

Langue/ language niveau/level of proficiency (basic) (intermediate) (fluent)

…… ☐ ☐ ☐

.....	☐	☐	☐
.....	☐	☐	☐
.....	☐	☐	☐

Langue / language	frequency of use	(rarely)	(monthly)	(daily)
.....		☐	☐	☐
.....		☐	☐	☐
.....		☐	☐	☐
.....		☐	☐	☐

Père de l'informant / Informant's father :

Année de naissance / year of birth: .....

Lieu d'origine / Place of origin:.....

Activité / Occupation: .....

Education / Education:.....

Langues ou dialect local parlé / Languages or local dialect spoken:.....

.....

Mère de l'informant / Informant's mother :

Année de naissance / year of birth: .....

Lieu d'origine / Place of origin:.....

Activité / Occupation: .....

Education / Education:.....

Langues ou dialect local parlé / Languages or local dialect spoken:.....

.....

Conjoint de l'informant / Informant's partner :.....

Lieu d'origine / Place of origin:.....

Activité / Occupation: .....

Education / Education:.....

Langues ou dialect local parlé / Languages or local dialect spoken:.....

.....

Nombre d'enfants, âge / Number of children, age : .....

.....

……

Personnes ayant joué un rôle important pour l'informant lors de l'acquisition de la langue / People who played an important role during the informant's acquisition of the language:

……

……

……

Endroit où l'informant réside (maison, appartement, quartier résidentiel, banlieue, etc…) Type of accommodation of the informant (house, flat, in a residential area, housing estate, block of flats, etc.): ……

Intégration et relation avec le voisinage / Integration into the area, relationships within the neighbourhood:

……

……

Groupe ethnique / Ethnic group: ……

Activités culturelles et de loisirs, voyages / Cultural and leisure activities, travels:

……

……

Information supplémentaire / Additional information: ……..

……

……

Remarques sur l'enregistrement / Remarks on the recording : ……..

……

……

……

Annexe 3 : Fiche signalétique IARI

NOM :

Prénom :

Sexe (entourez votre réponse) : masc. / féminin

Date de naissance (jj/mm/aaaa) :

Lieu de naissance (Pays et région) :

Adresse personnelle:

Langue(s) première(s) (si plusieurs, le dire):

Lieu de naissance de la mère:

Langue(s) première(s) de la mère :

Lieu de naissance du père:

Langue(s) première(s) du père:

Lieu de domiciliation actuel :

Lieu de scolarisation (au niveau de la région, et non des villes et villages) **et de domiciliation depuis le primaire**

Région (*Bretagne, Nord, Bavière*)	Pays (*France, Allemagne*)

Niveau d'études/ d'instruction (entourez votre réponse) :
(Si vous êtes, bac +4, mettez le niveau inférieur Bac + 3)

BEPC
BEP
BAC
BAC + 30
BAC + 5
Doctorat et plus
Autre Précisez

Profession (si besoin, mettre étudiant, écolier, mère/ père au foyer, etc.) :

Fait le

Signature

Annexe 4 : Consentement éclairé LPL

Je soussigné(e), ……..,

- autorise par la présente Mme/M. …… à m'enregistrer en audio/vidéo ;

- autorise l'utilisation de ces données, sous leur forme enregistrée aussi bien que sous leur forme transcrite et anonymisée

- à des fins de recherche scientifique (mémoires, thèses, articles scientifiques, exposés à des congrès, séminaires)
- à des fins d'enseignement universitaire
- pour une diffusion dans la communauté des chercheurs sous la forme d'éventuels échanges et prêts de corpus à des chercheurs, moyennant la signature d'une convention de recherche ou d'une licence non-commerciale
- pour une diffusion sur le site du CRDO (http://crdo.fr) dédié à la recherche

- prends acte que, pour toutes ces utilisations scientifiques, les données ainsi enregistrées seront anonymisées, cela signifie que les bandes audio qui seront présentées à des conférences ou des cours sous forme d'extraits seront nettoyées de toute information personnelle (nom, adresse, n° de téléphone…)

-souhaite que la contrainte supplémentaire suivante soit respectée :

……

……

……

Lieu et date : Signature

CONSENT FORM (LPL)

I the undersigned, ……..,

- give permission to Ms./M. …… to record me (audio/video)

- give permission to use the recordings and their transcriptions anonymously.

- For the purposes of scientific research (dissertations, PhD, scientific articles, conference talks, seminars)
- For the purposes of university teaching
- For distribution amongst the academic community by means of exchanging or lending corpora, dependent upon the signature of a research agreement or a non commercial licence.
- For broadcast on the CRDO website (http://crdo.fr) devoted to research

- note that for all scientific uses, the recorded data will be treated anonymously, meaning that all personal information (name, address, telephone n° etc.) will be deleted from any extracts on the audio tapes presented at conferences or in class.

- wish for the following supplementary constraint to be respected :

……

……

……

Place and date : Signature

Annexe 5 : Consentement éclairé IARI

UFR de Linguistique

U.F.R. Linguistique
case 7003
Université Paris Diderot-Paris7
75205 PARIS CEDEX 13

CNRS/ UMR 7110

Laboratoire de Linguistique Formelle
Case Postale 7031
5, rue Thomas Mann
F-75205 Paris Cedex 13

UNIVERSITAT POMPEU FABRA

Departament de Traducció i Ciències del llenguatge
Universitat Pompeu Fabra

Campus de la Comunicació-Poblenou
C/ Roc Boronat, 138
08018 Barcelona

Fiche de consentement

Je soussigné, M. Mme ....., personne majeure, ayant le numéro INSEE ou le numéro d'identité ..... domicilié à ....., né à ..... en ..... (année), déclare avoir été informé(e) de la finalité des données enregistrées que j'ai fournies à ..... (nom de la personne ayant fait l'enregistrement, université et pays d'origine), conformément à la loi espagnol 15/1999 du 13 décembre 1999 portant sur la Protection des données d'ordre personnelles et au décret royal 1720/2007 du 21 décembre 2007 demandant l'application de cette dernière, conformément à la législation française, et plus particulièrement à la loi *Informatique et Liberté*, et conformément à la directive européenne 95/46/CE.

Je manifeste mon accord :

1) Pour être enregistré(e) ;
2) Pour que ces enregistrements soient transcrits, analysés et utilisés à des fins scientifiques et pédagogiques, notamment pour mettre au point des projets de recherche dans le domaine de la linguistique. cela en préservant mon anonymat ;
3) que ces données soient traitées dans l'archive « *Projectes de recerca, desenvolupament i innovaci* » (Projets de Recherche/ Développement et Innovation) dont sont titulaires l'université ..... (université d'origine du chercheur ayant collecté les données, et indiquée dans l'en-tête) et l'Universitat Pompeu Fabra , en préservant mon anonymat.
4) que mes productions soient diffusées ou représentées en préservant mon anonymat.

J'autorise donc l'Université ..... par le biais des chercheurs .....et ....., ou leurs représentants, et l'Université Pompeu Fabra, par le biais du *Grup d'Estudis de Prosòdia* (coordonné par le Dr. Pilar Prieto), à reproduire et publier par quelque moyen les enregistrements auxquels j'ai participé, ceci avec des fins purement académiques et scientifiques sans un esprit lucratif. Ces enregistrements feront partie du site web de l'*Atlas interactif de l'intonation dans les langues romanes*, consacré à l'apprentissage et à l'étude des langues romanes.

J'autorise donc 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 et l'Institution bénéficiées n'en profitent aucunement pour faire des activités lucratives.

Je déclare également avoir été informé(e) que l'organisme responsable des archives de l'Université est le Président et que les droits d'accès, de rectification, d'annulation et d'opposition à ces données peuvent s'exercer par une demande écrite accompagnée d'une photocopie d'un papier d'identité ou d'un autre document justificatif équivalent, adressée au Président de l'Universitat Pompeu Fabra, Plaça de la Mercè, 10-12, 08002 Barcelona

Pour que les effets de cette autorisation aient lieu,

je signe ce document à..... (ville), le ..... 20...

.....

(Signature)

Annexe 6 : Table des matières provisoire Delais e*t al.* à paraître

Delais-Roussarie, E., Avanzi, M. & Herment, S. (eds.), (à paraître), *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*, New York : Springer Verlag.

Introduction to the thematic and presentation of the volume (Elisabeth DELAIS-ROUSSARIE, Mathieu AVANZI & Sophie HERMENT)

PART 1: Prosodic characteristics of language spoken in contact situation

Chap. 1: Markedness considerations in prosodic focus marking, Sabine ZERBIAN (Universität Potsdam, Germany)

Chap. 2: Prosody in African French: Transfer and Typological Variation, Guri BORDAL (University of Oslo, Norway)

Chap. 3: The intonation of Malay speakers of English, Ulrike GUT (Universität Münster, Germany) & Stefanie PILLAI (Universiti of Malaya)

Chap. 4: Prosody in language contact: Occitan and French, Raféu SICHEL-BAZIN, Caroline BUTHKE & Trudel MEISENBURG (Universität Osnabrück, Germany)

Chap. 5: Falling yes/no questions in Corsican French and Corsican: evidence for a prosodic transfer, Philippe BOULA DE MAREUIL (LIMISI, CNRS, Paris, France) Albert RILLARD (LIMSI, CNRS, Paris), Iryna Lekha-Lemarchand, Paolo Mairano (GIPSA LAB, Grenoble) & Jean-Pierre Lai (Grenoble)

Chap. 6: Revisiting the French accentuation/intonation syncretism. Evidence from 'Français Fédéral', Mathieu AVANZI (Université de Neuchâtel, Suisse) Sandra SCHWAB (Université de Genève, Suisse) & Pauline DUBOSSON (Université de Neuchâtel, Suisse)

Chap. 7: Socio-phonetic variation in a language-contact situation: the case of Welsh and Welsh English realisation of stress, Ineke MENNEN (Bangor University, UK) & Robert MAYR (Cardiff Metropolitan University, UK)

Chap. 8: Rhythmic properties of a contact variety: Comparing read and spontaneous speech in Argentinean Porteño Spanish, Christoph GABRIEL & Elena KIREVA (universität Hamburg, Germany)

Chap. 9: You're not from around here, are you? - A dialect identification experiment with speakers of British and Indian English, Robert FUCHS (Westfälische Wilhelms-Universität Münster, Germany).

Chap. 10: Title TBA, Sam HELLMUTH (University of York, UK)

PART 2: Prosody and languages in contact: specific processes such as L2 acquisition and L1 attrition

Chap. 1: Theoretical and practical aspects of studying cross-language and L2 intonation, Ineke MENNEN (Bangor University, UK)

Chap. 2: Language Attrition and Articulation-based Account of Tonal Change in Hailu Hakka, Chia-Hsin YEH & Chi-Jui LU (Michigan State University, USA)

Chap. 3: An Investigation of Prosodic Features in the German Speech of Chinese Learners Hongwei DING (Tongji University, Shangaï, China) & Rüdiger HOFFMANN (TU Dresden, Germany)

Chap. 4: The acquisition of question intonation by Mexican Spanish learners of French, Fabian SANTIAGO-VARGAS & Elisabeth DELAIS-ROUSSARIE (LLF, Université Paris-Diderot, France)

Chap. 5: Tba. Elaine SCHMIDT & Brechtje POST

ANNEXES

Symbols used (IPA, etc.)

Index

Annexe 7 : Exemples sonores du document de synthèse

Le tableau ci-dessous indique de quel corpus et de quel fichier du corpus chaque exemple sonore du document de synthèse est extrait : « Question Time » et « BBC » sont des émissions de télévision et de radio respectivement sur lesquelles j'ai travaillé pour ma thèse (voir Herment-Dujardin 2001).

N° ex. sonore	*Corpus*	*Fichier*
fs 1	AixOx	E_F_B_037-O0
fs 2	AixOx	FRENB_F03_T08
fs 3	AixOx	FRENB_F01_T01
fs 4	AixOx	E_F_H033-P0
fs 5	AixMarsec	A0101B
fs 6	AixOx	FRFR_F02_T15
fs 7	AixMarsec	BO101B
fs 8	AixMarsec	B0302B
fs 9	AixMarsec	A703G
fs 10	Question Time	P1.9S04
fs 11	Féron 2013	fichier audo n° 81
fs 12	AixMarsec	A101B
fs 13	PAC Donegal	Helen
fs 14	PAC Donegal	Magella
fs 15	AixMarsec	A0101B
fs 16	COEC	P1.2S07
fs 17	AixMarsec	J614B
fs 18	AixMarsec	J 608G
fs 19	Wells 2006	ex_Wells
fs 20	ICE-GB	s1b-046F010
fs 21	AixNonce	TC
fs 22	COEC	P2.5S03

fs 23	BBC	P3.2S03
fs 24	COEC	P2.4S01
fs 25	BBC	P3.1S06
fs 26	COEC	P2.4S03
fs 27	Question Time	P1.1S03
fs 28	BBC	P3.1S03
fs 29	COEC	P2.7S04
fs 30		P2.7S04Manip
fs 31	ICE-GB	s1a_058F171
fs 32	ICE-GB	s1b-041F049
fs 33	ICE-GB	s1a-002F120.
fs 34	ICE-GB	s1a-086 F219
fs 35	ICE-GB	s1b-016F123
fs 36	ICE-GB	s1b-047F117
fs 37	ICE-GB	s1a-091F347
fs 38	ICE-GB	s1a-002F006
fs 39	ICE-GB	s2a-011F091
fs 40	ICE-GB	s2b-033F025
fs 41	ICE-GB	s1a-033F179
fs 42	ICE-GB	s1b-027F131
fs 43	ICE-GB	s1b-009F151
fs 44	ICE-GB	s1a-041F030
fs 45	ICE-GB	s1a-035F014
fs 46	AixMarsec	F404 G
fs 47	AixMarsec	J103G
fs 48	ICE-GB	s1b-030F111
fs 49	ICE-GB	s1a-002F147
fs 50	ICE-GB	s1b-024F055
fs 51	ICE-GB	s1a-067F048
fs 52	ICE-GB	s2b-038F023
fz 53	ICE-GB	s1a-056F040
fs 54	ICE-GB	s2b-032F063

fs 55	ICE-GB	s1a-047F262
fs 56	ICE-GB	s1b-015F193
fs 57	ICE-GB	s2a-029F062
fs 58	ICE-GB	s1a-078F030
fs 59	ICE-GB	s1b-066F190
fs 60	ICE-GB	s1a57F283
fs 61	ICE-GB	s1a65F273
fs 62	ICE-GB	s1a30F148
fs 63	ICE-GB	1a94F135
fs 64	ICE-GB	s1a085F127
fs 65	ICE-GB	P1.1S07
fs 66	AixMarsec	F0304G
fs 67	AixMarsec	J607G
fs 68	Féron 2013	fichier audio n°119
fs 69	Rodrigues Da Mota 2010	B_HRT_scat,
fs 70	Rodrigues Da Mota 2010	B_Q_september
fs 71	Rodrigues Da Mota 2010	K_HRT_semester,
fs 72	Rodrigues Da Mota 2010	K_Q_plan,
fs 73	IViE Bradford	maptask

Annexe 8 : Contenu du CD-ROM

Le CD-ROM qui accompagne ce dossier d'habilitation à diriger des recherches contient trois dossiers correspondant aux trois volumes de l'habilitation :

- Curriculum Vitae,
- Volume I synthèse,
- Volume II travaux.

Dans chaque dossier, le fichier au format PDF du volume est disponible. Dans les deux derniers dossiers, les fichiers son auxquels les volumes font référence sont disponibles. Ils sont listés en annexe 7 pour le volume I. Dans le volume II, trois articles ont été publiés dans des revues électroniques permettant d'intégrer du son. Les fichiers son pour ces trois articles se trouvent dans le dossier « Volume II travaux » du CD.

La fenêtre ci-dessous donne un aperçu du contenu du CD-ROM :

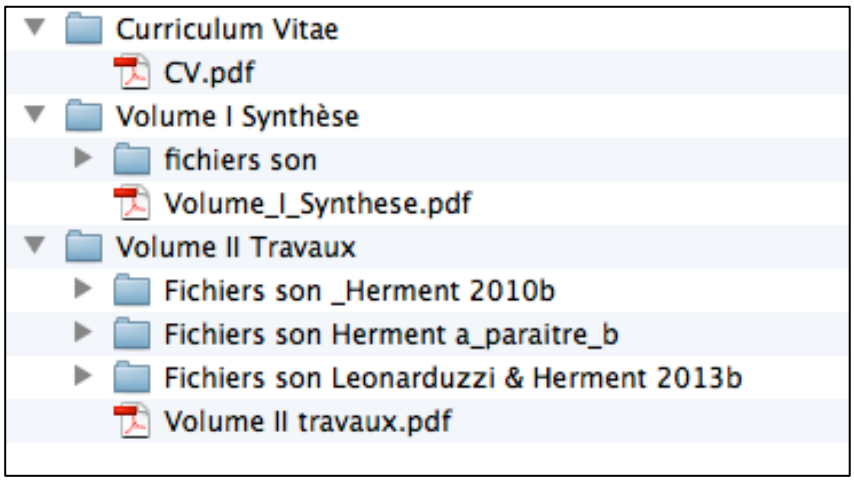


Figure 38 : contenu du CD-ROM