UNIVERSITÉ DE POITIERS

Faculté des Lettres et Langues

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE POITIERS

LES MOTS COMPOSÉS :

ANALYSE DE SCHÉMAS ACCENTUELS DE L'ANGLAIS BRITANNIQUE STANDARD

en deux volumes, accompagnée d'un CD-ROM : VOLUME 1

présentée et soutenue publiquement par

Susan MAUROUX

le 14 décembre 2002

sous la direction de

Monsieur Jean-Louis DUCHET

Professeur à l'université de POITIERS

JURY

Monsieur Nicolas BALLIER, Maître de conférences à l'université de Rouen

Monsieur Jean-Louis DUCHET, Professeur à l'université de Poitiers

Madame Ruth HUART, Professeur à l'université de Lille III

Monsieur Michel PAILLARD, Professeur à l'université de Poitiers

Monsieur Peter ROACH, Professeur à l'université de Reading, Angleterre

AVERTISSEMENT

LA QUALITÉ DE CET EXEMPLAIRE DÉPENDANT DE L'ÉTAT GÉNÉRAL, L'A.N.R.T. NE PEUT GARANTIR UN RÉSULTAT IRRÉPROCHABLE.

Au passé, toujours présents

Mes parents

Au présent et à l'avenir

Pierre

Julie Anna Charlotte

REMERCIEMENTS

Je voudrais tout d'abord remercier Jean-Louis Duchet pour sa direction, pour le temps qu'il m'a consacré. Ses connaissances et son enthousiasme pour le domaine ont toujours rendu les séances de travail fascinantes et instructives.

Je dois aussi un grand merci à tous ceux et toutes celles qui m'ont aidé sur le plan matériel : Christine pour le don de sa 12ème édition du "Jones" (et merci aussi pour ton soutien !) ; les collègues qui m'ont prêté des livres et repéré des références ; Oumar pour avoir scannérisé l'œuvre littéraire du corpus ; Abel pour le temps qu'il a consacré à m'initier à ACCESS ; Nicolas pour la numérisation d'une bonne partie du corpus oral, et pour avoir préparé le CDROM qui accompagne ce volume.

Je suis également très reconnaissante à toutes les autres personnes qui m'ont conseillée et aidée en me faisant partager leur savoir : je voudrais n'oublier personne même si je ne peux pas vous nommer tous. Je remercie ma famille, mes amis et collègues pour l'intérêt qu'ils ont montré pour ce travail et pour les mots sympathiques qui m'ont soutenue dans les passages difficiles, plus particulièrement à Christine et à Anne pour leurs encouragements constants.

Merci aussi à mes relectrices, Catherine et Elisabeth, qui ont eu la patience de corriger mes erreurs et étourderies. Leurs remarques et leurs corrections m'ont été précieuses, et bien plus efficaces que le correcteur automatique ! Bien entendu, j'assume la reponsabilité de toutes les erreurs qui ont subsisté, malgré leur vigilance.

L'attribution d'un aménagement de mon service d'enseignement en tant que PRAG à partir de la troisième année de la thèse m'a permis de terminer ce travail dans un délai raisonnable et dans de bonnes conditions. Merci aux membres de la Commission de la Recherche, du Conseil de Gestion et du Conseil Scientifique de l'université de Limoges de me l'avoir accordé.

Je termine avec la formule conventionnelle, mais si vraie « *Last but not least* ». Merci à toi, Pierre, d'avoir toujours cru en moi, et de m'avoir aidée à aller jusqu'au bout. Merci à Julie, Anna et Charlotte de votre soutien, et d'avoir supporté une maman beaucoup moins disponible qu'elle ne l'aurait voulu.

SOMMAIRE

Remerciements 3

Sommaire 5

Abréviations et symboles 6

Introduction 7

Chapitre 1 14

Chapitre 2 99

Annexes du chapitre 2 234

Chapitre 3 292

Annexes du chapitre 3 356

Chapitre 4 366

Annexes du chapitre 4 530

Conclusion 588

Glossaire des termes 600

Bibliographie thématique 602

Références bibliographiques 616

Index des notions 628

Index des auteurs 633

Table des matières 636

Table des figures 645

Contenu du CDROM 648

Liste des abréviations et symboles

Abréviations	Correspondance
« d »	Accentuation double
« s »	Accentuation simple
CSR	*Compound Stress Rule*
EPD	*English Pronouncing Dictionary* (Jones, 1963, 1967, 1977 & 1997)
EPD12	12ème édition
EPD13	13ème édition
EPD14	14ème édition
EPD15	15ème édition
F_0	Formant zéro, représentant la fréquence fondamentale
I	Intensité
IPA	International Phonetics Association (Alphabet)
LPD	*Longman Pronouncing Dictionary* (Wells, 1990 & 2000)
LPD90	Edition parue en 1990
LPD00	Edition parue en 2000
NPD	*New Pronouncing Dictionary* (http://www.mshs.univ-poitiers.fr/Forell/Guierre/index.htm,
NSR	*Nuclear Stress Rule*
SAMPA	Speech Assessment Methods Phonetic Alphabet (codage de Wells (1992 & 1995) compatible avec tout support informatique)
SPE	*Sound Pattern of English* (Chomsky & Halle, 1968)

Symboles	
#	Borne morphologique de frontière entre deux éléments de composition
{ }	Parenthèses autour de séquences relevant des requêtes (MPW et ACCESS, voir Glossaire).
/1230/	Le schéma accentuel où /1/ est l'accent primaire, /2/ l'accent secondaire, /3/ l'accent tertiaire et /0/ la syllabe non accentuée (voir Glossaire).
/æ/ ou [æ]	Transcription phonétique utilisant les symboles IPA. Nous employons les crochets lorsqu'il s'agit de réalisations effectives, et une transcription large, conformément aux indications de LPD. Les barres obliques / / sont utilisées pour représenter des séquences de phonèmes, hors discours.
*	Généralement, ce symbole est utilisé pour désigner « 0 ou n caractères » dans une expression régulière.

INTRODUCTION

Introduction

Cette thèse s'inscrit dans le cadre de la recherche sur l'anglais oral. Le domaine d'étude que nous avons choisi est très vaste, et d'autres chercheurs, notamment Boisson (1980) et Huart (1984, 1989), l'ont déjà exploré ; nous ferons souvent référence à leurs travaux. Toutefois, si le sujet n'est pas neuf, notre façon de l'aborder (travail sur corpus, utilisation d'outils informatiques récemment mis au point) a l'ambition de l'être. Ainsi, notre analyse apportera peut-être un éclairage nouveau sur les hypothèses qui ont déjà été formulées ; dans tous les cas, elle permettra de les vérifier en utilisant des méthodes nouvelles.

L'accentuation de l'anglais est notoirement difficile à maîtriser, et *a fortiori* l'accentuation des composés. Toutefois, il existe des règles et nos objectifs concordent avec ceux qu'annonce Duchet dans son introduction au *Code de l'anglais oral* (1994 : 1) : « Il existe, plus qu'on ne l'imagine, des règles qui régissent la prononciation des mots anglais. C'est à la recherche systématique de régularités, et à leur formulation claire et complète, que vise notre démarche ».

Nous avons pris conscience des difficultés que rencontrent les non natifs, et plus particulièrement les francophones, lors de notre enseignement. Une mauvaise accentuation des composés peut simplement produire sur l'auditeur une impression un peu bizarre qui trahit le non-anglophone ; mais elle peut aussi entraîner un véritable problème de compréhension. Même les locuteurs non natifs les plus compétents s'y trompent, et c'est la raison pour laquelle nous avons voulu approfondir ce sujet. Comment faire pour que cette compétence soit fondée sur une connaissance explicite des règles en jeu ?

A partir de ce souci d'ordre pédagogique, nous avons étudié un domaine qui a des enjeux pratiques, certes, mais qui implique aussi des considérations théoriques importantes. Notre approche se situe à la croisée des chemins : une partie de nos analyses relève de l'empirique car le corpus comprend des extraits sonores de discours ; l'autre partie relève d'une analyse plus traditionnelle de l'accent lexical tel qu'il est prévu dans les dictionnaires de prononciation. Boisson (1984 : 175) insiste sur la complexité de l'accentuation des composés et conclut : « Compte tenu de ces complexités, il me semble qu'il reste encore beaucoup de travail à faire sur les composés anglais. [...] l'analyse du comportement effectif des usagers de la langue montrera dans quelle mesure les caractérisations accentuelles des dictionnaires ne sont pas illusoires ». C'est pourquoi notre démarche consiste en la confrontation systématique des prévisions du dictionnaire et des réalisations effectives en discours, en utilisant des moyens informatiques de traitement automatique des données. Par ailleurs, cette étude nous permettra d'émettre des hypothèses concernant les extraits sonores non lexicalisés, donc absents des dictionnaires.

Dans le domaine de l'accentuation, il existe divers courants théoriques correspondant à des méthodologies différentes. Dans le premier chapitre, nous avons voulu confronter ces différentes approches afin d'évaluer leur apport en ce qui concerne l'accentuation, plus particulièrement dans les composés. Nous aborderons donc les problèmes de la définition de l'accent, en termes de ses corrélats acoustiques, qui ont toute leur pertinence pour l'analyse des extraits sonores, mais aussi par rapport à sa fonction dans l'organisation du discours, et la hiérarchie accentuelle, d'une portée plus théorique. Ensuite, nous situerons les

différents courants en ce qui concerne les règles de l'accentuation en général. Enfin, nous ferons état des caractéristiques spécifiques de l'accentuation des composés, et nous étudierons la manière dont on peut en rendre compte, surtout pour ce qui est de la notation de la hiérarchie accentuelle.

Dans le second chapitre, nous expliquerons quels ont été les objectifs et les méthodes de constitution du corpus à partir duquel se feront les analyses. Dans la première partie, nous posons la question de savoir ce que l'on inclut dans le corpus, c'est-à-dire, ce que nous entendons par « composé ». Cette question est épineuse et nous n'avons pas la prétention de l'avoir résolue, mais il est nécessaire de l'aborder afin de s'assurer de la cohérence du contenu du corpus. Ensuite, nous nous intéresserons au choix des sources elles-mêmes et à la manière dont les données sont extraites, d'abord pour les sources écrites puis pour les sources orales.

Comme nous l'avons expliqué, nous tenons à confronter les prévisions de l'accent lexical et les réalisations d'unités en discours. Les données des sources écrites correspondent donc aux entrées de dictionnaires lexico-phonétiques, notamment la version électronique du *Longman Pronouncing Dictionary* dont nous disposons au laboratoire FORELL, ainsi que le *New Pronouncing Dictionary* de Guierre, et une version électronique du *English Pronouncing Dictionary* comportant les entrées à trait d'union de la 12ème édition, également rassemblées par Guierre. Pour ce qui est des sources orales, nous avons décidé de constituer un corpus d'extraits sonores (appelé corpus oral d'analyse) provenant de domaines divers, et constitué d'unités lexicalisées figurant dans le dictionnaire ainsi que d'unités non lexicalisées (qui en sont donc absentes). Ces dernières unités nous

permettront d'évaluer les règles par rapport à la prévision de l'accent pour des entrées inconnues. Cet objectif nous a conduite à choisir deux types de sources orales : des documentaires de la radio et de la télévision ; et l'enregistrement d'une œuvre littéraire, devenue célèbre à présent, et qui est riche en néologismes, *Harry Potter and the Philosopher's Stone*.

La constitution d'un corpus nécessite des explications détaillées, parfois fastidieuses. La sélection des entrées des dictionnaires s'est faite en partie à partir de traitements automatiques qui permettront de tester des hypothèses concernant le processus de la composition, et cette démarche est riche d'enseignements : elle fera l'objet de la deuxième partie du chapitre 2. Dans la troisième partie, nous rendrons compte de la sélection des extraits sonores, toujours avec beaucoup de précisions, car la validité de nos résultats dépend du bien-fondé de la sélection des données à analyser et de la qualité sonore des extraits du discours.

Dans le troisième chapitre, nous expliquerons quelles seront les méthodes d'analyse de ces deux types de sources. Afin de comprendre les enjeux de l'analyse acoustique et d'être en mesure d'en tirer des conclusions pertinentes, nous étudierons en détail les différents corrélats que l'on associe à l'accent, et nous déterminerons, grâce à un échantillonnage d'extraits sonores, quels sont les outils les plus efficaces pour les interpréter. Ensuite, nous passerons à l'étude de la base de données que nous avons créée pour analyser l'ensemble des données. Les explications seront toujours très précises car nous travaillons à partir de traitements automatiques dont l'intérêt est fonction de la qualité et de la fiabilité des informations disponibles.

Enfin, le dernier chapitre comprendra les résultats de l'analyse du corpus qui nous permettront de tirer des conclusions sur les règles de l'accentuation des composés. Nous mettrons à l'épreuve les règles déjà énoncées et, là où c'est possible, nous formulerons quelques autres hypothèses. Cette étude s'appuiera sur les requêtes formulées à partir de la base de données, et elle mettra aussi en œuvre l'analyse acoustique des extraits sonores. Nous tirerons également des conclusions sur des prévisions de l'accent pour les unités non lexicalisées.

Nous avons volontairement limité notre étude à l'anglais britannique standard pour deux raisons. Premièrement, notre compétence est plus grande pour cette variété de l'anglais et une partie des analyses repose sur une appréciation de l'anglophone (nous-même). Deuxièmement, le sujet est déjà très vaste car le domaine de l'accentuation est complexe, tout comme celui de la composition. Une étude de la variation nécessiterait des sources orales plus variées et l'importance du travail mérite de faire l'objet d'une étude spécifique. C'est pourquoi nous avons préféré cibler nos analyses, tout en faisant référence à la variation dans nos conclusions.

Bien que nous n'ayons exclu aucune catégorie syntaxique, nos exemples sont souvent des composés nominaux. Ceci s'explique par le fait qu'il y a beaucoup plus de composés qui appartiennent à cette catégorie qu'à d'autres, et qu'elle est représentative des difficultés que nous pouvons rencontrer[1].

---

[1] Nous faisons une étude de l'accentuation selon la catégorie syntaxique dans le chapitre 4.

L'organisation des annexes, ainsi que le choix de les inclure à la fin de chaque chapitre concerné, correspond à notre souci de les intégrer au cœur de ces travaux. En effet, les indications de la programmation (qui permet l'extraction des données à partir de sources électroniques), par exemple, font partie de nos méthodes d'analyse. Nous avons pris la décision de les numéroter par ordre alphabétique, en recommençant la numérotation pour chaque nouveau chapitre.

Nous devons insister sur la nécessité d'une démarche qui peut sembler laborieuse car les outils d'analyse que nous avons choisis ne sont efficaces que s'ils sont maniés avec toutes les précautions requises. Nous devons à nos lecteurs des explications concernant ces précautions afin qu'ils puissent évaluer la portée des résultats obtenus. Ces précisions sont aussi importantes pour comprendre la démarche elle-même.

Nous avons également souhaité confronter des théories différentes, que ce soit en phonologie, en lexicologie ou en phonétique acoustique, afin de tirer le maximum de chacune d'entre elles et d'évaluer leur apport pour notre sujet. Ce sont les théories de la phonologie que nous allons aborder dans le premier chapitre qui traite de l'accentuation générale.

CHAPITRE 1

Chapitre 1 : L'accentuation et les composés dans les travaux phonologiques relatifs à l'anglais

Dans ce chapitre, nous allons étudier les différentes théories sur l'accentuation et situer notre travail par rapport à ces approches. Si les chercheurs s'accordent pour admettre l'importance du phénomène de l'accentuation dans la langue anglaise, elle n'a pas la même portée pour tous : certains lui donnent une place de choix dans leur théorie ; chez d'autres elle occupe une position moins centrale.

Gimson (1980) fait partie de ceux pour qui l'accent est au cœur du système de prononciation de l'anglais. Dans son ouvrage très complet, *An Introduction to the Pronunciation of English* (1980 : 318), il s'exprime clairement sur ce point.

> The stress-timed rhythm of the English utterance (be it a single word or connected speech) with the related obscuration of weak syllables is the prime distinguishing feature of the language's pronunciation.

C'est aussi le point de départ des travaux de Guierre (1979) qui l'inclut au sein des **contraintes rythmiques**. Pour lui, c'est l'accent qui dicte la prononciation des sons environnants ; il est la première opération dans la construction de la chaîne parlée. Cette approche est reprise par Duchet (1994) dans *Code de l'anglais oral* où il souligne la différence entre le français qui est une langue dont le rythme est dicté par les syllabes (*syllable-timed language*) et l'anglais (*stress-timed language*) où c'est l'accent tonique qui l'emporte. L'accent est la clé du système de prononciation de l'anglais, entraînant la réduction vocalique, et il est central pour la compréhension même de la langue orale.

Roach (1991) ne conteste sûrement pas l'importance que l'accentuation peut avoir pour la compréhension, mais ses propos sont beaucoup plus réservés quant à

l'appellation *stress-timed language*. En effet, dans *English Phonetics and Phonology*, il explique sa position: comme Ladefoged (1993: 249), il n'est nullement convaincu que l'anglais fasse partie de cette famille, mais sa position n'est pas arrêtée parce que ce domaine de recherches n'en est qu'à ses débuts (Roach, 1991 . 123).

> [...] one ought to keep an open mind on the subject, remembering that the large-scale, objective study of suprasegmental aspects of real speech is only just beginning, and there is much research that needs to be done.

Nous reviendrons à cette question dans l'analyse acoustique.

Bien entendu, l'importance que l'on attribue à l'accentuation doit être prise en compte par l'apprenant s'il veut bien prononcer l'anglais. Toutefois, si ces propos sont mal compris, et appliqués de manière trop étroite, il est probable que la prononciation de l'apprenant sera quelque peu caricaturale. Boisson (Boisson, 1980), qui parle de **filtre rythmique** quand il décrit le rythme naturel de l'anglais, est très conscient de ce risque. Dans *L'accentuation des composés en anglais*, il nous offre, avec humour, un portrait correspondant aux dérapages qu'il a dû souvent constater (Boisson, 1980: 102).

> [...] l'accent ne se réduit pas au martèlement des syllabes. Cette conception simplette est partiellement responsable de bizarreries de prononciation chez les francophones apprenant l'anglais. Ceux-ci s'imaginent qu'accentuer correctement l'anglais consiste à danser une espèce de rumba éléphantesque où les syllabes « accentuées » sont martelées *furioso* et obscènement dilatées, sans toutefois que les inaccentuées soient parcourues *allegro*, comme il convient, et sans que les inflexions de la voix et la centralisation des voyelles inaccentuées confèrent l'authenticité du rythme [...]

En effet, le francophone a tendance à jouer sur l'intensité et la longueur, ce qui entraîne cette impression de martèlement. Il ignore que la prononciation de la syllabe inaccentuée est aussi importante. Nous retrouvons les propos de Gimson (1980) qui insiste sur le lien entre les syllabes accentuées et les inaccentuées. Les

syllabes qui ne sont pas proéminentes s'accompagnent très souvent de réduction vocalique, ou bien d'un *tempo* rapide ce qui permet la mise en relief des syllabes accentuées qui alternent avec les inaccentuées.

Le rythme de l'anglais est dicté par cette alternance, qu'elle soit régulière, comme le suggère l'appellation « stress-timed », ou qu'elle soit plus approximative. L'importance de l'accentuation pour la compréhension ainsi que pour l'authenticité de la prononciation est incontestable ; toutefois, il reste à préciser ce que l'on entend par **accent**. Dans la première partie de ce chapitre, nous allons tenter d'en donner une définition, en étudiant quels sont ses corrélats acoustiques. Puis, nous aborderons d'autres aspects de l'accentuation : son rapport avec la compréhension, les différents types d'accent, et la hiérarchie accentuelle. Dans la quatrième partie, nous verrons comment ce phénomène est intégré dans les différentes théories concernant la prononciation, et nous aborderons les règles de l'accentuation qui seront significatives pour notre analyse des composés. Ensuite, les caractéristiques de l'accent dans le discours continu seront brièvement discutées, car nous nous intéressons au comportement en discours, non seulement à l'accent lexical. Enfin, pour clore ce chapitre, nous définirons la nomenclature que nous adopterons dans notre analyse.

1.1 Définition de l'accent

Dans cette première partie, nous étudierons les propos d'auteurs représentatifs de divers courants, qui citent des facteurs différents en ce qui concerne la définition de l'accent. Nous récapitulons ces facteurs dans le tableau (Figure 1.1) ; la liste des auteurs n'est pas exhaustive. Nos objectifs sont de montrer la variété des facteurs dont il faut tenir compte, non de résumer les

définitions de tous les chercheurs qui s'y sont intéressés. Nous avons cité Kingdon (1958b) qui s'est intéressé le premier au phénomène de l'accent ; puis des auteurs d'ouvrages généraux sur la phonologie de l'anglais, (Gimson (1980), Poldauf (1984), Giegerich (1992), Roach (1991 et 1992), Ladefoged (1993), Kager (Goldsmith, 1995), Durand (1998) et Hammond (1999)). Les ouvrages cités sont ou des manuels de phonologie et phonétique, ou traitent de la phonologie anglaise générale, mais ce ne sont pas des ouvrages spécialisés en phonétique acoustique[1] ; nous ferons référence à ces ouvrages dans le chapitre 3. Nous citons également les deux auteurs qui se sont intéressés de manière spécifique à l'accentuation des composés : Boisson (1980) et Huart (1984 et 1989).

1.1.1 L'accent et la proéminence : un ensemble de facteurs

Il est en effet essentiel que l'on arrive à définir ce qu'est l'accent, et pourtant, ce n'est pas une question simple, même si elle est préalable à toute discussion. Il est clair que l'accent tonique est un des phénomènes centraux du système phonologique et phonétique de l'anglais, mais, de l'avis de tous, il résulte non pas d'un seul mais d'un ensemble de facteurs. Il s'agit donc d'un phénomène complexe.

Pour désigner cet ensemble de facteurs, les auteurs ont souvent employé le terme **proéminence** (Boisson, 1980), « prominence » en anglais. Dans le livre de Kingdon (1958b : ix), *The Groundwork of English Stress*, qui est paru dès 1958,

---

[1] Dans la bibliographie, certains ouvrages de Ladefoged se trouvent dans la section « Manuels de phonologie et phonétique de l'anglais », d'autres en « Phonétique acoustique ». Nous faisons référence ici aux ouvrages généraux.

l'auteur donne une définition du mot *stress* (accent) où l'on retrouve ce terme «Stress. The force employed in uttering a syllable, giving it a certain degree of prominence ». Notons que la notion de **force** est également mentionnée. Elle correspond ici à la façon dont le locuteur va produire l'accent. Roach (1992 : 86) fournit une définition du terme **proéminence.**

> [...] experiments have shown that prominence is associated with greater length, greater loudness, pitch prominence (i.e., having a pitch level or movement that makes a syllable stand out from its context) and with 'full' vowels and diphthongs [...]. Despite the complexity of this set of interrelated factors, it seems that the listener simply hears syllables as more prominent or less prominent.

Nous retrouvons la notion de force dans la définition, mais dans le sens perceptuel *loudness*. En effet, *force* en anglais est souvent employé pour désigner l'intensité avec laquelle le locuteur produit son énoncé, par opposition à la notion perceptuelle de *loudness*. Il semble pourtant qu'il y a une certaine confusion dans l'emploi de ces termes. Ainsi, Hammond (1999 : 151) utilise le concept de *loudness* en se référant à la production :

> In the production domain, it is generally thought that prominence is a function of greater duration, loudness or pitch. Hence stressed syllables would be longer, louder or higher in pitch than unstressed syllables.

Comme le mot **force** recouvre les deux notions en français, il est nécessaire de déterminer une nomenclature claire, et de s'y tenir. Nous employons le terme **intensité** quand il s'agit de la perception, qui se mesure en décibels, et **force** quand il s'agit de l'énergie musculaire de la force de production.

La définition associe d'autres éléments puisque la description correspond plus à la manière dont l'accent est perçu qu'à la manière dont il est produit. Ces autres facteurs associés à la proéminence sont d'ailleurs nombreux : la longueur, le contexte sonore et la hauteur mélodique. Dans cette définition, il faut retenir

l'insistance de Roach sur la façon simple dont l'accent est perçu, par opposition à la complexité physique du phénomène. Cette idée est reprise quelques pages plus loin dans la définition du mot *stress* (Roach, 1992 : 103) :

> It seems likely that stressed syllables are produced with greater effort than unstressed [...]. These effects of stress produce, in turn, various audible results. One is pitch prominence, in which the stressed syllable stands out from its context. [...] Another effect of stress is that stressed syllables tend to be longer – this is very noticeable in English, less so in some other languages. Also, stressed syllables tend to be louder than unstressed, though experiments have shown that differences in loudness alone are not very noticeable in most listeners.

En lisant ces extraits, l'on commence à appréhender la complexité du phénomène car Roach nous parle tantôt de la façon dont on produit l'accent, tantôt de la façon dont on le perçoit – et ce n'est justement pas la même chose. L'on remarque aussi qu'il y a des tendances, mais pas toujours des constantes ni des certitudes. Dans un autre ouvrage (Roach, 1991 : 86), il résume ainsi ce qu'est la proéminence. «Prominence, then, is produced by four main factors : (i) loudness, (ii) length, (iii) pitch and (iv) quality ».

Tentons donc de résumer en quoi consiste la proéminence. Ce phénomène semblerait être plutôt une notion psychologique, résultant de facteurs physiques reliés entre eux de façon complexe. Il est particulièrement utile lorsque l'on veut faire référence à la manière dont l'auditeur perçoit l'accent. Par contre, la mesure objective est difficile à opérer, d'où le désaccord entre experts sur le nombre de degrés d'accent à déterminer. Nous devrons en tenir compte dans la discussion à venir. Mais avant de se lancer dans ce débat, nous voulons étudier les corrélats acoustiques et articulatoires qui aboutissent à la proéminence et voir comment ils peuvent nous aider à répondre à la question : comment définir l'accent ?

1.1.2 Les facteurs à prendre en compte pour une définition de l'accent

Sous cette rubrique, nous allons étudier les corrélats acoustiques qui sont associés à la proéminence et déterminer leur poids relatif dans le phénomène appelé accent. Nous commencerons par l'intensité et la force, puis la longueur, la qualité vocalique, la hauteur mélodique, et enfin la fo[illegible]e relative des syllabes.

1.1.2.1 Intensité et force (perception et production)

Ces deux caractéristiques de l'accent se situent à deux niveaux différents que l'on oppose souvent : l'intensité correspond à la perception alors que la force correspond à la production. Certains auteurs privilégient l'une par rapport à l'autre mais beaucoup citent les deux (voir Figure 1.1). Ladefoged (1993 : 14) explique le phénomène physique et les corrélats acoustiques qui en résultent :

> Variations in stress are caused by an increase in the activity of the respiratory muscles (so that a greater amount of air is pushed out of the lungs) and in the activity of the laryngeal muscles (so that there is a significant change in pitch).

Nous pouvons nous demander dans quelle mesure ces deux pôles sont distincts. Si Ladefoged explique de manière détaillée quels sont les corrélats articulatoires de l'accent, il a conscience de l'interférence entre la perception et la production (Ladefoged, 1993 : 113) et se réfère à la théorie motrice de la perception de la parole (TMPP) :

> It seems as if listeners sometimes perceive an utterance by reference to their own motor activities. When we listen to speech, we may be considering, in some way, what we would have to do in order to make similar sounds. This notion is called the motor theory of speech perception.

Il apparaît clairement que ces deux pôles ne doivent pas être considérés comme étant distincts, bien au contraire; ils sont complémentaires. Durand (1998 : 64), qui évoque le débat articulation / audition, affirme : « Tout chercheur qui

examine les travaux récents sans s'engager dans un modele psycholinguistique particulier constatera que le débat articulation/audition n'a pas connu de resolution. » et il conclut (1998 : 66):

> Puisque aucun système de traits n'a pu s'imposer sur une base seulement articulatoire ou acoustique, force est de constater que ces systèmes portent en leur sein les traces de la production et de la perception du langage. Affirmer que les traits phonologiques sont neutres quant à ces deux dimensions comme le fait Durand (1990), c'est en definitive reconnaître l'équivalence et la complémentarité du moteur et du perceptuel.

Ses conclusions s'appliquent au système phonologique dans son ensemble, mais au vu de l'énorme complexité que nous avons retrouvée dans les définitions de l'accent, elles revêtent un intérêt particulier. Retenons de ce débat que les deux pôles de la perception et de la production sont en complémentarité, non en opposition. Ainsi, nous devons tenir compte des deux pour effectuer une analyse pertinente.

Pour récapituler ce qui a été dit précédemment, voici un tableau qui reprend les différents éléments que les auteurs utilisent pour définir l'accent. Nous n'avons inclus dans le tableau que ceux qui ont évoqué de façon explicite la définition de l'accent. Y figurent également des auteurs auxquels nous n'avons pas encore fait référence.

Figure 1.1 : Récapitulatif des paramètres de l'accent

Paramètre / Référence	Proéminence	Intensité (perception auditive)	Force (de production)	Longueur	Qualité vocalique	Hauteur mélodique	Relativité
Kingdon 1958b	**OUI**		OUI			OUI	
Boisson 1980	**OUI**				OUI	OUI	
Gimson 1980	**OUI**	OUI	OUI	OUI	OUI	**OUI**	
Huart 1984 et 1989	**OUI**					**OUI**	
Poldauf 1984		OUI					**OUI**
Roach 1991 et 1992	**OUI**	OUI	OUI	OUI	OUI	OUI	
Giegerich 1992							**OUI**
Ladefoged 1993	OUI	OUI	OUI	OUI	NON	OUI	OUI
Kager 1995	OUI	OUI		OUI		OUI	
Hammond 1999	OUI		OUI	OUI		OUI	

Les éléments considérés comme étant les plus caractéristiques sont en caractères gras.

1.1.2.2 Longueur

Nous constatons que plusieurs éléments sont souvent cités, mais ne figurent jamais comme étant l'élément clé pour caractériser l'accent. C'est le cas de la longueur qui est un des corrélats acoustiques généralement reconnus. En effet, si nous mesurons la durée de la syllabe accentuée, elle dure plus longtemps que la même séquence sonore dans une syllabe non accentuée Ce phénomène est relié au rythme du discours et est partiellement responsable de la définition de l'anglais comme étant « stress-timed » (Gimson, 1980, 318). Une interprétation stricte de ce terme nous obligerait à une alternance régulière des syllabes accentuées et non

accentuées ce qui n'est manifestement pas le cas, mais il n'en reste pas moins que cette langue repose sur un rythme où des syllabes fortes sont rarement juxtaposées, et où il existe une certaine alternance. Le fait de passer moins de temps sur les syllabes non accentuées en augmentant la durée des syllabes accentuées renforce cette impression.

1.1.2.3 Qualité vocalique

La syllabe non accentuée subit le plus souvent non seulement une réduction de longueur mais aussi une réduction de la qualité vocalique (pour faire écho à Roach, cité en 1.1.1). Pour certains auteurs, ce processus est indépendant du processus de l'accent, pour d'autres il y est étroitement relié. Nous y reviendrons quand nous étudierons les différentes théories de l'accent. Tous s'accordent pour dire que la réduction vocalique ne peut se produire que sur une syllabe non accentuée. Ce changement de la qualité vocalique affecte le rythme du discours, comme le fait la réduction de la durée d'une syllabe. La présence d'une voyelle réduite est nécessairement la marque d'une syllabe non accentuée, mais nous ne pouvons pas dire l'inverse : une syllabe non accentuée peut retenir sa qualité vocalique pleine (bien que, dans certaines théories, une syllabe ayant une qualité vocalique pleine se voie attribuer un certain degré d'accent : nous y reviendrons). Ce n'est donc pas un facteur suffisant pour déterminer l'accent, mais il y contribue. Comme le dit Kager (Goldsmith, 1995 : 367) quand il définit l'accent : « [...] stress contrasts tend to be enhanced segmentally: stressed syllables may be strengthened by vowel lengthening or gemination, while stressless syllables may be weakened by vowel reduction ». La mise en relief de la syllabe accentuée peut donc être renforcée par la réduction vocalique.

1.1.2.4 Hauteur mélodique

Kager (Goldsmith, 1995 : 367), tout en attribuant à la proéminence une grande importance, a conscience de la difficulté qui existe à mesurer ce phénomène plutôt psychologique. Pour lui, les facteurs physiques mesurables que nous avons étudiés plus haut (longueur et intensité) dépendent [g]rande partie de la hauteur mélodique qui prend une place privilégiée dans la définition de l'accent :

> Although the mental reality of prominence is undisputed, an unambiguous phonetic correlate has not yet been discovered. Prominent syllables are potentially capable of bearing pitch movements with a strong perceptual load. They also tend to be of longer duration, as well as of higher intensity, but both of the latter factors are usually subordinated to pitch.

Nous l'avons vu dans le tableau, presque tous les auteurs font référence à la hauteur mélodique, mais ce n'est pour la plupart qu'un élément parmi d'autres qui contribue à la proéminence, caractéristique principale à leurs yeux du phénomène de l'accent. Huart (1984 et 1989) va plus loin. Pour elle, ce n'est pas la proéminence mais plutôt la hauteur mélodique qui est l'élément clé de l'analyse de l'accent.

Avant d'étudier son argumentation sur cet aspect de l'accentuation, nous voulons rappeler deux points essentiels concernant la hauteur mélodique, qui est la marque de l'intonation :

1/ Tout énoncé a une intonation

2/ Un mot isolé est énoncé : il a une intonation ; elle porte sur la syllabe accentuée.

Un mot composé a par conséquent une intonation. L'importance que Huart (1984 : 72) lui attribue fait que, dans son analyse, elle n'emploie pas le terme d'**accent primaire** qui est le terme conventionnel, mais lui préfère l'appellation **chute mélodique**. Ce parti pris est particulièrement intéressant pour nous car son étude porte sur les agrégats nominaux, dont les composés ; elle s'intéresse surtout au placement de l'accent sur le premier ou le second terme et l'élément accentué est celui qui porte la chute mélodique. Cette appellation présente bien des avantages pour ce genre d'analyse et nous reviendrons à ce terme ultérieurement dans la définition de la nomenclature que nous allons employer.

D'après le tableau, nous constatons que ce phénomène a été remarqué par d'autres, même si ce n'est pas le point central de leur description de l'accent. Boisson (1980 : 821), par exemple, qui a également étudié les composés, fait une analyse approfondie de ce qui constitue l'accent. Il aboutit à un ensemble fort complexe que nous avons résumé sous l'appellation **proéminence** et dans son analyse il n'utilise pas la notion de chute mélodique. Néanmoins, il n'écarte pas l'influence de l'intonation.

> A ces complexités de l'accent anglais s'ajoutent les problèmes soulevés par l'interaction entre accentuation et intonation, qu'il est impossible de passer complètement sous silence même si l'on se restreint à une étude de l'accent.

D'autres auteurs utilisent l'intonation pour décrire un certain type d'accent. Kingdon (1958b : ix) attribue une importance capitale à ce phénomène – ce qui correspond à ses préoccupations puisqu'il est l'auteur de *The Groundwork of English Intonation* (1958a). Pour lui, les deux notions sont reliées et il utilise l'appellation « *kinetic stress* » pour désigner l'accent qui est appelé conventionnellement **accent primaire**. Il explique le terme en faisant référence à

l'intonation. « [...] pitch of voice varies during whole duration of tone». Nous reviendrons sur ces appellations dans la discussion de la hiérarchie accentuelle.

Quant à Gimson (1980 : 225), dans sa définition de l'accent, il parle de *pitch change* et de *potential change of pitch* lorsqu'il décrit ce même accent. Pour lui, comme pour Huart, c'est un élément clé de la perception de l'accent qu'il met au même niveau que la proéminence. Encore une fois, nous constatons que la hauteur mélodique est au centre de la production et de la perception de l'accent.

1.1.2.5 Force relative des syllabes

Ladefoged (1993 : 113) explique d'abord toute la complexité du phénomène physique de l'accent, mais conclut de la manière suivante : « Stress can always be defined in terms of something a speaker does in one part of an utterance relative to another ». Ainsi, l'accent peut se définir en termes de force relative.

Nous retrouvons cette même idée dans Giegerich (1992 : 179) qui insiste sur le fait que les facteurs physiques complexes, auxquels les autres auteurs font référence et qui contribuent à l'accent, sont difficilement mesurables de manière absolue. Ses propos sont assez catégoriques : « It is not possible to quantify any of the physical correlates of stress in absolute terms. A syllable's property of being stressed is primarily a relative one ».

De même, Poldauf (1984 : 13) met ce phénomène au centre de son analyse de l'accent. Il apporte un éclairage nouveau en insistant sur la relativité des syllabes les unes par rapport aux autres. Dans *English Word Stress*, il explique sa théorie en ce qui concerne l'accent de mot :

> Word stress is the relative position of the force peak or peaks in a word [...] opposed to other positions the force peak or peaks could occupy in a word in consideration of the number of its syllables. Hence a monosyllable cannot have word stress.

Cette définition de l'accent en termes de force relative des syllabes les unes par rapport aux autres est très pertinente pour nos travaux. Il y a une sorte d'opposition entre syllabe accentuée et syllabe non accentuée ; elles se définissent mutuellement par contraste. L'accent n'a d'importance que parce que la syllabe accentuée est plus perçue que d'autres. Ceci a une grande signification quand on étudie les composés en particulier.

Ce parti pris théorique fait dire à Poldauf (1984) que les monosyllabes ne peuvent pas porter un accent. Cette conclusion est contestable : premièrement, parce qu'un monosyllabe est [illegible]né à faire partie de la chaîne de discours où il sera environné d'un ou plu[illegible]tiques et donc il sera lui-même accentué par rapport à ceux-ci ; mais surtou[illegible]mment expliquer autrement le placement de l'accent quand on ajoute des terminaisons neutres ? On pourrait peut-être dire que tous les monosyllabes ont un accent potentiel, qui ne se réalise que lorsque le mot se trouve dans la chaîne parlée, ou encore quand il s'enrichit en formant des dérivés. Le calcul de l'accent du dérivé se fera en fonction du radical monosyllabique, ce qui explique pourquoi Giegerich (1992) dit que le fait d'attribuer un accent à un mot monosyllabique a du sens sur le plan phonologique, même si sur le plan phonétique il n'apporte aucune information. En effet, il n'y a pas d'autre syllabe et donc l'accent ne peut pas entraîner les modifications associées à la syllabe non accentuée. La prononciation d'un mot lexical monosyllabique isolément ne sera aucunement affectée par l'attribution ou non d'accent. Mais comme notre analyse se limite à des composés qui sont par

définition formés d'au moins deux éléments, nous n'entrerons pas davantage dans ce débat.

Par ailleurs, Giegerich (1992) utilise un autre niveau d'analyse : le pied (*foot* en anglais). Dans ce que l'on appelle *metrical phonology*, on exprime ainsi les rapports relatifs de proéminence (*prominence relations*) à l'intérieur des mots. Ces pieds (*feet*) seraient plus ou moins forts et servent à mesurer le rythme (*timing*) de la phrase. Dans le chapitre 3, nous reprendrons les deux niveaux d'analyse, la syllabe et le pied, pour déterminer lequel est le plus adapté à notre analyse des composés ; ici, nous préférons parler d'accent par rapport à la syllabe.

Roach (1992 : 94), dans *Introducing Phonetics* lorsqu'il définit le rythme, fait référence à cette théorie, mais il paraît réticent :

> Most evidence from the study of real speech suggest that such rhythms only exist in very careful, controlled speaking, but it appears from psychological research that listeners' brains tend to hear timing regularities even where there is little or no physical regularity.

Cela rejoint ses réticences vis-à-vis de l'appellation « stress-timed language » pour l'anglais. Ce qui est à retenir dans cette citation, c'est l'écart important qui peut exister entre la réalité physique de la production et les impressions - beaucoup plus simples - au niveau de la perception. Nous avons déjà suggéré que la proéminence est un phénomène qui doit être appréhendé sur le plan psychologique ; il est pertinent sur le plan de la perception, et nous suggérons que l'impression de force relative est également à analyser de ce point de vue.

1.1.3 Une définition de l'accent

Nous devrons à présent être en mesure de répondre à la question, « Qu'est-ce que l'accent ? » Ayant considéré un certain nombre de points de vue, nous

retenons le principe suivant : l'accent est la résultante d'un ensemble de facteurs qui permet de faire ressortir une syllabe d'un mot par rapport aux autres syllabes. Appelons cet ensemble, comme le font tant d'auteurs, proéminence. Pour nous, les éléments les plus significatifs pour le repérage de l'accent sont certainement la hauteur mélodique, puis la relativité de la syllabe accentuée par rapport aux autres. Les caractéristiques des syllabes non accentuées nous semblent également avoir une grande importance et nous reviendrons sur la valeur de ces syllabes dans le cadre de notre analyse des composés. La suite de syllabes proéminentes et non proéminentes, par alternance, confère à l'anglais son rythme caractéristique et nous tenons là la clé du phénomène sur le plan de la production et celui de la perception. L'analyse acoustique nous permettra de mieux comprendre ce fonctionnement.

Nous avons, à plusieurs reprises, marqué la distinction qui est faite entre les deux niveaux : la production et la perception. Dans notre analyse, l'accent sera étudié surtout sur le plan de la perception. L'accent est perçu par l'auditeur sans pour autant qu'il en ait vraiment conscience. Ce n'est souvent que lorsque l'accent est déplacé (par un non-anglophone) que le locuteur natif se rend compte de son existence et plus précisément de l'importance de son placement. Son déplacement fait apparaître son incidence sur la compréhension. C'est cet aspect de l'accent que nous voulons aborder à présent.

1.2 Accent et compréhension

Nous nous sommes tous retrouvés dans des situations gênantes lorsqu'un locuteur non natif pose une question à laquelle nous ne comprenons pas grand-chose. Ceci est particulièrement fréquent quand un locuteur demande son chemin

en déplaçant l'accent, et qu'il est impossible de reconnaître le nom du lieu évoqué dans la question. Comme le fait remarquer Fudge (1984: 4): « a wrong stressing will lead to a wrong and misleading rhythm »[2].

Le locuteur natif s'attend à un certain rythme et quand il ne le retrouve pas il lui faut du temps (et même de l'imagination) pour retrouver le raisonnement qui lui permettra de deviner ce qu'on lui demande. O'Connor ([1973] 1991: 194) précise : « Words easily become unrecognizable if the stress is wrongly placed » et il poursuit: « it is an essential part of word shape ». Ceci explique pourquoi le locuteur natif va devoir retrouver l'orthographe – son lexique mental – pour repartir sur sa propre prononciation du mot.

Manifestement, le placement de l'accent fait donc partie de l'identification même d'un mot. Fournier (1992: 74) parle en ces termes du rôle de l'accentuation : « Du point de vue de la compréhension comme de la production, l'identité phonologique d'une unité repose donc essentiellement sur sa syllabe accentuée (ou ses syllabes accentuées dans le cas d'un accent /2/) ».

L'accent aurait donc un rôle d'identification. Il permet à l'auditeur de comprendre en focalisant son attention sur la syllabe qui est mise en relief par rapport aux autres. C'est d'ailleurs le mot *identifieur* qu'utilise Fournier: « L'accent s'analyse ainsi comme **un identifieur d'unités sémantiques,**

---

[2] Voir aussi les propos de Taylor (1981 : 225) : « Effective, recognizable rhythm is essential for the listener. if he is unable to perceive a rhythm which he can recognize by using his native speaker's competence [...] then he lacks the means to break up the message into its coherent parts. He cannot recognize the natural divisions and groups which are essential cues to meaning, and considerable impairment of meaning may result ».

confirmant le point de vue constant de Lionel Guierre que l'accent est un résultat, qu'il se déduit. » (Fournier, 1992, 88). Comme nous le verrons plus loin, lorsque nous rappelons les règles de l'accentuation de Guierre, l'accent en anglais tend à remonter à gauche. D'ailleurs une des règles fondamentales de cette langue, c'est qu'il doit y avoir un accent sur l'une des deux premières syllabes d'un mot. Fournier, dans ce même article (1992 : 94-5), lorsqu'il étudie les préfixés, suggère que l'accent signale où trouver l'information, et il argumente qu'il y a une concentration de l'information à gauche. Sa conclusion donne à réfléchir :

> Comme nous l'avons vu, l'accentuation anglaise est fondamentalement un identificateur sémantique : le caractère inacceptable d'une perte d'information à gauche, au début du mot, et donc le caractère accessoire, de ce point de vue, de la droite du mot : ces faits conduisent ainsi à suggérer que l'anglais présente une localisation de l'information pertinente.

Cette hypothèse est intéressante pour l'analyse des composés que l'on divise en général en deux classes, l'une avec l'accent effectivement à gauche (sur le premier élément), l'autre où les deux éléments portent un accent. Cela constitue également un argument en faveur des critères sémantiques de définition d'un composé que nous étudierons dans le chapitre suivant.

Il est important de se poser la question : « Quelle est l'importance du placement ? » Il y a des cas où l'accentuation du composé joue un rôle important. Son placement permet de distinguer des sens différents, et donc de désambiguïser ; le composé sera interprété selon l'accent. Un exemple classique, emprunté à Boucher (1998), c'est la paire « *ˈbrick ˌfactory* » et « *ˌbrick ˈfactory* », où l'accent à gauche indique l'interprétation « usine où on fabrique des briques », alors qu'avec l'accent sur les deux éléments avec chute mélodique à droite s'interprète comme « une usine qui est construite en briques ». Un cas similaire est

étudié par Huart (Huart, 1984 : 247) : *ˈtoy factory*, désignant une usine où l'on fabrique des jouets, et *ˌtoy ˈfactory* qui représente une usine qui est en fait un jouet. Nous nous intéresserons à ce genre de cas dans notre analyse[3].

Il est vrai que dans la vaste majorité des cas, s'il y a un déplacement anormal de l'accent (les deux éléments sont accentués à la place d'un accent unique sur le premier élément, ou *vice versa*) l'effet produit sera simplement bizarre, trahira le locuteur non natif, mais ne posera pas de réels problèmes de compréhension, contrairement à l'effet d'un déplacement d'accent sur un mot non composé, qui nuira forcément à la compréhension. Toutefois, l'interprétation du sens du composé est influencée par le type d'accentuation. Dans le cas d'un composé, formé d'un adjectif suivi d'un nom, avec un accent uniquement sur le premier élément, si les deux éléments portent l'accent, il sera interprété comme un syntagme. Nous citons la paire bien connue : *ˈblackbird* qui est un composé désignant un espèce d'oiseau et *ˌblack ˈbird* qui est un syntagme désignant un oiseau qui est noir. Dans le cas d'un composé portant un accent sur les deux éléments, le fait de mettre le premier en relief aux dépens du second risque de troubler l'interprétation de l'unité, même si les cas d'ambiguïté sont rares ; ainsi, si l'on prononce *ˈheadmaster* au lieu de *ˌheadˈmaster*, nous conclurons à un accent de contraste sur *head* qui l'oppose à *deputy headmaster*, par exemple. Nous étudierons ces différents types d'accent dans la partie suivante.

---

[3] Voir 4.6.1.

La distinction entre les composés accentués à gauche et ceux qui sont accentués à droite, pour reprendre la nomenclature de Boisson (1980), est donc pertinente sur le plan de la compréhension. Dans ses travaux, il fait aussi remarquer le comportement différent de l'accentuation et de l'intonation au niveau de l'interprétation de l'énoncé. Nous l'avons vu, l'accent en anglais tend à être à gauche ; or, le noyau intonatif, c'est-à-dire la syllabe où se trouvera la chute mélodique au niveau de l'énoncé dans son ensemble, sera le plus à droite possible. Il explique (Boisson, 1980 : 689):

> Il est vraisemblable que ceci est dû au désir de mettre en valeur par nucléarisation ce qui se trouve en fin de phrase, et qui tend, en anglais, à être l'information nouvelle par opposition à l'information donnée (« new » / « given »).

Nous retrouvons le problème de la localisation de l'information soulevé par Fournier, mais nous voyons que le comportement de l'intonation est donc opposé à celui de l'accent. Cependant, nous sommes persuadée que les deux jouent un rôle au niveau sémantique. Ces deux phénomènes sont-ils contradictoires ou bien peut-on les concilier ?

D'abord, il faut rappeler que lorsque l'on parle d'accent, on se réfère en général à des mots pris isolément ; l'intonation n'a de sens que dans un contexte d'un énoncé (d'un ou plusieurs mots). Les remarques faites précédemment sur la hauteur mélodique et l'accent se réfèrent à ce qui se passe si la syllabe accentuée du mot est le noyau de l'énoncé ; il s'agit donc d'une chute mélodique potentielle (ce que Gimson (1980 : 225) a appelé *potential change of pitch*). Dans la chaîne parlée, l'intonation permet de faire passer un certain « message » ; elle est significative pour la compréhension de l'énoncé que l'on va interpréter en fonction de l'intonation utilisée. Dans cette chaîne parlée, il y aura le plus souvent

beaucoup de syllabes accentuées, mais une seule (par unité tonale) porte la chute (ou montée) mélodique associée au noyau. Ce n'est pas contradictoire avec notre hypothèse que l'accent et l'intonation sont des marqueurs sémantiques tout simplement parce que, pour comprendre l'énoncé, il ne suffit pas de comprendre le mot où se trouve le noyau. Les deux notions suprasegmentales agissent à des niveaux différents et sont plutôt complémentaires que contradictoires.

Nous avons été sensible à l'interaction de l'accent et de l'intonation. Les deux jouent un rôle sémantique, l'accent au niveau du mot, l'intonation au niveau de l'énoncé. Cependant, il y a souvent une confusion entre ces deux niveaux et pour qu'il n'y ait pas de confusion possible, nous voulons définir les différents types d'accent et déterminer sur quel plan nous nous situons pour notre analyse.

1.3 Les différents types d'accent

Pour l'interprétation de l'énoncé, il faut insister sur le rôle crucial que joue le noyau au niveau informationnel. La tendance effectivement est de le mettre à droite[4], comme dans l'exemple suivant :

John likes ↘*cars.*

Dans cet énoncé, le mot lexical le plus à droite, *cars*, portera l'accent nucléaire, indiqué par la flèche descendante. Cependant, cet accent peut se déplacer dans certaines circonstances, pour des raisons bien précises et qui tiennent au sens que l'on veut donner à l'énoncé. En effet, si le noyau remonte à

---

[4] Voir la règle LLIR, « Last Lexical Item Rule », Deschamps et al., 2000 : 195.

gauche, c'est pour attirer l'attention sur le mot sur lequel est la chute (remontée) intonative. Ainsi, si le locuteur veut insister sur le fait que c'est John, et personne d'autre, qui aime les voitures, l'accent nucléaire se trouvera sur *John*.

↘*John likes cars.*

Dans ce cas, *likes cars* est déjà posé et *John* est considéré comme l'information nouvelle. Ce phénomène est caractéristique d'un certain type d'accent : l'accent de phrase. Boisson (1980 : 689) fait référence à ce type de contexte : « un terme normalement accentué peut subir une désaccentuation anaphorique lorsqu'il a déjà été mentionné ».

Il nous rappelle d'autres contextes où l'on trouvera ce genre d'accent : « un terme normalement inaccentué peut recevoir un accent anormal à des fins contrastives [...] » (p. 689), comme dans l'énoncé suivant, ou la partie soulignée porte l'accent nucléaire (également indiqué par la flèche descendante) :

I prefer tea ↘*bags*

Ce composé serait accentué à gauche normalement, mais dans cet énoncé, l'élément *bags* est accentué par opposition à *tea leaves*. Il s'agit dans tous les cas de l'accent de phrase (aussi appelé accent de contraste) qui est attribué en fonction du sens que l'on veut donner à l'énoncé tout entier. Nous en avons eu un autre exemple avec *headmaster* (voir 1.2).

Kingdon (1958b) donne une description détaillée des accents de l'anglais et propose un grand nombre de catégories[5]. Il appelle l'accent de phrase *semantic stress*, mais il le met sur le même plan que les accents rythmiques. C''est justement ce que nous voulons éviter, car l'accent de phrase et l'accent rythmique (lexical) sont de nature très différente. L'accent de phrase a sa place dans une analyse du discours, mais il n'est attribué qu'en fonction du sens d'un énoncé particulier. Il ne met pas en jeu l'identité du mot.

Troubetzkoy (Troubetzkoy, 1949: 312) attribue une fonction très particulière à l'accentuation qui est un des signes aphonématiques permettant de marquer la frontière d'un mot ou d'un morphème dans certaines langues. Il emploie le terme «fonction délimitative». Ce type de fonction de l'accent se rapproche du *silent stress* proposé par Giegerich (1992: 272). Il s'agit en fait d'une sorte de pause et relève encore d'une analyse du discours puisqu'il sert à désambiguïser un certain type de syntagme. Dans *old men and women*: cet accent (ou pause) se placera après *old* si *old* s'applique aux deux noms, mais si *old* ne s'applique qu'à *men*, la pause sera après *men*. Parler d'accent plutôt que de pause ici nous semble souligner le rôle démarcatif que joue l'accentuation. Il est intéressant de noter que le schéma accentuel de ce syntagme sera également différent. Dans le premier cas, l'accent primaire portera sur la syllabe accentuée de *women*, *men* se trouvant avec un accent affaibli, puisque le syntagme *men and women* est considéré comme un ensemble, alors que dans le deuxième cas, *men*

---

[5] Le système accentuel qu'il propose est fort complexe. Les accents rythmiques tiennent compte de l'intonation (kinetic stress), de la hauteur mélodique (high level / low level), et toutes les combinaisons possibles (multiple stress, triple stress). voir son introduction (Kingdon, 1958b : ix).

portera un accent aussi important que *women*. Dans ce cas, il s'agit de deux ensembles distincts : *old men* d'une part, *and women* d'autre part, auxquels on attribue une égale importance, ce qui est reflété par le schéma accentuel. Nous retrouverons ces considérations de hiérarchie accentuelle et l'accent dans le discours ultérieurement.

Il faut donc distinguer les accents qui portent sur des énoncés ou des syntagmes (accent de phrase, *silent stress*) où ils jouent en fait le rôle de structuration syntaxique qui dissipe l'ambiguïté de la chaîne linéaire, et l'accent lexical (également appelé *word stress*). Même s'ils peuvent tous servir de repère au niveau du sens, il ne faut pas les confondre. Les accents qui opèrent au niveau de l'énoncé permettent d'attribuer un sens spécifique à celui-ci. Avec les mêmes mots, ayant le même sens, on peut formuler des énoncés ayant des interprétations différentes. Bien au contraire, pour l'accent lexical, l'accent est constant selon le sens du mot et, pour un sens donné, ne peut être modifié. Notre analyse s'intéresse à l'accent lexical, qui rend compte du rythme essentiel de la langue, et, dans l'analyse acoustique, nous serons amenée à tenir compte de facteurs contextuels ou situationnels. Dans ces cas, le mot « accent » sera qualifié par « accent d'insistance » ou « accent de contraste »[6].

---

[6] Voir 4.5.1.2.4 « Les contextes non neutres ».

1.4 La hiérarchie accentuelle

L'accent lexical caractérise le rythme de l'anglais qui est une série de syllabes fortes et faibles qui se suivent en alternance plus ou moins stricte. La syllabe forte est une syllabe accentuée, caractérisée par un certain degré de proéminence, par rapport aux syllabes intermédiaires qui sont donc considérées comme étant des syllabes faibles. Ces syllabes jouent un rôle rythmique et contribuent à la mise en relief des syllabes fortes.

La hiérarchie accentuelle correspond au nombre de degrés d'accent que l'on distingue. Nous l'avons vu quand nous avons étudié la définition même de l'accent, les points de vue sont variés, et nous faisons le même constat en ce qui concerne le nombre de degrés d'accent qui sont reconnus par les auteurs. Précisons d'abord la nomenclature que nous allons utiliser. Le terme *degré d'accent* est parfois utilisé pour désigner les différents accents forts, sans inclure l'accent faible (le non-accent). Dans cette étude, nous incluons l'accent faible (le non-accent) dans le décompte des degrés d'accent. Ainsi, lorsque nous parlons d'une syllabe non accentuée, cette syllabe porte le degré zéro (généralement noté /0/).

Ainsi, tous les auteurs s'accordent pour désigner au moins deux degrés d'accent (un accent fort et un accent faible) ; mais combien d'accents forts faut-il distinguer ? Pour tenter de confronter les points de vue sur cette question, nous avons conçu deux tableaux récapitulatifs. D'abord, nous nous sommes intéressée aux usages dans les ouvrages de référence dans le domaine : les dictionnaires de prononciation de l'anglais (britannique) *The English Pronouncing Dictionary* (appelé EPD avec le numéro d'édition accolé suivi de l'année d'édition) et

Longman Pronunciation Dictionary (appelé LPD avec l'année d'édition). Il s'agit ici de la notation de l'accent de l'ensemble du lexique, et non seulement des composés que nous étudierons plus loin, à la lumière de cette discussion générale.

Dans le tableau, nous indiquons le signe employé par le dictionnaire pour désigner l'accent, puis sa valeur. Les valeurs /1/ et /2/ correspondent aux accents dits « forts », alors que la valeur /0/ correspond au non-accent, à l'accent dit « faible ». Les syllabes non marquées par le dictionnaire ont cette valeur par défaut. Dans les ouvrages d'aujourd'hui, l'accent primaire (portant la chute mélodique) correspond à l'accent /1/, et l'accent secondaire (rythmique) correspond à l'accent /2/. Cette nomenclature sera reprise et expliquée plus loin dans la discussion sur les critères de détermination de la hiérarchie accentuelle.

Figure 1.2 : Conventions de notation des dictionnaires de prononciation

Nombre de degrés Ouvrage de référence	3 degrés	4 degrés	Signes & Appellations
EPD12 1963 EPD13 1967 EPD14 1977 EPD15 1997 LPD 2000	oui		ˈ = /1/ (main/strong stress) ˌ = /2/ (secondary stress) non marqué = /0/
LPD 1990		oui	ˈ = /1/ (primary word stress) ˌ = /2/ (secondary stress) ˌ = /3/ (tertiary stress) non marqué = /0/

Nous constatons qu'il y a un consensus quasiment unanime à l'exception de la première édition de LPD (1990). Ce dictionnaire a introduit l'accent tertiaire qui se distingue de l'accent secondaire dans plusieurs contextes, ce qui implique

qu'il distingue quatre degrés d'accent[7]. A l'heure actuelle, ces ouvrages s'accordent pour en indiquer trois : deux accents forts – accent primaire et accent secondaire – et un accent faible (le non-accent). La deuxième édition de LPD a simplement remplacé l'accent tertiaire par l'accent secondaire.

Ces ouvrages sont conçus essentiellement pour des usagers non-anglophones et les symboles utilisés doivent être suffisamment détaillés pour permettre aux apprenants de réaliser une prononciation fidèle des mots dont la prononciation leur est inconnue. On peut conclure de ce consensus que trois degrés d'accent sont donc nécessaires et suffisants pour guider l'apprenant. Cependant, le débat sur le nombre de degrés « réels » est loin d'être clos.

Dans le second tableau récapitulatif de la notation de la hiérarchie accentuelle, nous nous sommes intéressée aux auteurs d'ouvrages pédagogiques en phonologie et phonétique anglaise ainsi qu'à la notation de Chomsky et Halle (1968), auteurs du premier ouvrage théorique marquant en phonologie générative, qui se démarque nettement de la notation habituelle. Nous avons également inclus d'autres auteurs qui proposent des hiérarchies différentes. Enfin, nous avons inclus la notation de Boisson (1980) et Huart (1984 et 1989) qui tient compte de manière spécifique des composés. Le tableau regroupe les auteurs selon le nombre de degrés reconnus : comme ces degrés correspondent à des réalités différentes, la troisième colonne comporte un commentaire qui explique à quoi correspond la notation. Celle des dictionnaires est incluse à titre de référence. Ensuite, nous

[7] Voir la discussion dans la dernière section de cette partie, 1.4.1.5 « Rythme et qualité vocalique : l'accent /ə/ en question ».

pourrons étudier les différentes prises de position sur la question, afin d'être en mesure de définir pour notre analyse le nombre de degrés qui sera utile.

Figure 1.3 : Synthèse des différentes hiérarchies accentuelles

Nombre de degrés	Référence	Commentaire de la hiérarchie choisie
Plus de 4 degrés	Kingdon (1958b)	Nombreuses appellations spécifiques *kinetic stress*, appellations relatives aux tons et l'interaction des différents types d'accent, *static stress*
2 degrés et plus de 4 degrés	Chomsky & Halle (1968)	2 niveaux d'analyse Niveau superficiel > 4 degrés Niveau profond 2 degrés (accent ou non-accent)
3 degrés et plus de 4 degrés	Viel (1981, 1994)	Pas de limite supérieure au nombre de degrés /0/ : non-accent avec réduction vocalique Mêmes indications que les dictionnaires sur le plan pratique
3 degrés	EPD12-15 et LPD2000	/1/ accent primaire /2/ accent secondaire /0/ non-accent
	Guierre (1979 & 1984), Roach (1992), Ladefoged (1993), Duchet (1994 et 1998), Ginésy (1995), Deschamps et al (2000)	Mêmes indications que les dictionnaires[8]
	Gimson (1980)	Mêmes indications que les dictionnaires mais des appellations spécifiques : accent primaire : *primary / nuclear / tonic accent* accent secondaire : *secondary / rhythmic / non-tonic accent*
	Giegerich (1992)	/1/ accent primaire /2/ accent secondaire (ou non-accent) sans réduction vocalique /0/ non-accent avec réduction vocalique
4 degrés	LPD1990	/1/ accent primaire /2/ accent secondaire /3/ accent tertiaire /0/ non-accent avec ou sans réduction vocalique
	Fudge (1984)	/1/ : accent primaire en position nucléaire : *nuclear stress*, noté ˈ /2/ : accent secondaire, noté ˈ devant /3/ : accent tertiaire correspondant à /3/ ou /0/, noté , /0/ : accent non marqué
	Lilly & Viel (1993 & 1999)	/1/ : accent primaire /2/ : accent secondaire /3/ : accent tertiaire[9] correspondant au non-accent sans réduction vocalique /0/ : non-accent avec réduction vocalique
2 degrés et 4 degrés	Ladefoged (1993)	2 degrés : syllabe accentuée ou inaccentuée 4 degrés : syllabe accentuée, position nucléaire ou non syllabe inaccentuée, voyelle pleine ou voyelle réduite
	Boisson (1930).	Indications pour les composés /2-1/ : D (droite, accent sur les 2 éléments) /1-0/ : G (gauche, accent sur 1er élément)
	Huart (1984 & 1989)	Indications pour les composés /1/ : accent primaire en position nucléaire, chute mélodique, noté ˈ /1/ : accent primaire ne correspondant pas à la syllabe nucléaire, noté ˈ /2/ : accent secondaire, noté , /0/ : accent non marqué

Le tableau montre clairement les conclusions des auteurs en matière de hiérarchie accentuelle, ainsi que les correspondances entre leur analyse et les notations des

[8] L'accent primaire est noté /2/ et l'accent secondaire /1/ dans l'ouvrage de Guierre, 1979.

[9] Cet accent est marqué dans l'ouvrage de 1993, mais ne l'est plus en 1999.

dictionnaires de prononciation. Il nous appartient maintenant d'étudier les critères utilisés pour déterminer cette hiérarchie.

1.4.1 Critères de détermination de la hiérarchie accentuelle

La première question est, en effet, de savoir de quelle manière on peut déterminer cette hiérarchie. Les problèmes de repérage de l'accent ont déjà été évoqués dans la partie concernant la définition de l'accent où nous avons constaté que les auteurs ne s'accordent pas, mais qu'il y a un certain nombre de facteurs auxquels ils se réfèrent presque tous, auxquels ils attribuent une importance plus ou moins grande. Certains de ces facteurs sont repris plus loin en tant que critères pour la détermination de la hiérarchie accentuelle. Nous commencerons par situer la hiérarchie accentuelle par rapport au débat perception / production dont nous avons discuté (voir 1.1.1.1).

1.4.1.1 Perception et production

Nous avons déjà étudié le phénomène de proéminence qui est la clé de la définition de l'accent pour la plupart des auteurs, même s'ils ne s'accordent pas toujours sur la question de savoir quels sont les facteurs qui la composent et quel est leur poids relatif. Ces facteurs correspondent tantôt au niveau de la perception, tantôt à la production, mais la proéminence en est la résultante et c'est elle qui permet aux auditeurs de percevoir l'accent.

Sous la rubrique 1.1.1.1, nous avons conclu qu'il n'y avait pas d'opposition entre les niveaux de perception et de production, mais plutôt une complémentarité. Toutefois, en ce qui concerne la hiérarchie accentuelle, notre souci concerne la manière de noter l'accent, non la description de la production

des énoncés. Ceci explique pourquoi la plupart des auteurs ne retiennent que trois degrés d'accent. En effet, Gimson (1980 : 58) souligne que le niveau de la production, qui correspond à la réalité physique, est fort complexe mais que la réalité perceptive est beaucoup plus simple et il conclut qu'une notation simplifiée suffit. L auditeur n'en percevra pas davantage.

> [...] stress [...] rarely requires more than two or three degrees to be shown, although the speaker's intensity will show an infinite number of degrees of variation [...] the ultimate notation will present functional categories, abstracted from the concrete expression.

Une telle simplification motivée par le souci de mettre en avant l'information pertinente nous rappelle la distinction entre phonologie et phonétique, entre transcription large et transcription serrée. Duchet (1981, 1998 : 14) cite Sweet :

> Henry Sweet (1845-1912) sent bien qu'il n'y a pas de limite à la finesse des descriptions phonétiques, et il ressent la nécessité d'introduire une distinction entre les traits phonétiques qui importent à la langue et les autres, entre une transcription qu'il appelle « large » et une transcription « étroite ».

Gimson fait référence à deux ou trois degrés d'accent et le tableau de synthèse nous permet de constater que la majorité des auteurs ainsi que les dictionnaires de prononciation distinguent effectivement trois degrés dans la hiérarchie accentuelle correspondant à des syllabes fortes (degrés /1/ et /2/) : deux accents « forts » et une syllabe faible (degré /0/). C'est également le point de vue de Roach (1992 : 104) qui arrive aux mêmes conclusions que Gimson ; il limite le nombre de degrés à trois pour des raisons pratiques :

> In English, we can argue that if we take the word *indicator* as an example, the first syllable is the most strongly stressed, the third syllable is the next most strongly stressed and the second and fourth syllables are weakly stressed, or unstressed. This gives us three levels: it is possible to argue for more, though this rarely seems to give any practical benefit.

Même si Fudge (1984 : 135) propose quatre degrés d'accent, donc un degré de plus, il tient sensiblement les mêmes propos :

> Notice that our notation appears to have run out of means of differentiating between stress levels: [...] The notation could be refined so as to maintain the marking of different stress levels, but this would in fact be misleading: the physical properties which signal stress in English [...] do not enable hearers, even trained phoneticians, to distinguish consistently more than three degrees of strength.

La difficulté qui existe à percevoir des degrés d'accent différents est donc décisive pour déterminer une hiérarchie accentuelle relativement simple. Ladefoged (1993) propose un système[10] qui tienne compte de l'intonation et de la qualité vocalique, ce qui impliquerait 4 degrés d'accent, mais il est réservé vis-à-vis de son application car, pour lui, ces phénomènes, qui peuvent nous aider à distinguer l'accent, sont des phénomènes distincts et il en conclut que deux niveaux suffisent (Ladefoged, 1993 : 116) :

> In summary, we can say that English syllables are either stressed or unstressed. If they are stressed, they may or they may not be the tonic syllables, which carry the major pitch change in the tone group. If they are unstressed, they may or may not have a reduced vowel.

Par ailleurs, il fait remarquer que API n'offre que trois possibilités pour indiquer l'accent : « ˈ » (/1/) « ˌ » (/2/) et non marqué[11].

Ladefoged est également cité par Boisson (1980 : 357-8) pour souligner la difficulté qui existe à percevoir les différences entre les degrés d'accent :

> [...] selon Lagefoged, il est pratiquement impossible de distinguer la force de l'accent dit « secondaire » de celle d'un accent dit « primaire ».
>
> Il faut donc prendre garde au fait que l'expression « écho d'accent » ou « accent subsidiaire », « accent secondaire », désigne bien un accent qui, dans certains cas, dépend de l'accent principal pour le calcul de sa localisation, mais qui ne lui est pas nécessairement subordonné par la force.

---

[10] Il établit un tableau comprenant les trois critères suivants pour décrire la syllabe : *tonic accent, stress, full vowel*. En fait, si les trois conditions sont remplies, nous aurons une syllabe avec la valeur /1/, si deux conditions seulement sont remplies, nous aurons la valeur /2/, et si une seule condition est remplie, nous aurons la valeur /3/. Dans le cas où aucune condition n'est remplie, nous aurons la valeur /0/.

[11] En fait, il existe une troisième possibilité, « ˌˌ », utilisée par Wells dans la première édition de LPD (1990), qui correspond à l'accent tertiaire.

Boisson (1980 : 78) suggère que le niveau de la perception est justement celui qui est significatif et non celui de la production. Il le dit, en commentant les propos de Jones qui s'inquiétait du fait que ses enregistrements ne pourraient pas être utilisés pour déterminer le nombre de degrés d'accent :

> On pourrait toutefois objecter à Jones qu'il n'est pas impensable de déterminer des degrés d'accent « objectivement » à partir, non pas des enregistrements « objectifs » des machines, mais des impressions « subjectives » des auditeurs, qui nous importent plus.

Nous retiendrons donc pour notre analyse le niveau de la perception. Nous allons nous tourner dans la suite de cette discussion vers d'autres types de critères qui sont tous d'ordre perceptif : la mélodie, le rythme et la qualité vocalique.

1.4.1.2 Mélodie

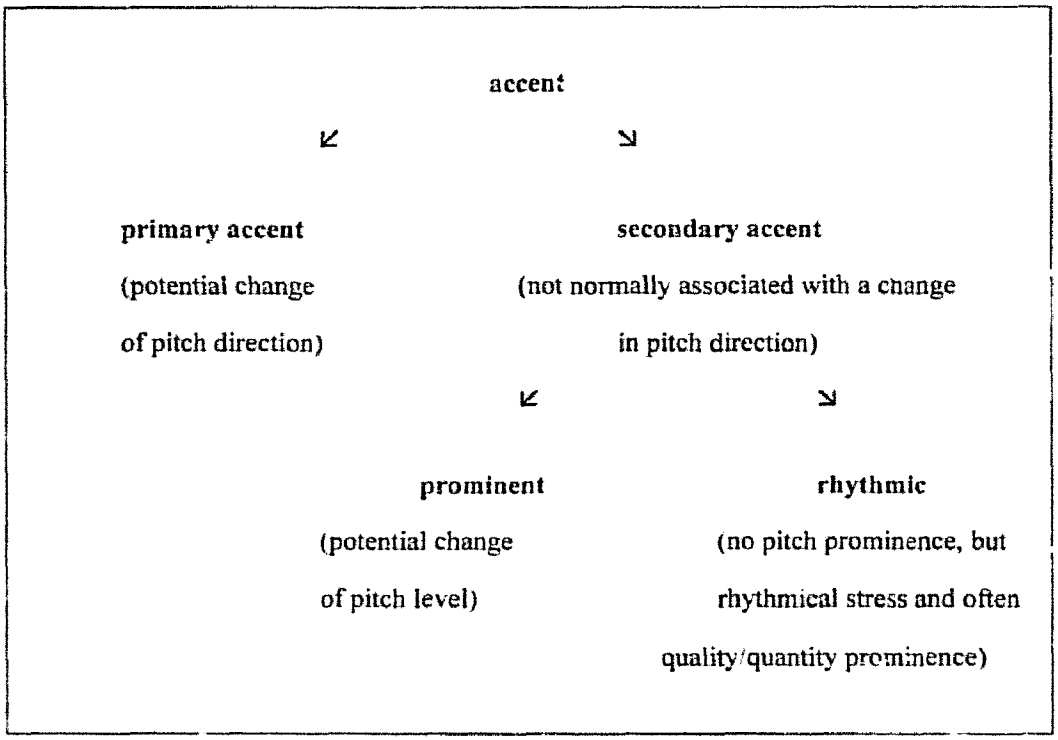
Autant la différence entre syllabe forte et syllabe faible paraît relativement évidente, autant la distinction entre les degrés des syllabes effectivement accentuées (accent /1/ et /2/) ne semble pas facile à faire, au point que Ladefoged (1993) propose de considérer que la distinction entre degrés n'est systématiquement perceptible qu'en fin de phrase ou dans les mots pris isolément. L'accent /1/ correspond, dans ces contextes, à l'accent nucléaire. Pour Goldsmith (1990 : 185), c'est ce qui différencie l'accent primaire des autres accents :

> In most languages, what is called primary, or word-level stress, is the assignment of a special tonal (or intonational) melody to the syllable in question, in addition to the stress effects perceived on the other secondarily stressed syllables. In such languages, including English, there is in this way a natural distinction between a primary stress and a secondary stress.

De même, Gimson propose deux accents forts. L'accent /1/ correspond à ce qu'il appelle *primary, nuclear, or tonic* et il est associé au changement mélodique, *pitch change* (mot isolé) ou *potential change of pitch* dans le contexte du discours

continu[12] ; l'accent /2/ correspond à ce qui est appelé *secondary, rhythmic, or non-tonic* et il est associé à un changement potentiel de niveau *potential change of pitch level*. Le tableau suivant, inspiré de Gimson[13], récapitule la hiérarchie des accents.

Figure 1.4 : Récapitulatif des accents, d'après Gimson



On note les diverses appellations des accents /1/ et /2/, qui indiquent clairement le rôle que Gimson attribue à la mélodie comme facteur permettant de déterminer la hiérarchie accentuelle. L'analyse acoustique nous permettra de le vérifier. D'autres auteurs le rejoignent pour insister sur l'importance de la mélodie

---

[12] Seule une syllabe par unité tonale correspond au noyau intonatif : cette nomenclature distingue donc entre le mot isolé dont une syllabe correspond au noyau, et le mot dans un contexte de discours où il est possible que la syllabe soit associée à un changement de la hauteur mélodique, selon son placement.

[13] Gimson, 1980, p. 84

pour distinguer entre les deux accents forts. Boisson (1980 : 357-8) commente les propos de Ladefoged sur la difficulté à distinguer entre accent primaire et accent secondaire : « Il faut que d'autres composantes de la proéminence viennent s'associer pour le faire ressortir, et spécialement l'inflexion mélodique du noyau de groupe phonologique ».

Cependant, c'est certainement Kingdon, auteur d'un des premiers ouvrages sur l'accent en anglais, qui attribue le plus d'importance à la mélodie dans la distinction entre les degrés d'accent. Son appellation de l'accent /1/ est à ce titre révélatrice ; c'est *kinetic stress* (Kingdon, 1958b : ix), ce qui nous rappelle l'importance que lui et d'autres accordent à ce phénomène. Le terme rappelle le mouvement (donc l'intonation). Le système d'accents qu'il propose est sans doute le plus complexe de tous en raison de son souci de donner une description détaillée et fidèle du phénomène de l'accent. Cependant, la complexité de ses appellations nuit quelque peu à la clarté de l'analyse. Ses références aux accents secondaires sont très floues (Kingdon, 1958b : xiii) : « The question of what degree of prominence a syllable must have before it deserves to be marked as partially stressed is, of course, largely a matter of opinion ». Une formulation plus précise serait souhaitable.

Nous terminons cette rubrique qui traite de la mélodie en nous référant à deux auteurs qui ont analysé les composés : Boisson (1980) et Huart (1984 et 1989). Nous constatons que, d'après le tableau, Boisson est parmi les rares a n'utiliser que deux degrés d'accent. Il pratique ce système parce que cela correspond aux besoins de son analyse, mais par ailleurs il reconnaît un nombre de degrés supérieur. Il se limite dans son analyse à la distinction entre

l'accentuation du premier terme (accent à gauche) où du deuxième terme (accent à droite), mais il reconnaît en fait quatre degrés d'accent en tenant compte de l'intonation. Ce phénomène permet de distinguer entre ce que l'on appelle l'accent primaire et l'accent secondaire. Quant à Huart, elle utilise surtout la notion d'accent initial (chute mélodique sur le premier terme) et accent final (chute mélodique sur le dernier terme), donc se restreint, comme Boisson, à une analyse n'utilisant effectivement que deux degrés d'accent. Toutefois, une des caractéristiques marquantes de son travail (sur le plan accentuel) est la signification qu'elle attribue à l'intonation puisqu'elle a préféré parler de chute mélodique que d'accent primaire ou d'accent /1/ et ce terme reflète clairement son parti pris. Nous en reparlerons plus loin, lorsque nous définirons la nomenclature appropriée à notre analyse. Retenons toutefois que ces deux auteurs s'accordent pour distinguer quatre degrés d'accent au niveau des composés, en tenant compte de la hauteur mélodique, même s'ils n'utilisent que deux degrés d'accent dans leur notation.

Dans notre étude des composés, il faudra déterminer s'il suffit de repérer lequel des deux éléments porte l'accent primaire comme Boisson et Huart, ou s'il est utile de distinguer une hiérarchie plus complexe en tenant compte de la mélodie.

1.4.1.3 Rythme

Nous avons constaté que nous pouvons reconnaître l'accent /1/ grâce à la mélodie quand un mot est isolé parce que la syllabe accentuée porte la chute

nucléaire[14] Néanmoins, en discours cette differenciation entre deux accents forts est beaucoup plus difficile à faire. Rappelons d'abord la conclusion de Boisson (1980 : 56) sur la question des deux accents :

> Mais, à en croire Ladefoged, à l'intérieur d'un groupe phonologique, on accentuera *inspiration* avec une égale force sur la première et sur la troisième syllabe [...]. Nous verrons pourtant que, d'une manière ou d'une autre, la notion d'accent subsidiaire reste nécessaire en anglais.

En effet, deux accents (et non un seul) sont généralement attribués à l'anglais car cela correspond à son rythme. Se succèdent syllabes fortes et syllabes faibles en alternance. C'est le point de vue de beaucoup d'auteurs qui distinguent trois niveaux d'accent, comme le font les dictionnaires de prononciation (à l'exception de LPD90 qui en distingue quatre) : l'accent primaire (accent/1/) l'accent secondaire (accent/2/), étant des accents forts, et la syllabe faible non accentuée (accent/0/). Ainsi, les avis de Gimson (1980), Guierre (1979 & 1984), Roach (1992), Ladefoged (1993), Duchet (1994 et 1998), Ginésy (1995), et Deschamps et al. (2000)[15] concordent, malgré les problèmes de reconnaissance de ces deux accents dont nous avons fait état plus haut. En effet, cette opposition reste nécessaire en raison des contraintes rythmiques de la langue anglaise. Le rythme est l'un des phénomènes, sinon le phénomène, le plus caractéristique de l'anglais ; les descriptions de la langue citées au début de ce chapitre nous le prouvent. Comme le dit Fudge (Goyvaerts & Pullum, 1975 : 291), en commentant l'ouvrage de Chomsky & Halle (1968) :

---

[14] Dans le cas d'une intonation descendante, nous parlons de chute, mais il existe d'autres contours intonatifs qui sont étudiés dans le chapitre 4, en 4.5.1.2.3. La syllabe nucléaire est associée au changement de la hauteur mélodique.

[15] Giegerich distingue le même nombre de degrés mais l'accent secondaire n'a pas la même signification.

> Differences between degrees of non-main stress are best handled within the phonetic component (see Fudge, 1972, 144) since the principle underlying them appears to be rhythmic – secondary stresses next to main stress are weaker, the ones next to these secondary stresses are stronger, the ones next to these are weaker, and so on.

Kager (Goldsmith, 1995 : 367) dans sa description de l'accent fait remarquer qu'il peut être rythmique dans certains cas : « [...] stress is rhythmic in systems where stressed and stressless systems alternate, and where clashes (adjacent stresses) are avoided. Naturally, stress does not assimilate to adjacent syllables, as this would produce clashes ». C'est le cas de l'anglais, et le système de phonologie métrique que Kager et Goldsmith (Goldsmith, 1995) défendent tente justement de représenter clairement cet aspect de la langue. Nous y reviendrons quand nous étudierons les différentes approches théoriques.

Quant à Guierre (1979 : 319), nous citons ses règles qui démontrent les caractéristiques rythmiques de l'anglais :

règle 1	Tout mot comporte un et un seul accent primaire.
règle 2	Aucun mot ne commence par une séquence de deux accents 0/.
règle 3	Les séquences de deux accents forts[16] sont non grammaticales

Ces contraintes expliquent comment, lorsque l'accent primaire se trouve sur la troisième syllabe ou au-delà, c'est un accent secondaire qui est attribué automatiquement à l'une de ces deux syllabes (règle 2). De même, nous constatons l'alternance entre syllabes fortes et syllabes faibles puisque deux accents forts ne se juxtaposent pas (règle 3). Avec ce raisonnement, auquel nous souscrivons, nous aboutissons forcément à deux accents forts. La question de

---

[16] Le terme « accent fort » recouvre l'accent primaire et l'accent secondaire.

savoir s'il faut encore distinguer entre des degrés différents d'accent non primaire reste ouverte.

1.4.1.4 Qualité vocalique

Dans le tableau, nous remarquons que le degré /0/ est attribué selon deux critères différents. Pour la plupart des auteurs, l'accent /0/ correspond à la valeur de la syllabe non accentuée, quelle que soit sa qualité vocalique. L'attribution de /0/ repose donc sur des critères rythmiques, essentiellement. Toutefois, quelques auteurs utilisent la réduction vocalique comme seul critère de détermination d'une syllabe non accentuée (Giegerich, 1992; Viel, 1991 & 1994; Lilly & Viel, 1993 & 1999). Toute syllabe comportant une voyelle pleine se voit attribuer un accent de degré supérieur, mais qui diffère selon les hiérarchies, les systèmes pouvant être plus ou moins nuancés. Nous allons étudier ces différents systèmes.

Giegerich (1992) reconnaît effectivement trois degrés d'accent, mais n'entre pas dans le détail, et part du principe que toute syllabe sans réduction vocalique, ne portant pas d'accent /1/, porte un accent /2/ (*mayhem, gymnast, compost.*) Les degrés reconnus le sont uniquement à partir de ce critère acoustique. Ce système ne nous paraît pas rendre compte de la réalité et de la complexité des faits car l'accent /2/ ici n'a rien de rythmique. D'après Giegerich (1992), le rythme se situe au niveau du pied, non de la syllabe et il l'analyse comme une série de S/W (*strong and weak feet*) ; nous détaillerons ultérieurement cette théorie quand nous aborderons la théorie de la phonologie métrique.

Quant à Viel, il propose plus de quatre degrés d'accent ; seule la syllabe comportant une voyelle réduite porte l'accent /0/. Ses propositions sont proches

des travaux de Chomsky & Halle (1968), qui feront l'objet de la discussion dans la partie suivante (1.5). En ce qui concerne Lilly & Viel, le système d'hiérarchie accentuelle qui est proposé est relativement complexe, avec quatre degrés d'accent, déterminés en fonction de critères rythmiques et acoustiques (et non en fonction de la hauteur mélodique, comme c'est le cas pour Huart (1984) et Boisson (1980)). Les accents primaires et secondaires correspondent à ceux des autres auteurs, car ils reposent sur des critères rythmiques, mais ce qui correspond à l'accent /0/ (l'accent faible) des autres auteurs est décomposé comme suit : un accent dit « nul » portant le degré /0/, et l'accent dit « tertiaire », portant le degré /3/. Dans l'édition la plus récente, l'accent /3/ n'est pas noté pour les raisons suivantes (Lilly & Viel, 1993 : 4) : « Nous n'avons pas noté l'accent tertiaire : on le distinguera de l'accent nul par l absence de réduction vocalique ». On en conclut que l'accent /3/ (ou accent tertiaire) sert à distinguer la syllabe non proéminente dont la qualité vocalique reste inchangée (il n'y a pas de réduction vocalique), par opposition à l'accent /0/ qui est réservé pour les syllabes avec une réduction vocalique. Par exemple, la terminaison *-ate* d'un verbe aura un accent tertiaire car elle ne se réduit pas alors que l'adjectif aura un accent nul car la voyelle se réduit.[17] Le fait de ne pas le noter nous fait penser que ce degré supplémentaire n'est pas d'un intérêt primordial : la question essentielle étant celle du rythme.

Ces systèmes de notation de la hiérarchie accentuelle soulèvent la question du rôle de la qualité de la voyelle dans la détermination des degrés d'accent. En

---

[17] Voir la rubrique suivante et la discussion de l'accent tertiaire : 1.4.1.5.

effet, la qualité vocalique, et surtout l'opposition entre la voyelle réduite / non réduite, sont des caractéristiques importantes en ce qui concerne l'accentuation. Toutefois, il nous semble que l'utiliser pour la détermination de la hiérarchie accentuelle, comme le font Giegerich et Lilly et Viel, est lui attribuer un rôle qui n'est pas le sien. La qualité vocalique dépend de l'accentuation, elle en résulte[18]; il faut donc déterminer la hiérarchie d'abord en fonction d'autres facteurs. Les systèmes faisant appel à la mélodie et au rythme nous paraissent plus appropriés.

1.4.1.5 Rythme et qualité vocalique : l'accent /3/ en question

Nous terminons ce débat sur la hiérarchie accentuelle en faisant à nouveau référence aux dictionnaires de prononciation, ou plutôt à l'unique exception dans l'inventaire de ces ouvrages. Il s'agit bien entendu de l'édition 1990 du LPD où Wells a introduit la notation d'un accent tertiaire (accent /3/).

Si nous retrouvons le même nombre de degrés chez Wells que chez Lilly & Viel ((1977) 1993), l'accent tertiaire ne s'emploie pas de la même manière. Wells le place dans une hiérarchie de degrés d accent et lui fait jouer un rôle plus ou moins rythmique. Lilly & Viel associent l'accent tertiaire à la seule qualité vocalique. Pour prendre un exemple concret, pour Lilly & Viel, la dernière syllabe de *deliberate* (verbe) porte un accent /3/ car la voyelle n'est pas réduite, contrairement à l'adjectif qui portera un accent /0/ puisqu'elle est réduite : le schéma accentuel du verbe est donc /0103/ alors que pour l'adjectif il est /0100/. Wells n'attribue aucun statut particulier au verbe, donc la terminaison porte un

---

[18] Voir 1.2, Fournier explique le point de vue de Guierre.

accent /0/ dans les deux cas, et il aura donc le schéma accentuel /0100/ comme l'adjectif.

Par contre, nous trouverons un accent /3/ dans des mots préfixés tels que *indefensibility* dans les deux ouvrages, mais il est attribué pour des raisons différentes. Chez Wells, ce mot aura le schéma accentuel /2030100/ parce que la suite de trois syllabes faibles entre l'accent secondaire et l'accent primaire va à l'encontre du rythme faible / fort en alternance et une certaine proéminence sera donc associée à la troisième syllabe. Bien que Wells ne mentionne pas la règle dérivationnelle, la présence de l'accent /3/ ici s'explique par l'isomorphisme, puisque la deuxième syllabe du radical *defense*, (comme le dérivé *defensible*), porte un accent primaire ; lorque le dérivé *defensibility* est formé à partir de la base *defersible*, l'accent secondaire se place sur la deuxième syllabe pour cette raison ; enfin, lorque le préfixe est rajouté, nous retenons cet accent sur la base, mais en diminuant sa valeur dans la hiérarchie accentuelle. Dans ce cas, il s'agit en fait d'un mot préfixé dont la base contient déjà un accent secondaire. L'accent /3/ est attribué en fonction d'arguments fondés sur le rythme. Nous retrouverons le même schéma chez Lilly & Viel, mais en raison de la qualité vocalique pleine de la troisième syllabe, et non pour des raisons rythmiques. De même, ils attribueraient le schéma /2310/ à *incantation* sur ces mêmes critères, alors que chez Wells, nous avons /2010/ pour ce mot car il n'y a pas de raison rythmique pour inclure un accent /3/. Chez Wells, cet accent ne se juxtapose pas aux autres accents forts dans des mots non préfixés et non composés

Dans les composés, Wells emploie cet accent dans des cas très précis. Si le composé porte un accent /1/ sur le premier élément, comme *blackbird*, Wells

attribue le schéma /10/, partant du principe que le deuxième élément d'un composé accentué sur son premier élément n'a pas besoin d'être marqué s'il est monosyllabique. Toutefois, si le composé accentué sur le premier élément a un deuxième élément plurisyllabique, comme *weightlifter*, Wells attribue le schéma accentuel /130/ pour des raisons rythmiques, car le deuxième élément comporte deux syllabes, et il est nécessaire de marquer laquelle de ces syllabes a une proéminence par rapport à l'autre. Dans le cas de *blackbird*, Lilly & Viel attribue le schéma /13/ en raison de la qualité vocalique pleine de la deuxième syllabe. Pour les mêmes raisons, ils noteront le schéma /130/ pour *weightlifter*.

Pour certains, la notation supplémentaire de Wells (1990) complique sans raison les indications accentuelles, pour d'autres, elle apporte une solution intéressante sur le plan intellectuel à certains problèmes de rythme (par exemple, la juxtaposition d'accents forts /1/ et /2/ dans des composés) et de différenciation et hiérarchisation des syllabes. Néanmoins, rappelons les avis de phonéticiens tels que Ladefoged (1993) en la matière, qui déclare que la différence entre accent /1/ et /2/ est déjà difficilement perceptible, alors comment justifier un accent /3/ ? Dans l'édition 2000, Wells a, du reste, abandonné la notation de l'accent /3/.

En fait, cette question nous intéresse de près car l'accent /3/ était utilisé, nous l'avons vu dans nos exemples, pour certains composés alors que le EPD15, dans les mêmes cas, propose un accent /2/ qui suit l'accent /1/ (ce qui n'est pas non plus conforme aux usages habituels pour les mots lexicaux non composés). La présence de ces usages exceptionnels pour indiquer l'accent suggère que pour cerner la question de l'accentuation sur les composés, il sera intéressant d'utiliser des appellations parfois un peu différentes de celles que l'on utilise

habituellement. Nous reviendrons sur ces questions quand nous définirons notre nomenclature et la hiérarchie accentuelle que nous utiliserons pour notre étude.

1.5 Statut de l'accentuation dans la théorie phonologique

Nous avons pu constater la grande diversité d'opinions en ce qui concerne la définition de l'accent ainsi que sa notation, ce qui reflète à la fois la complexité du phénomène et les partis pris théoriques différents. Nous passons à présent aux approches théoriques de la phonologie et particulièrement de la place de l'accent dans ces travaux, pour nous permettre de situer notre travail. Nous commençons notre étude par l'ouvrage de Chomsky & Halle, *The Sound Pattern of English* (1968), précurseur dans ce domaine.

1.5.1 L'approche de *The Sound Pattern of English* (SPE) (1968)

Dans la discussion de la hiérarchie accentuelle, le seul ouvrage cité dans le tableau (Figure 1.3) auquel nous n'avons pas fait référence est celui de Chomsky & Halle (1968), *The Sound Pattern of English* (dorénavant SPE). Leur traitement de ce phénomène est tellement imbriqué dans les bases mêmes de leur théorie qu'il faut aborder cette question en même temps que nous étudierons leur approche théorique.

Le point de départ de toute la théorie générative est qu'elle s'intéresse à des concepts linguistiques universels. La langue anglaise n'est donc qu'un exemple parmi d'autres et certains aspects de la langue sont davantage étudiés car ils permettent de révéler des fonctionnements qui ont une application générale. Ainsi, Chomsky & Halle (1968 : 162) annoncent :

> As we stated in the Preface, we are interested in an analysis of the facts of English insofar as this analysis serves to reveal general principles of linguistic structure. Thus

> we have been concerned with details of stress contour and vowel reduction because of the way they bear on such general notions as the principle of the transformational cycle. We are not, in this book, interested in these details for their own sake.

L'accent est donc étudié par rapport au cycle transformationnel. L'intérêt principal de cette approche est que nous nous intéressons au mot dans le discours et non seulement au niveau lexical. Ainsi, les « internal relations of stress» (Chomsky & Halle, 1968 : 23) sont mises en avant. L'approche est également très ancrée dans la syntaxe.

Deux grandes règles transformationnelles génèrent des contours accentuels : la *Nuclear Stress Rule* (NSR) et la *Compound Stress Rule* (CSR). La *Nuclear Stress Rule* opère au niveau du syntagme sur des séquences étiquetées [NP], alors que la *Compound Stress Rule* opère au niveau du composé sur des séquences étiquetées [N]. Elles fonctionnent de la manière suivante (Chomsky & Halle, 1968 : 92) :

> [NSR] assigns primary stress to the second sonority peak, reducing by one all other stress levels in the phrase under consideration; the Compound rule, on the other hand, assigns primary stress to the first of the two peaks, reducing all other stresses by one.

L'attribution ou non d'un accent (syllabe ayant le marqueur [+stress]) est fonction du type de syllabe (une syllabe lourde porte l'accent, une syllabe légère n'en porte pas). Ce typage dépend en grande partie de la qualité vocalique. En fait, les syllabes susceptibles de porter l'accent correspondent à celles contenant une voyelle tendue ou diphtonguée (ou finissant par deux consonnes dans certaines circonstances).

Goyvaerts & Pullum (1975 : 201) dans leur introduction au chapitre sur l'accent expliquent de manière très synthétique comment les règles fonctionnent. Ils expliquent qu'elles sont ordonnées, avec la CSR qui s'applique au [N] et la

NSR s'appliquant à [NP]: «applying cyclically (to progressively larger constituents up to the full phonological phrase) to the syntactically bracketed surface structure». Ils précisent que la théorie repose sur l'hypothèse suivante: « Main stress position is to be predicted from segmental composition and syntactic constituency. » C'est pourquoi chaque mot doit être étiqueté en fonction de sa catégorie syntaxique.

En ce qui concerne la hiérarchie accentuelle, elle se situe à deux niveaux d'analyse bien distincts. Au niveau profond, il n'y aurait que deux accents: un fort et un faible. Chomsky & Halle (1968) ne se préoccupent pas à ce niveau des problèmes de la réalisation effective ou de la perception, car l'analyse se fait au niveau profond, avant même la production (qui représente le niveau de la surface). Voilà pourquoi ils peuvent proposer un système aussi simple. Par contre, au niveau de la surface, le nombre de degrés peut être très élevé, car, à chaque passage du cycle, l'accent est attribué (ou révisé), donc plus la phrase phonologique est longue, plus le nombre de degrés est important.

Chomsky & Halle (Chomsky & Halle, 1968: 20-21) s'appuient, par exemple, sur les composés *blackboard* et *board-eraser* qui sont opposés au syntagme *black board (eraser)* ou *black* fonctionne en tant qu'adjectif qualificatif. Les composés sont soumis à la *Compound Stress Rule* (CSR) puisqu'ils sont étiquetés $[\text{blackboard}]_N$ et $[\text{board eraser}]_N$[19]. L'accent /1/ est donc attribué au

---

[19] Voir Chomsky & Halle, 1968: 20-21. Ici nous avons simplifié la démonstration, comme le font Halle & Keyser (1971). Chez Chomsky & Halle, chaque lexème porte une étiquette ce qui implique un nombre beaucoup plus important de parenthèses.

premier élément dans les deux cas. La valeur de la deuxième syllabe de *eraser* n'est pas mentionnée dans cette démonstration puisque les règles de la CSR et de la NSR s'appliquent à partir du mot. Le schéma accentuel donné pour chacun de ces composés à deux éléments est donc /1-2/. Comme les règles sont cycliques, une de leurs forces est leur capacité à traiter les surcomposés et les syntagmes, et à les désambiguïser le cas échéant. Ainsi, si nous avons *black board-eraser*, ou *blackboard eraser*, selon le système de parenthèses, l'accent sera attribué différemment selon s'il s'agit d'un effaceur de tableau de couleur noire ou bien d'un effaceur pour tableaux noirs, parce que les règles opéreront différemment. Dans le cas de [black [board-eraser]$_N$]$_{NP}$, le schéma de [board-eraser]$_N$ est attribué au premier passage du cycle parce que les parenthèses sont les plus imbriquées: le schéma est /1-2/ car c'est l'élément de gauche qui se voit attribuer l'accent /1/ selon la CSR ; puis, au cours du second cycle, c'est la NSR qui opère puisque les parenthèses sont étiquetées [NP] ; c'est donc l'élément de droite qui se verra attribuer à nouveau l'accent /1/, donc l'accent primaire reste sur *board*, mais la valeur de l'accent sur l'autre élément, *eraser*, sera réduit d'un degré (donc *board-eraser* aura le schéma /1-3/), l'élément de gauche (*black*) ayant un accent inférieur d'un degré (/2/). La séquence dans son ensemble aura le schéma /2-1-3/.

Dans le second cas, [[blackboard]$_N$ eraser]$_N$, comme la structure est différente, le résultat le sera aussi. Le premier cycle attribue le schéma /1-2/ à *blackboard*, selon la CSR qui attribue l'accent /1/ à l'élément de gauche. Au second tour du cycle, c'est encore la CSR qui opère, donc l'accent /1/ est encore attribué à l'élément de gauche, en réduisant les autres valeurs d'un degré

(*blackboard* a le schéma /1-3/, *eraser* a la valeur /2/). La séquence dans son ensemble porte le schéma /1-3-2/.

Cette approche était révolutionnaire à son époque car elle posait des règles pour prédire des contours accentuels et tentait de démontrer le processus créatif du langage. Cependant, beaucoup de critiques lui ont été adressées. Tout d'abord, l'intérêt de l'universel prend le dessus au point que des variantes et des exceptions qui ne correspondent pas au moule théorique sont considérées comme étant en dehors du débat. Chomsky & Halle parlent du phénomène de l'accent dans la citation ci-dessus comme étant un *détail*. Beaucoup d'exemples ne sont valables que pour l'anglais américain (les mots en *-ory*[20] qui seraient accentués en anglais américain, mais qui se réduisent en anglais britannique, *baboon* dont la première syllabe serait accentuée en anglais américain, mais se réduit en anglais britannique). Les règles du cycle transformationnel précèdent les règles concernant la phonologie du mot; cependant, les règles du cycle reposent sur les valeurs sonores (supposées) des segments qui font partie des mots. Enfin, le fait que le cycle n'ait pas de limites aboutit à des contours accentuels invraisemblables. A chaque passage du cycle, le degré d'accent se réduit d'un degré, ce qui aboutit à un nombre théoriquement illimité de degrés. Dans l'exemple que nous avons cité, nous constatons que *blackboard* a le schéma /1-2/ isolément, mais /1-3/ lorsqu'il entre dans un surcomposé. Si la phrase phonologique était plus longue, les valeurs changeraient à chaque passage du

---

[20] Chomsky & Halle, 1968: 91-3. La démonstration de la règle CSR *chemistry research laboratory* ne peut pas être reprise pour l'anglais britannique puisque le schéma accentuel de *laboratory* est différent.

cycle. C'est le point de la théorie qui a été le plus souvent critiqué. Ainsi Lee (Goyvaerts & Pullum, 1975 : 244) attire l'attention sur le fait que, d'après les prédictions des règles de SPE, les relations internes entre les différents éléments d'un composé seront changées selon la longueur de la phrase :

> This prediction that there should be a radical difference between the contours of a compound when said in isolation and when said as part of a larger construction seems to me to be quite wrong.

Il ajoute (Goyvaerts & Pullum, 1975 : 245):

> I have predicted that it should be impossible to maintain an intonational contrast between two compounds of different internal structure; e.g. *bush-black-bird* (i.e. a black-bird found in bushes) and *black-bird-food*. I think this matter is unclear, but if such a contrast does exist I suspect that it can be easily neutralized.

C'est un peu expéditif car nous avons là un des points intéressants de l'interprétation des composés. Nous y reviendrons.

La complexité de ce système a aussi été critiquée par Boisson (1980, 196) qui dit :

> Chomsky & Halle 1968 [propose un] nombre excessif de degrés d'accent (c'est un peu curieux d'y voir seulement deux degrés d'accent en profondeur, mais un nombre de degrés superficiels qui n'est pas perceptuellement plausible).

C'est l'effet pervers de leur approche qui privilégie l'abstrait au profit de l'intérêt général et au détriment de considérations pratiques. Ils insistent pourtant sur la simplicité de leurs règles par opposition à la complexité des réalisations, mais nous pouvons nous demander si ces réalisations ne sont pas bien plus simples que leurs règles ne le laissent supposer.

Chomsky & Halle ont conscience du problème, puisqu'ils proposent (Chomsky & Halle, 1968 : 23) de limiter le nombre d'applications du cycle et

formuler un principe pour l'interprétation phonétique qui neutralise certaines distinctions[21].

Par ailleurs, les règles dépendent d'un étiquetage syntaxique qui devra produire le contour souhaité. Ceci n'est pas non plus sans problèmes car on ne voit pas clairement selon quel critère *'Christmas cake* sera traité comme un composé et, *Christmas 'pudding* comme un syntagme. C'est une des questions que nous devons nous poser dans notre analyse, mais si l'étiquetage est attribué en fonction du résultat souhaité, cette solution présente le risque d'un raisonnement circulaire. Halle & Keyser (1971 : 59-60) reprennent ces règles et tentent de résoudre ce genre de problème en fondant la CSR et la NSR en une seule règle, mais leur fonctionnement reste tributaire d'un étiquetage syntaxique.

Quant à la réduction vocalique, elle intervient en même temps que l'accent, et elle opère à la fin du cycle transformationnel. Il est intéressant de noter que dans cette analyse, une fois réduite, la voyelle est irrécupérable. Pour arriver au bon résultat, des règles qui s'appliquent une fois, en fin de cycle, sont mises en application. Pour ne pas se réduire, la syllabe doit être marquée d'un accent quelconque, mais parfois, l'on attribue [-tense, -stress] pour permettre la réduction de certaines voyelles. Inversement, ils ajoutent des règles qui attribuent [+ stress] à des syllabes ayant [- stress] pour les empêcher de se réduire. Ces ajustements (appelés précisément Adjustment Rules) sont certainement la partie la moins convaincante de la théorie.

---

[21] La formulation est très imprécise : « distictions that go beyond a certain degree of refinement ».

Retenons pour nos travaux deux axes intéressants : d'abord, le parti pris théorique que les composés s'accentuent de manière générale sur le premier élément (le principe de base de la *Compound Stress Rule*), ce qui s'oppose au comportement des syntagmes qui, eux, s'accentuent à la fin, sur la tête du syntagme. Toutefois, les règles de SPE sont conçues de sorte qu'elles attribueront un accent au deuxième élément d'un composé (considéré comme tel par SPE, donc avec accent primaire sur le premier élément) même si celui-ci est monosyllabique, donc *blackboard* a le schéma /1-2/. C'est une question à laquelle nous reviendrons. Ensuite, les règles fonctionnent au niveau de la phrase phonologique et l'on s'intéresse aux éléments par rapport à leur fonction dans la phrase plutôt qu'en tant qu'éléments lexicaux isolés. Ce type d'analyse peut fournir des indications à développer en ce qui concerne des phénomènes de discours tels que le déplacement de l'accent.

1.5.2 La phonologie métrique

The Sound Pattern of English de Chomsky & Halle, paru en 1968, était un des premiers ouvrages à s'intéresser à la phonologie accentuelle de l'anglais. Cet ouvrage a eu une influence considérable sur le domaine de la phonologie jusqu'à aujourd'hui. Cette première approche de la phonologie s'intègre dans le contexte de la théorie générative qui cherche à rendre compte du processus créatif des phénomènes langagiers. Giegerich (1992 : 310) rappelle la puissance de ces théories, même si quelques réserves sont exprimées :

> Generative phonologies are powerful. Abstract underliers, feature-changing rules, rule ordering, deletion rules, insertion rules, reference to non-phonological information – such a battery of devices adds up to a phonological theory that seems able to account for almost anything, even if in a way that is at times rather less than insightful.

Il fait bien sûr référence au niveau d'abstraction qui résulte souvent de leur souci d'universalité et de simplification, d'où le décalage entre des niveaux d'analyse profonde (simple) et en surface (avec des représentations complexes) auquel nous avons fait référence plus haut.

Depuis l'époque de SPE, les règles ont été affinées et d'autres voies ont été explorées qui s'intègrent dans le courant de la grammaire générative pour améliorer cette première tentative. Ces nouvelles théories tentent, elles aussi, d'exprimer formellement des généralisations qui s'appliquent en même temps à la phrase phonologique et au mot. Les chercheurs se sont intéressés à une des failles du système de SPE. Jensen (1993 : 15-16) explique le sens de ces recherches :

> [...] stress was an n-ary feature [...] Not only was stress n-ary, the number of possible values for this feature was unlimited. The reason is that stress is assigned to words and phrases, and in principle there is no upper limit to the length of words or phrases.

La théorie générative actuelle qui présente le plus d'intérêt pour l'analyse de l'accent est la phonologie métrique. Deux représentations formelles sont proposées pour représenter les caractéristiques de l'accent : l'arbre métrique et la grille métrique. Kager (Goldsmith, 1995: 368-9) explique :

> In the metrical tree, stress is represented as a hierarchy of binary branching structures, each of which is labeled *strong-weak* (*sw*) or *weak-strong* (*ws*) [...] Stress, as represented in the metrical tree, is a *relational* property: a node is strong only by virtue of the fact that it is the sister of a weak node [...] Liberman (1975) introduced the metrical grid as a representation of rhythmic structure.

Ces deux représentations ont chacune leur intérêt puisque l'arbre permet de montrer la hiérarchie accentuelle alors que la grille (introduite par Liberman) montre clairement les problèmes de juxtaposition d'accents forts (*stress clash*) et donc sert pour analyser le phénomène du déplacement de l'accent (*stress shift*).

Si nous reprenons notre exemple, *blackboard eraser* sera représenté dans une grille métrique de la manière suivante :

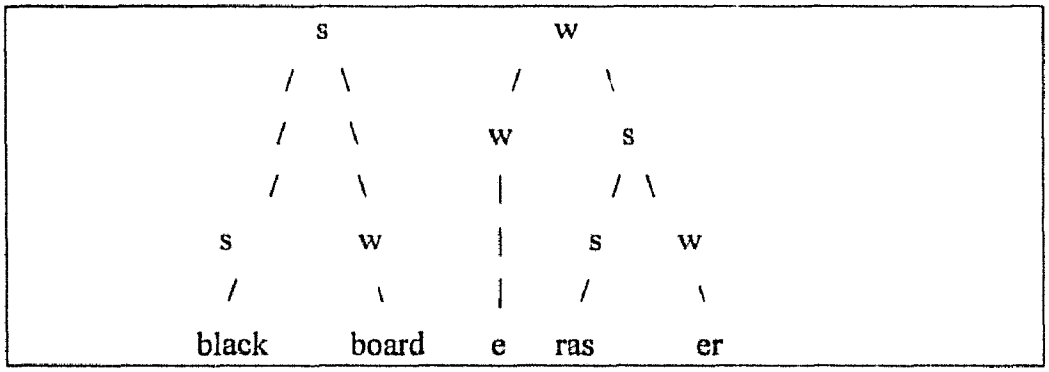
Figure 1.5 : Grille métrique : *black board eraser*

X					Ligne 2
X			X		Ligne 1
X	X	X	X	X	Ligne 0
black	board	e	ras	er	

Dans le tableau ci-dessus, la syllabe qui est associée aux trois croix (qui comporte une croix en lignes 0, 1 et 2) correspond au noyau intonatif : ici *black* ; les autres syllabes proéminentes sont associées à deux croix (donc comportent une croix en lignes 0 et 1) : ici, la deuxième syllabe de *eraser*. Le tableau montre qu'il n'y a pas de conflit car les syllabes accentuées sont séparées par des syllabes non accentuées, ce qui est représenté par le fait qu'il n'y a pas de croix qui se suivent en ligne 1, comme l'explique Selkirk (Goldsmith, 1995 : 564). Dans le cas où des croix se suivent en ligne 1, il y a des accents adjacents (et donc « *stress clash* »). Dans une phrase phonologique plus longue, la pertinence de cette représentation apparaît pleinement, car elle permet de voir très clairement le rythme de la phrase, la hiérarchie accentuelle, et les éventuels cas de conflit (s'il y a des accents forts adjacents). C'est une représentation fort utile pour tenter d'expliquer des phénomènes de discours tels que le déplacement accentuel (*stress shift*). Nous notons que, contrairement à SPE, plusieurs syllabes peuvent se trouver au même niveau de proéminence relative puisqu'il n'y a que trois lignes.

Sous forme d'arborescence, le graphique permet de visualiser l'ensemble comme une suite de pieds forts et faibles en alternance, comme suit :

Figure 1.6 : Arbre métrique *black board eraser*



Nous voyons la relation entre les éléments, mais dans cet exemple, comme l'arbre est binaire, ces relations sont moins claires que dans la grille. Toutefois, les deux représentations visent à rendre compte d'une certaine vue de l'accent ; Goldsmith (1990 : 170) résume :

> What the two versions of metrical theory share in common is the idea (i) that the study of stress is different in kind from the study of other phonological characteristics; (ii) that it is fundamentally a study of the relative prominence of syllables and higher-level units, such as the foot; (iii) that the most important characteristics determining stress patterns are rhythm (i.e. alternating prominence) and sensitivity to inherent syllable (or rhyme) weight; and (iv) that stress representations are hierarchical. The precise formal way in which the hierarchical nature of stress is expressed differs in the two approaches, but both share the goal of expressing the hierarchical relationship found among syllable rhymes, secondary stress patterns, primary word patterns, and ultimately phrase-level prominences.

Nous retrouvons toujours ce souci d'arriver à un formalisme qui puisse s'appliquer au niveau de la phrase phonologique. La théorie s'appuie sur l'unité du pied auquel nous reviendrons dans le chapitre 3. Notons également le principe (i) qui met l'étude de l'accent sur un plan différent des autres caractéristiques phonologiques. Liberman est à nouveau cité par Halle & Idsardi (Goldsmith, 1995 : 403)

> The fundamental insight into stress systems remains that expressed by Mark Liberman in his 1975 dissertation. Liberman suggested that stress is not a simple phonetic feature, as had been assumed by most linguists and phoneticians, but rather that it is a phonetic means for marking various kinds of groupings of linguistic elements.

Nous l'avons déjà vu chez Troubetzkoy (Troubetzkoy, 1949 : 312). Que l'accent ait un véritable rôle de démarqueur est un point de vue auquel nous souscrivons, mais il nous reste à déterminer de quel type de démarquage il s'agit en ce qui concerne les composés. Dans leur conclusion (Goldsmith, 1995 : 439), ils déclarent :

> Once these groupings have been established, greater prominence is supplied to certain elements in the group than to the others. The prominence is thus a by-product of the grouping of elements into constituents.

Nous retrouvons le parti pris théorique de SPE qui est fondé sur la syntaxe. Nous émettons des réserves vis-à-vis de ces propos car il nous semble que la proéminence est un phénomène qui fait partie intégrale de l'accent et qui est attribuée au départ et non, comme le suggèrent Halle & Isardi, presque par accident. Cependant, le phénomène de déplacement de l'accent pourrait être analysé ainsi car il y a effectivement un changement de la hiérarchie accentuelle, qui modifie la valeur de départ de l'accent.

1.5.3 L. Guierre et ses successeurs

Les théories que nous venons d'étudier reposent essentiellement sur un souci de généralisation qui utilise des systèmes formels (parfois abstraits) pour expliquer les phénomènes langagiers. Les exemples donnés sont précis, mais limités, et sont tirés de toutes les langues; l'anglais n'étant qu'un exemple. Bien au contraire, la théorie de l'accentuation exposée par Guierre (1979) utilise une méthode différente qui repose sur une analyse faite d'après un vaste corpus lexical de l'anglais contemporain. C'est une des raisons pour lesquelles nous nous intéressons particulièrement à son analyse.

Nous avons déjà vu dans l'introduction de ce chapitre que, pour Guierre, l'accent est une des caractéristiques les plus importantes de la langue anglaise. Dans son analyse de l'accent, la hiérarchie accentuelle qu'il utilise correspond à ce qu'utilisent les dictionnaires de prononciation. Il faut préciser que les arguments de sa thèse sont le résultat d'une analyse de la 12ème édition de EPD. Il utilise un système de chiffres pour désigner les degrés d'accent, ce qui clarifie nettement la hiérarchie accentuelle. Dans sa thèse (1979), l'accent primaire correspond au chiffre /2/, l'accent secondaire à /1/ et le non-accent à /0/. Cette numérotation implique une hiérarchie qui monte en puissance vers l'accent primaire /0/-/1/-/2/. Par la suite, la numérotation a été inversée, ce qui permet la correspondance entre la nomenclature que nous utilisons pour désigner l'accent primaire /1/ et l'accent secondaire /2/ et, le cas échéant, cela permet d'introduire l'accent tertiaire /3/.

L'approche de Guierre se démarque de celle de SPE où l'accent est placé en premier, à chaque tour du cycle, en fonction du type de la syllabe (qui dépend très largement de la qualité vocalique, comme nous l'avons vu). SPE distingue seulement deux types de voyelles qui leur permettent d'attribuer l'accent : voyelles tendues (attirant l'accent) et voyelles relâchées (ne portant pas d'accent). Au contraire, Guierre (1979 : 387) quant à lui conteste le bien-fondé de ces règles en faisant intervenir un troisième type de voyelle : la voyelle réduite. Il écrit :

> La question, selon nous, n'est pas de savoir si le relâchement de la voyelle en dernière syllabe implique /10/; cette voyelle, on l'a dit, n'est qu'exceptionnellement « relâchée » ; quand elle n'est pas tendue, elle est presque toujours réduite. Comment, dès lors, se fonder sur le caractère réduit de la voyelle post-tonique pour engendrer l'accentuation /10/, alors que la réduction, elle-même, résulte essentiellement de l'accentuation.

D'après Guierre, le raisonnement de SPE n'est pas convaincant, voire circulaire, puisque ce qui est utilisé pour déterminer l'accent est en réalité le résultat. La voyelle est réduite non pas parce qu'elle est de nature relâchée, mais parce qu'elle se trouve en position non accentuée. Si, comme Guierre, nous faisons la distinction entre voyelle relâchée (qui est une qualité, une valeur en soi) et voyelle réduite (qui est le résultat d'un processus, et non une valeur en soi), nous sommes amenée à adhérer au même point de vue.

Pour Guierre, l'accent est déterminé avant la valeur des segments, et c'est l'attribution de cet accent en premier lieu qui peut ensuite influencer très largement le segment en entraînant ou non une réduction vocalique. Selon Guierre (1979 : 564), la non-accentuation d'une syllabe conduit souvent à une réduction de la voyelle non accentuée. « Les règles d'accentuation précèdent les règles d réduction ». Cet argument nous paraît très puissant, et c'est pourquoi, comme Guierre, nous partirons du principe que **le placement de l'accent lexical précède la détermination phonétique des sons**.

Nous allons étudier les règles de l'accentuation élaborées par Guierre et qui tiennent compte de ce parti pris théorique. Notons que, pour Guierre, l'accent est tantôt considéré comme un état (sur le radical) tantôt comme un processus (dans le dérivé). Nous situerons ses règles par rapport aux composés et il sera intéressant de poser la question de savoir si l'accent des composés fonctionne comme un état ou comme un processus.

1.5.3.1 Les règles générales de l'accentuation de L. Guierre

Guierre (1979 : 100) dans sa thèse d'Etat pose d'abord le problème de ce qu'il appelle la grammaticalité de l'accentuation. L'accent relève-t-il ou non de la grammaire ? « Cette donnée peut appartenir à la grammaire si elle suit des règles, ou au lexique si l'accentuation est rarement ou jamais prévisible ». Pour lui, elle fait partie de la grammaire car elle suit effectivement des règles : elle est dans une large mesure prévisible.

Les règles qu'il propose sont des règles d'abord contextuelles, s'appuyant sur ce qu'il appelle la grammaire de lecture. Il a élaboré une série de règles graphophonématiques et rappelle l'importance de les ordonner (Guierre, 1979, 58) : « les règles de réduction sont postérieures aux règles assignant l'accent, et concernent aussi bien les voyelles tendues que les voyelles relâchées ». Il fait également remarquer que les classes productives sont aussi les plus régulières, remarque qu'il sera intéressant de tester lors de notre analyse des composés.

Le second point fait référence à l'évolution historique de la langue. Il s'agit de la règle de la « rétraction germanique ». L'accent a tendance à remonter vers le début du mot (Guierre, 1979 : 95) :

> L'un des principes directeurs auxquels on fait appel en matière d'accentuation et que confirme l'évolution historique de l'anglais depuis plusieurs siècles, est que l'accent a tendance à remonter vers la syllabe initiale (tendance germanique), sauf si cette remontée est bloquée.

C'est ce que Kingdon appelle «*the Teutonic rule*» (Kingdon, 1958b, 30). Dans sa conclusion, Guierre dit même : « S'il faut user de formules lapidaires, nous dirions volontiers que l'anglais contemporain est une langue latine (85% de son lexique) que l'on prononce avec un accent germanique ». (Guierre, 1979 : 784).

L'idée est reprise par Duchet (1998) qui propose en fait deux logiques pour le placement de l'accent en anglais : la rétraction iambique d'une part mais aussi l'isomorphisme. Cette tendance forte explique pourquoi l'accent de certains mots ne remonte pas, parce que l'accent est placé en fonction de la morphologie du mot par analogie avec d'autres de même type et se trouve ainsi « bloqué» dans une autre position. Des conflits sont d'ailleurs toujours possibles et des variantes accentuelles courantes sont souvent dues à ces conflits. Comme Guierre, Duchet donne à la morphologie une place importante.

Ayant posé ces principes généraux, Guierre (1979) propose quatre règles. Nous allons examiner quelles sont ces règles, comment elles sont reprises par d'autres auteurs et enfin comment elles s'appliquent aux composés.

Guierre (1979 : 319) nous rappelle que : « Les contraintes rythmiques limitent de façon drastique le nombre de schémas accentuels licites », ayant postulé ses quatre règles de base (1979 : 317) :

règle 1 Tout mot comporte un et un seul accent primaire.
règle 2 Aucun mot ne commence par une séquence de deux accents /0/.
règle 3 Les séquences de deux accents forts sont non grammaticales,
(sauf composés et préfixés à préfixe séparable).
règle 4 La voyelle réduite ne peut apparaître que sous accent /0/,
(mais la voyelle en /0/ n'est pas forcément réduite).

Nous allons commenter ces quatre règles et déterminer leur signification au niveau des composés.

1.5.3.1.1 Règle 1

Tout mot comporte un et un seul accent primaire.

Comme nous l'avons signalé (1.1.2.5), pour la règle 1, on voit que par rapport à Poldauf (1984), il y a désaccord puisque Poldauf rejette l'idée de l'existence d'un accent sur un mot monosyllabique Toutefois, dans le cadre de notre travail, ce débat n'est pas utile car tout composé comporte forcément au moins deux éléments, donc il n'y a pas de composé monosyllabique.

1.5.3.1.2 Règle 2

Aucun mot ne commence par une séquence de deux accents /0/.

L'importance de la règle 2 avait déjà été remarquée par Kingdon (Kingdon, 1958b : 16) : « [...] all words must have a stress on one of the first two syllables ».. Elle est incontournable, qu'il s'agisse de radicaux, de dérivés ou de composés, à tel point qu'elle peut servir de test d'erreur pour un fichier contenant des schémas accentuels. La séquence initiale /00/ indique de manière infaillible une erreur. Duchet (1994), à son tour, dans *Code de l'anglais oral*, rappelle la même règle. Notons que c'est elle qui explique la nécessité d'un deuxième accent dans certains mots, car s'il y a un accent sur la troisième syllabe ou au-delà, il faut introduire un autre accent en début de mot, que l'on appelle accent /2/ habituellement. Nous avons déjà fait référence à cette règle dans la discussion concernant les degrés d'accent.

1.5.3.1.3 Règle 3

Les séquences de deux accents forts sont non grammaticales.

C'est certainement la règle 3 qui nous intéresse le plus. C'est elle qui est responsable du fait que l'on considère l'anglais comme «*a stress-timed language*», car l'interdiction de juxtaposer deux accents forts implique la présence intermédiaire d'accent(s) faible(s), que nous avons appelés jusqu'ici, accent /0/. Duchet (1998) insiste également sur la règle 3 dans l'ouvrage cité ci-dessus, rappelant que les schémas /21/, /12/ ou /11/ sont des schémas illicites *a priori*. Kingdon (1958b) indique l'existence de «*adjacent stress* » dans le cas de préfixes accentués (*de- mis-*) et fait la même remarque sur ce type de schéma (Kingdon, 1958b : 18) : « (they) violate the rhythmic principle ».

L'idée est formulée par Deschamps dans son "Introduction phonétique" (Bouscaren, C. et al., 1995 : vii) :

> Cet accent secondaire ne peut être sur la syllabe qui précède l'accent principal. Il y aura donc toujours une syllabe inaccentuée entre l'accent secondaire et l'accent principal [...][ˌ–ˈ-], sauf mots composés et mots à préfixes séparables.

Notons l'exception que constituent les composés, la même que pour Guierre (1979). Marquer un intérêt particulier pour cette règle n'est pas incohérent avec cette restriction, bien au contraire. Il s'agit d'un des phénomènes marquants des composés qui ne fonctionnent pas comme les autres mots de la langue. Dans les exemples que nous avons pris (voir 1.5.1), la pertinence de cette règle n'apparaît pas clairement car dans les deux cas (*ˈblackboard* et *ˈboard-eˌraser*), les syllabes accentuées sont séparées. Par contre, si nous prenons l'exemple de *fruit market*, ce composé sera accentué /1-20/ par Guierre, et /1-2/ par SPE. Dans la grille métrique, nous aurons la représentation suivante :

Figure 1.7 : Grille métrique *fruit market*

X			Ligne 2
X	X		Ligne 1
X	X	X	Ligne 0
fruit	mar	ket	

Nous avons donc un cas d'accents adjacents, puisque la ligne 1 comporte deux croix consécutives. Le phénomène est encore plus clair lorsque l'on considère les composés qui s'accentuent sur leur deuxième élément : *red tape* (au sens de « tracasseries administratives ») où l'on aura un accent secondaire qui précède l'accent primaire sans qu'une syllabe faible ne s'intercale. Il est vrai que dans la théorie générative, ces composés ne sont pas traités comme tels, et suivent les règles qui s'appliquent aux syntagmes.

Dans son analyse des préfixés et suffixés, Fournier (1992 : 74) attire l'attention sur le même phénomène, et insiste sur le corollaire qui suppose que, si deux accents forts ne doivent pas se suivre, les syllabes environnantes seront affectées par ce principe :

> On observera en effet que ce principe suppose la *désaccentuation des syllabes voisines d'une syllabe accentuée, et donc la concentration de l'information dans la syllabe accentuée aux dépens des syllabes voisines, un phénomène amplifié par la réduction fréquente de ces syllabes inaccentuées *Noter qu'il s'agit d'un état et non d'un processus.

Voilà donc une explication claire des effets de cette règle sur les syllabes voisines. Il fait aussi référence à la réduction vocalique qui laisse prévoir la règle 4.

1.5.3.1.4 Règle 4

La voyelle réduite ne peut apparaître que sous accent /0/.

Enfin, la règle 4 reprend une idée fondamentale dans la théorie de Guierre (1979), à savoir que la réduction vocalique est le résultat du placement de l'accent. Elle est formulée différemment par Lilly & Viel qui n'ont pas la même

approche (Lilly & Viel, (1977) 1993 : 49), mais c'est le même phénomène qui est en cause :

> Principe général. Lorsqu'une voyelle reçoit un accent nul, elle subit une transformation appelée 'réduction'. Au contraire, toute voyelle qui reçoit un accent non nul, si faible soit-il, n'est pas affectée par la réduction.

On voit à nouveau pourquoi ces auteurs ont besoin d'utiliser plusieurs degrés d'accent, puisque pour eux l'accent /0/ se confond avec la réduction vocalique. Leurs conclusions en ce qui concerne les voyelles susceptibles de réduire sont également précisées (Lilly & Viel, (1977) 1993 : 50) :

> Seules les voyelles tendues représentées dans l'orthographe par une voyelle simple, c'est-à-dire dérivées d'une voyelle relâchée par tension sont susceptibles de réduire. Exemple: *to deliberate* = di'libə,reit ~ *deliberate* (adjectif) = di'libərət).

Ce point de vue n'est pas partagé par Guierre (1980) qui part du principe que toutes les voyelles peuvent subir une réduction. Cependant, tous s'accordent sur le fait que seule une syllabe avec un accent /0/ peut subir une réduction vocalique. Cependant, pour Guierre, l'accent /0/ peut entraîner la réduction vocalique, alors que pour Lilly & Viel, la voyelle correspondant à l'accent /0/ subit forcément la réduction (voir la discussion sur la hiérarchie accentuelle en 1.4).

En ce qui concerne notre travail, il faut bien insister sur le comportement atypique des composés puisque ce qui tend à les caractériser, c'est que les syllabes non accentuées ou moins accentuées ne subissent pas de réduction vocalique, alors que, dans d'autres mots de la langue, les mêmes voyelles subiraient

effectivement une réduction[22]. En effet, dans l'exemple de *blackboard*, le second élément non (ou moins) accentué ne se réduit pas. Il garde son identité, contrairement au composé historique, *cupboard*, qui n'a plus que l'orthographe pour nous rappeler ses origines, puisqu'il a subi une réduction vocalique (et même la consonne du premier élément est élidée); il n'est plus ressenti comme un composé, d'ailleurs[23]. La question de la non-réduction de l élément non accentué d'un composé sera étudiée attentivement dans l'analyse.

1.5.3.1.5 La règle de Fournier

Enfin, Fournier (Fournier, 1992 : 74) nous propose une autre règle qui sera à tester dans l'analyse des composés : « il n'y a pas d'accent à droite de l'accent principal ». Cela revient à dire que l'accent /2/ n'est possible que sur l'une des deux premières syllabes du mot, pour respecter la règle 2. Il n'apparaîtra à aucun autre emplacement. *A priori*, cette règle pourrait bien subir les mêmes restrictions que la règle 3 si l'on se réfère aux propos de Roach & Hartman dans leur introduction à la 15e édition du EPD (Jones, 1997 : xii). L'accent /2/ est indiqué après l'accent /1/ dans certains cas bien précis. « Such a secondary stress is only marked in closed or hyphenated words where the second element is polysyllabic (e.g. *ˈfish,monger*). On peut trouver d'autres exemples : le quasi-lexème *ˈtele,vision* et le composé *ˈbank ,robber*. Ces cas rappellent *ˈfish,monger* et *ˈfruit*

---

[22] Nous nous référons aux deuxièmes éléments monosyllabiques des composés à accentuation simple.

[23] Voir Duchet, 1993 : 246.

ˌmarket, qui ont des accents adjacents, ce qui va à l'encontre de la règle 3 en plus de la règle de Fournier.

1.5.3.2 Les règles qui concernent les composés

Nous sommes donc concernée par les règles 2, 3 et 4 et la règle de Fournier que nous appellerons la règle 5. Nous récapitulons :

2 Le schéma /00/ est interdit.

3 Les schémas /11/, /21/ sont interdits.

4 S'il y a un accent /1/ ou /2/, la voyelle ne peut pas être /ə/.

S'il y a un accent /0/, la voyelle peut être /ə/.

5 Le schéma /1(0)2/ est interdit.

Des réserves concernant les composés ont déjà été formulées en ce qui concerne la troisième règle. Quant à la cinquième, nous émettons la même réserve pour les composés et les quasi-morphèmes. Nous y reviendrons dans notre analyse ou nous étudierons avec attention les composés par rapport à ces règles. Ce sera peut-être un moyen de définir un certain type de composés par la grammaire orale.

Dans tous les cas, le locuteur natif n'a aucun mal à appréhender ce système rythmique. Il faut croire que l'appellation de Boisson (1980 : 755) est très juste. Il parle du « homo computator» qui est capable en toute circonstance de calculer. L'image est-elle bien choisie ? S'agit-il d'un calcul ? En tout cas, c'est un phénomène auquel on est particulièrement sensible quand on regarde le comportement de l'accent en discours.

1.6 L'accent dans le discours continu

C'est le dernier point que nous voulons considérer avant d'aborder la conclusion de ce chapitre. Il est tentant d'étudier le lexique comme si les mots se prononçaient toujours pris isolément, d'une part parce que nous disposons d'informations au niveau du mot isolé que l'on peut traiter de manière claire, d'autre part parce que l'étude du discours continu implique un grand niveau de complexité parce que les mots sont sujets à l'influence du contexte intonatif, accentuel et sonore et que l'analyse risque d'aboutir à des résultats complexes, voire confus. Nous avons d'ailleurs déjà fait la distinct on entre les différents types d'accent.

A la suite de cette présentation des règles de l'accentuation, quelques remarques sur le comportement de l'accent lexical dans le discours continu s'imposent. Dans l'étude des règles génératives, ce phénomène a déjà été abordé, puisque celles que nous avons étudiées fonctionnent au niveau du mot et, au-delà, de la phrase phonologique. Nous savons donc que la hiérarchie accentuelle à laquelle nous avons fait référence est susceptible d'être changée en fonction du contexte accentuel de l'énoncé. Ce type de changement se fait tout naturellement par notre «homo computator» mais le fait que de tels changements existent pose quelques problèmes aux éditeurs de dictionnaires de prononciation. En effet, il ne faut pas perdre de vue que les mots vont se dire dans un contexte et lorsqu'il y a des constantes quand on passe du mot isolé au mot en contexte, il ne faut pas manquer de les signaler. C'est ce problème, ainsi que ceux de la variation et de la hiérarchie accentuelle, que soulèvent Roach & Hartman dans leur introduction à la 15e édition du EPD (Jones, 1997 : xii) :

> Stress patterns present one of the most difficult problems in a pronouncing dictionary. One reason for this is that many polysyllabic words have more than one possible stress pattern, and one must consider carefully which should be recommended. Secondly, the stress of many words changes in different contexts, and it is necessary to indicate how this happens. Thirdly, there is no straightforward way to decide on how many different levels of stress are recognizable.

Lorsque l'on parle, certaines distinctions que l'on fait sur un mot isolé vont se perdre, et ce proportionnellement au débit. Kingdon (Kingdon, 1958b : 18) a fait remarquer que les préfixés forment des exceptions à la règle 3 qui interdit deux accents forts successifs, mais il fait aussi remarquer que ces préfixés perdent l'accent « anti-rythmique » en discours continu. En effet, le débit rapide favorise les désaccentuations.

Un autre phénomène se produit également ; il s'agit du déplacement de l'accent, appelé *stress shift* en anglais. Ce phénomène affecte la hiérarchie accentuelle lexicale (accent primaire, accent secondaire). Nous constatons que dans certaines phrases nominales de type N+N ou A+N, l'affaiblissement de l'accent primaire sur l'élément qui fonctionne comme un adjectif est indiscutable. Dans des mots comportant deux accents, l'inversion de la hiérarchie est également constaté. SPE capturait déjà ce processus dans l'application cyclique de l'attribution de l'accent qui affaiblit l'accent de l'élément de gauche, en attribuant la valeur /1/ à l'élément de droite (*Nuclear Stress Rule*), même si le calcul aboutissait à des valeurs invraisemblables au bout de plusieurs applications du cycle.

Ce phénomène [illegible] généralement expliqué comme étant le fait de contraintes rythmiques, puisque [illegible]me naturel de l'anglais est l'alternance entre syllabes fortes / faibles et la te[illegible] à éviter des juxtapositions d'accents (*stress clash*). C'est ce que montre la ligne 1 de la grille en phonologie métrique. Ainsi, dans les

composés à double accentuation qui se trouvent en position de l'adjectif épithète, nous sommes devant un cas où l'accent du premier élément va se renforcer et l'accent /1/ du deuxième élément se perdra ou s'atténuera (devenant plutôt un accent /2/, voire /0/) en discours continu. Pour citer l'exemple de Pountney (1999), *an ˌarmˈchair* qui a le schéma /21/ (avec une variante en /10/) deviendra /10/ (ou peut-être /20/) dans *an ˈarmchair deˈtective*.

Selkirk (Goldsmith, 1995), qui étudie le phénomène de l'accent dans le discours, n'est pas convaincue que ce phénomène soit dû à des contraintes rythmiques (*the rhythm factor*). Il est vrai que dans l'exemple ci-dessus, *ˈarmchair deˈtective*, même sans déplacement d'accent, il n'y aurait pas d'accents adjacents mais plutôt une accumulation d'accents proches. Pour expliquer le déplacement accentuel dans ce type d'exemple, Selkirk suggère que le placement des accents est relié à la position des éléments dans l'unité tonale[24] (ce qu'elle appelle le *phrasing factor* qui implique ce qu'elle appelle *Phrase Edge Prominence*, voir 4.5.3.2), tout en reconnaissant le rôle que joue la hauteur mélodique (*the accent factor*) dans la reconnaissance de la proéminence (Goldsmith, 1995 : 562) :

> The relative prominence of a syllable within a sentence has been attributed to three different factors: (1) the presence or absence of a pitch accent (the *accent* factor), (2) the position of the syllable within a constituent structure (the *phrasing* factor), and (3) the presence or absence of other prominent syllables in the immediate vicinity of the syllable (the *rhythm* factor). A fuller rôle for phrasing based constraints on prominence than is conventionally assumed might remove motivation for the rhythmic component of a phonological theory of phrase stress.

---

[24] Le terme « phonological phrase » est employé en anglais. L'unité tonale peut être un mot, un syntagme ou une phrase complète. Voir aussi 1.1.1.5.

Sa conclusion va peut-être trop loin, mais nous devons tenir compte de ses propos qui prennent toute leur importance dans l'analyse acoustique du discours lorsque nous rencontrons des unités lexicales qui se forment comme des composés. Nous reviendrons à ces considérations dans l'analyse acoustique (4.5.3).

Si la question du déplacement accentuel n'est pas centrale dans notre analyse, nous ne voulons pas la passer sous silence, et elle a le mérite de souligner l'importance du rythme en anglais, ainsi que la hiérarchie accentuelle. Elle laisse également entrevoir quelques problèmes spécifiques en ce qui concerne la hiérarchie accentuelle dans les composés. C'est vers cette question que nous allons nous tourner maintenant pour ensuite définir notre nomenclature.

1.7 Nomenclature et notation adoptée

Dans la discussion qui a précédé, nous avons pu apprécier la diversité de la notation et de la nomenclature employées par les auteurs pour désigner les différents degrés d'accent, et le désaccord concernant le nombre de degrés à indiquer. Par ailleurs, nous nous sommes aperçue que le comportement des composés est souvent différent du lexique dans son ensemble, et pose des problèmes spécifiques. Nous allons résumer ce qui caractérise les composés du point de vue accentuel, puis nous pourrons déterminer la nomenclature et la notation que nous allons utiliser dans notre analyse.

1.7.1 Nomenclature

Les composés sont donc formés à partir de deux éléments (au moins), chacun pouvant porter ou non un accent. En fait, l'on peut trouver deux types accentuels. Le premier type de composés est constitué de ceux qui s'accentuent

avec une syllabe proéminente sur le premier élément, appelé *early stress* par Wells dans les notes du LPD 1990 et LPD 2000. Boisson utilise l'appellation « G » pour gauche puisque c'est l'élément qui se trouve à gauche qui porte l'accent /1/ (et la chute mélodique). Dans notre étude, nous parlerons d'**accentuation simple**. L'autre type de composés est constitué par ceux qui ont une proéminence sur les deux éléments avec une qui est plus ou moins égale, appelé *late stress* par Wells. Boisson, lui, donne l'appellation « D » pour droite puisque c'est l'élément de droite (le deuxième) qui porte la chute mélodique. Nous préférons le terme **accentuation double** qui montre bien que les deux éléments portent un accent, contrairement à l'accentuation simple où il n'y a qu'un accent. Dans ce cas, ces deux accents sont à différencier puisqu'ils ne fonctionnent pas de la même manière ; il faudra donc disposer d'une notation capable de faire cette distinction.

1.7.2 Notation : la hiérarchie accentuelle dans les composés

Avant d'étudier la question centrale de savoir comment noter ces distinctions dans notre analyse, nous voulons faire le point sur la hiérarchie accentuelle qui s'applique aux composés telle qu'elle est présentée dans les ouvrages de référence, à savoir les dictionnaires de prononciation EPD et LPD. Cette hiérarchie reprend la même notation que celle qui est étudiée en 1.4 « La hiérarchie accentuelle ». Toutefois, certains symboles revêtent une signification spécifique lorsqu'il s'agit d'un mot composé. Ensuite, nous proposerons un système qui tiendra compte de cette notation ainsi que de celles d'autres auteurs, afin d'aboutir à un système correspondant à nos objectifs.

1.7.2.1 Notation de la hiérarchie accentuelle dans les dictionnaires de prononciation

1.7.2.1.1 Tous les composés

Comme un composé est formé à partir de deux éléments, nous devons tenir compte de deux niveaux de notation accentuelle en même temps : celle qui concerne le poids relatif des syllabes à l'intérieur de chacun des éléments, puis celle qui concerne le poids des éléments les uns par rapport aux autres. C'est à ce deuxième niveau d'analyse que nous nous référons dans la section précédente, en définissant deux grands types de composés : les composés à accentuation simple, et les composés à accentuation double. La discussion de la notation de la hiérarchie accentuelle portera essentiellement là-dessus. Toutefois, le premier niveau d'analyse de chacun des éléments est pertinent là où un élément est plurisyllabique. Dans ce cas, l'identité phonologique de l'élément est forcement caractérisée par la présence d'au moins une syllabe forte et une syllabe faible ; la syllabe faible n'est pas notée dans les dictionnaires. Il s'agit de ce qui a été appelé l'accent /0/. Nous reviendrons à cette notation dans la section suivante.

Au niveau de la notation des composés en tant qu'unités, les dictionnaires de prononciation, dont le but est essentiellement prescriptif, ont adopté des manières de noter les accents qui varient selon les auteurs et selon les époques. Nous avons fait un tableau récapitulant ces différentes notations pour mieux pouvoir juger des avantages et inconvénients de chacune, et pour nous permettre de définir la nôtre.

Nous avons présenté les composés selon leur type : composés à accentuation simple et composés à accentuation double, en divisant chaque catégorie par

rapport au nombre de syllabes des éléments. Les numéros utilisés dans le tableau à des fins de clarification correspondent aux indications suivantes dans les dictionnaires :

accent /1/ = ˈ

accent /2/ = ˌ

accent /3/ = ₀

accent /0/ = pas d'indication accentuelle

Figure 1.8 : Dictionnaires de prononciation, notation des schémas accentuels dans les composés

Type d'accentuation & caractéristiques syllabiques des éléments de composition	EPD12 1963	EPD13 1967	EPD14 1977	EPD15 1997	LPD 1990	LPD 2000
Accentuation simple						
1er élément monosyllabique *goldsmith*	/1-0/	/1-0/	/1-0/	/1-0/	/1-0/	/1-0/
1er élément plurisyllabique *yellow hammer*	/10-20/	/10-20/	/10-20/	/10-20/	/10-30/	/10-20/
2ème élément monosyllabique *goldsmith*	/1-0/	/1-0/	/1-0/	/1-0/[25]	/1-0/	/1-0/
2ème élément plurisyllabique *yellow hammer*	/10-20/	/10-20/	/10-20/	/10-20/	/10-30/	/10-20/
Accentuation double						
1er élément monosyllabique *back door*	/1-1/	/1-1/	/2-1/	/2-1/	/2-1/	/2-1/
1er élément plurisyllabique *yellow fever*	/10-10/	/10-10/	/20-10/	/20-10/	/20-10/	/20-10/
2ème élément monosyllabique *back door*	/1-1/	/1-1/	/2-1/	/2-1/	/2-1/	/2-1/
2ème élément plurisyllabique *yellow fever*	/10-10/	/10-10/	/20-10/	/20-10/	/20-10/	/20-10/

En cas d'élément plurisyllabique comportant deux syllabes fortes, le premier accent a toujours la valeur /2/. Nous constatons que les différences de la notation

[25] Dans le cas de composés non compacts, EPD15 attribue un accent /2/ sur le deuxième élément désaccentué : *chicken pox* /10-2/ et *coal tar* /1-2/

des composés à accentuation simple sont minimes : seul LPD90 utilise l'accent /3/ pour les deuxièmes éléments plurisyllabiques. En revanche, la notation de l'accentuation double a évolué : avant EPD14, les deux éléments portaient un accent /1/ alors qu'à présent le premier élément porte un accent /2/. Nous allons commenter ce tableau par rapport aux deux types d'accentuation.

1.7.2.1.2 Composés à accentuation simple

En ce qui concerne les composés à accentuation simple, nous constatons que toutes les éditions s'accordent pour attribuer un accent /1/ au premier élément, ce qui est logique, puisque l'accentuation simple se définit par la présence d'un accent sur le premier élément par rapport au second, correspondant au noyau intonatif. Nous constatons que le deuxième élément monosyllabique n'est pas marqué pour l'accent : il est considéré comme étant « non accentué » et il porte donc par défaut un accent /0/. Cet accent correspond à la syllabe faible de tous les mots plurisyllabiques mais nous pouvons nous poser la question de savoir si cet accent est effectivement de même nature dans un composé et si cette convention très répandue est complètement satisfaisante. La convention qui est adoptée dans EPD15 est intéressante car un accent secondaire est attribué aux deuxièmes éléments pour les composés en deux mots distincts (figurant dans les sous-entrées, comme *ˈfish ˌknife*). Toutefois, il nous semble important d'adopter le même système de notation, quelle que soit l'orthographe du composé.

Si le second élément est plurisyllabique, nous nous retrouvons dans un autre cas de figure car cet élément comporte (au moins) deux syllabes, l'une ayant une proéminence par rapport à l'autre. La syllabe accentuée de l'élément en tant que

mot isolé porterait un accent /1/, mais quand il s'agit de l'élément « non accentué » d'un composé la question de savoir comment la noter se pose. EPD a toujours utilisé un accent /2/ ce qui se démarque déjà de la notation habituelle de l'accentuation où un accent /2/ ne suit pas un accent /1/ – il le précède. Dans l'introduction du EPD14 (Jones, 1977: xxii), Ramsaran écrit: « [...] post-tonic secondary stresses are shown only in the case of polysyllabic second elements of compounds, e.g. "season ticket" /ˈsiːzən ˌtɪkɪt/ » Roach & Hartman tiennent sensiblement les mêmes propos dans leur introduction de l'EPD 15 (Jones, 1997: xii):

> Secondary stresses have only limited occurrence after a primary stress: such a secondary stress is only marked in closed or hyphenated compound words where the second element is polysyllabic (e. g. ˈfishˌmonger).

LPD, lors de l'édition 1990, va plus loin encore et introduit l'accent /3/. Cette notation est abandonnée lors de l'édition 2000. Wells (Wells, 2000: x) explique ce changement dans l'introduction: « The separate symbol for "tertiary stress", which was not well received, has been replaced by the secondary stress mark ». La raison pour laquelle Wells renonce à l'accent /3/ n'est pas fondée sur des principes théoriques mais simplement parce que l'introduction de l'accent tertiaire a été contestée. L'objectif premier du dictionnaire est de servir ses utilisateurs, tels qu'ils sont perçus par l'éditeur: ainsi la standardisation de la notation présente certainement des avantages. Cependant, il est clair que, quelle que soit la notation de ces dictionnaires, un problème théorique est soulevé, que nous devrons aborder dans notre analyse.

1.7.2.1.3 Composés à accentuation double

Nous constatons que, pour l'accentuation double, l'élément qui reçoit l'accent correspondant à la chute mélodique est toujours doté d'un accent /1/, ce qui suit la même logique que celle des composés à accentuation simple. Cette notation correspond à l'appellation de Wells *late stress* par opposition à *early stress* et a le mérite de faire bien ressortir l'élément portant la chute mélodique. En ce qui concerne le premier élément – celui qui ne porte pas la chute mélodique – les éditions récentes de EPD et les deux éditions de LPD s'accordent pour lui attribuer un accent /2/. Il est intéressant de constater que cette notation est récente, EPD ayant changé de système entre la 13e et la 14e édition. Rappelons que EPD utilise encore deux accents /1/ pour des composés (ou mots) avec de multiples syllabes dans la 14e édition, mais dans la 15e, Roach & Hartman (Jones, 1997 : xii) expliquent dans l'introduction leur décision de ne plus utiliser ce système. « for the present edition only one primary stress may occur in a word or a compound ».

L'ancienne notation où un accent /1/ est indiqué sur les deux éléments d'un composé souligne le phénomène que nous désirons faire apparaître dans notre appellation **accentuation double**. Néanmoins, l'attribution de l'accent /1/ dans les dictionnaires actuels correspond précisément à la syllabe qui portera la chute mélodique. Ainsi cette logique est tout à fait satisfaisante. Par contre l'utilisation de l'accent /2/ devra être précisée en ce qui concerne les composés. Nous constatons que les dictionnaires attribuent un accent /2/ dans des cas où la fonction de l'accent n'est pas comparable. Il est utilisé pour l'accent rythmique du deuxième élément d'un composé à accentuation simple lorsque celui-ci est plurisyllabique (sinon, l'élément porte un accent /0/ puisque le rythme ne nécessite

pas d'indication accentuelle), par exemple *ˈbox ˌoffice*. Il est également utilisé pour le premier élément des composés à accentuation double aussi bien lorsque l'élément est monosyllabique que plurisyllabique, par exemple *ˌback ˈdoor* ou *ˌyellow ˈfever*. La nature de cet accent est très différente de l'accent rythmique porté par des dérivés, qui est placé en fonction du rythme. Sur ce mot *ˌderiˈvation*, l'accent secondaire est obligatoire afin de préserver le rythme de la langue. L'accent primaire est contraint sur la troisième syllabe, et comme l'une des deux premières doit porter un accent (règle 2 de Guierre, voir 1.5.3.1.2) et que nous devons éviter des suites de deux accents (règle 3 de Guierre, voir 1.5.3.1.3), l'accent secondaire est placé sur la première syllabe. Cette fonction de l'accent secondaire est très différente de sa fonction en ce qui concerne les composés : l'accent secondaire doit être indiqué sur le premier élément d'un composé à accentuation double en raison de la proéminence de l'élément (d'où l'appellation « accentuation double »), même s'il ne porte pas l'accent nucléaire ; l'accent fait partie de son identité phonologique et sémantique.

1.7.2.1.4 Notation du déplacement de l'accent sur les composés

Dans le cadre d'un dictionnaire de prononciation, il importe d'abord de présenter les entrées le plus simplement possible, donc d'adopter un système avec un minimum de niveaux d'accentuation permettant à l'apprenant de comprendre comment prononcer les mots donnés. Les dictionnaires actuels se limitent donc volontairement à un accent /1/ qui correspond toujours à la chute mélodique (du composé pris isolément) et un accent /2/ qui désigne l'accent qui se placera sur toute autre syllabe, pour des raisons différentes. Cependant, ces ouvrages, qui sont

des ouvrages de référence au niveau lexical, donnent des indications concernant le phénomène du *stress shift* (changement de la hiérarchie accentuelle dans le discours) dont nous avons parlé précédemment. LPD place le symbole « ◀ » après l'entrée; EPD15 indique «*stress shift*» en toutes lettres et en italiques. Bien entendu, l'objet des entrées du dictionnaire est l'accent lexical (donc celui des mots pris isolément), mais le phénomène du déplacement de l'accent est tellement répandu que les éditeurs donnent des informations portant sur ce point. Les composés à accentuation double sont particulièrement susceptibles de subir ce genre de changement. Peut-être que le fonctionnement de ce phénomène se comprendrait mieux si l'on utilisait une notation un peu différente pour les composés. Nous passons à présent à celle que nous proposons.

1.7.2.2 Notation de la hiérarchie accentuelle retenue

Ayant souligné les problèmes spécifiques des composés, et la manière dont les dictionnaires de prononciation notent leur accentuation, nous sommes en mesure de déterminer quel niveau de détail nous voulons utiliser, et quels termes employer. Nous préférons emprunter dans une large mesure les appellations habituelles, surtout celles des dictionnaires de prononciation, pour des raisons de lisibilité. Toutefois, là où cette nomenclature ne nous paraît pas rendre compte de la complexité du phénomène de l'accentuation dans les composés, nous apporterons des modifications qui permettront de fournir les précisions nécessaires.

1.7.2.2.1 Tous les composés

Rappelons que, pour déterminer les degrés d'accent, nous nous intéresserons essentiellement aux impressions des auditeurs, c'est-à-dire à la perception. Nous

avons déjà établi que le phénomène de l'accent est avant tout un phénomène de force relative des syllabes les unes par rapport aux autres. Il est donc évident que nous devons différencier accent et non-accent dans un premier temps. De même, dans des composés comme dans des mots non composés, l'on peut aussi être obligé d'intégrer deux accents sur le même mot pour les raisons rythmiques étudiées précédemment, puisque l'anglais interdit une séquence initiale de deux syllabes non accentuées. Il faut donc disposer d'au moins trois notations correspondant aux deux accents des syllabes fortes et au non-accent.

Ainsi, selon la majorité des auteurs, à l'intérieur d'un élément, l'on peut attribuer jusqu'à deux accents – /1/ et /2/ – que l'on différencie selon des critères de contraintes rythmiques, et à toutes les autres syllabes (qu'il y ait ou non réduction), on n'indique rien, c'est-à-dire que par défaut ce sont des syllabes avec un accent /0/. Nous retiendrons le terme **accent /0/**, mais par ailleurs nous pourrons être amenée à modifier la hiérarchie afin de rendre compte de la mise en relation des éléments les uns par rapport aux autres. Nous commençons par l'étude de la hiérarchie accentuelle dans les composés à accentuation simple.

1.7.2.2.2 Composés à accentuation simple

1.7.2.2.2.1 Deux éléments monosyllabiques

Prenons d'abord le cas le plus simple : un composé avec deux éléments monosyllabiques. S'il s'agit d'un composé à accentuation simple, nous proposons d'emprunter le terme d'**accent /1/**, en concordance avec les dictionnaires de prononciation pour le premier élément ; cet accent correspond au noyau intonatif.

Quant au deuxième élément, nous avons décidé de reprendre le terme d'**accent /0/**, qui correspond au non-accent ou à la désaccentuation[26], mais qui n'a pas de rapport avec la non-réduction de la voyelle[27]. En effet, si l'on considère que la non-réduction indique qu'il y a un accent (ne serait-ce qu'un accent /3/), quasiment tous les éléments de tous les composés auraient un accent /1/, /2/ ou /3/, l'accent /0/ étant restreint à certaines syllabes des éléments plurisyllabiques. De plus, cela se conforme aux usages des dictionnaires. Il n'en reste pas moins que la qualité vocalique de l'élément non accentuée dans les composés est une question épineuse que nous devrons aborder dans cette étude.

1.7.2.2.2.2 Un élément plurisyllabique

Dans d'autres composés, un (ou les deux) élément(s) peuvent être plurisyllabiques. Nous sommes donc obligée, quel que soit le statut de l'élément (accentué ou non) d'indiquer un accent fort sur l'une des syllabes au moins pour indiquer la proéminence relative par rapport aux autres syllabes de l'élément. Si l'élément plurisyllabique est le premier élément, il porte la chute mélodique, et donc les conventions habituelles peuvent fonctionner clairement. Nous utiliserons pour cet élément les termes d'accent /1/ et d'**accent /2/** (l'accent /2/ précédant l'accent /1/), le second élément monosyllabique ayant la valeur /0/.

Par contre, si le deuxième élément est plurisyllabique, nous souhaitons réintroduire, comme l'a fait Wells dans LPD 1990, l'**accent /3/**. L'accent /3/ se

---

[26] Voir la discussion en 4.2.2.2.1 « Désaccentuation d'un élément ».

[27] Nous nous démarquons donc du système de la théorie générative où la non-réduction entraîne un degré d'accent autre que /0/ qui est calculé en fonction de la structure syntaxique de la phrase phonologique dans son ensemble. Sa valeur est relative et non absolue.

situe en dessous des deux accents forts (/1/ et /2/) dans la hiérarchie accentuelle et indique que l'élément en question n'est accentué que par rapport aux autres syllabes de l'élément en question, non par rapport au premier élément. Cela permet de différencier cet accent par rapport à l'accent secondaire, celui-ci étant attribué en fonction des contraintes rythmiques à l'intérieur d'un élément. Si cet élément comporte lui-même plusieurs syllabes et si l'accent /3/ se situe sur la troisième syllabe (ou au-delà), nous aurons besoin d'indiquer l'accent secondaire, rythmique. Nous proposons le terme /R/ (pour accent rythmique). Ces cas seraient extrêmement rares au point que nous n'avons pas trouvé d'exemples lexicalisés ; mais nous avons voulu inclure cette éventualité sur le plan de la cohérence de l'ensemble.

1.7.2.2.3 Composés à accentuation double

1.7.2.2.3.1 Deux éléments monosyllabiques

Reprenons le cas le plus simple : un composé avec deux éléments monosyllabiques. Dans le cas d'un composé à accentuation double où il y a une proéminence sur les deux éléments quel que soit le nombre de syllabes, nous proposons d'emprunter le terme d'**accent /1/** pour le deuxième élément puisqu'il porte l'accent nucléaire et d'indiquer l'accent du premier élément avec la lettre **P** (pour primaire). /P/ indique un accent fort de même intensité que l'accent /1/ mais la syllabe ne portera pas la chute mélodique. Nous constatons que dans ce cas il n'y a pas d'accent /0/. Pour *back door*, nous aurons le schéma /P-1/ (voir la figure suivante).

1.7.2.2.3.2 Un élément plurisyllabique

Si le deuxième élément est plurisyllabique, nous pourrons utiliser les conventions habituelles, c'est-à-dire un **accent /2/** puisque la syllabe nucléaire de cet élément est notée /1/, comme pour les composés à accentuation simple. Par contre, si le premier élément est non seulement plurisyllabique, mais aussi pluriaccentuel, nous avons besoin d'indiquer un autre accent. Nous avons indiqué /P/ pour cet élément. S'il a besoin en plus d'un accent secondaire (rythmique), nous proposons à nouveau le terme /R/ qui a été proposé pour les composés à accentuation simple pour l'élément pluriaccentuel ne contenant pas la syllabe nucléaire.

1.7.2.2.4 Récapitulatif de la notation

L'accent /1/ désigne donc pour les deux types de composés la syllabe qui porte la chute mélodique (voir la nomenclature d'Huart (1984) et de Kingdon (1958b)), appelée aussi syllabe nucléaire. **L'accent /2/** correspond aux conventions habituelles et sert à noter l'accent rythmique de l'élément qui porte déjà l'accent /1/. **L'accent /P/** sert pour noter l'accent du premier élément d'un composé à accentuation double (correspondant à l'accent /1/ si l'élément était isolé). **L'accent /R/** sert pour noter l'accent secondaire de l'élément du composé qui ne contient pas la syllabe nucléaire (le deuxième élément d'un composé à accentuation simple, le premier élément d'un composé à accentuation double). **L'accent /0/** sert à noter le non-accent. Dans le tableau ci-après, nous reprenons ces notations avec des exemples précis de composés

Figure 1.9 : Récapitulatif de la nomenclature proposée

Type d'accentuation & caractéristiques syllabiques des éléments de composition	Exemple avec signes pour désigner le schéma accentuel	Schéma accentuel Proposition de notation des composés	Schéma accentuel LPD 1990	Schema accentuel LPD 2000 et EPD15
Accentuation simple				
1er élément monosyllabique	*ˈgoldsmith*	/1 - 0/	/1 – 0/	/1 – 0/
1er élément plurisyllabique	*ˈyellowˌhammer*	/10 –30/	/10 – 30/	/10 – 20/
2ème élément monosyllabique	*ˈsecond hand*	/10 – 0/	/10 – 0/	/10 – 0/
2ème élément plurisyllabique	*ˈgodˌdaughter*	/1 –30/	/1 – 30/	/1 – 20/
1er élément pluriaccentuel	*ˌcigarˈetteˌholder*	/201 – 30/	/201 – 30/	/201 – 20/
2ème élément pluriaccentuel	pas d'exemple lexicalisé	/*1* – R030/		
Accentuation double				
1er élément monosyllabique	*ˌback ˈdoor*	/P – 1/	/2 – 1/	/2 – 1/
1er élément plurisyllabique	*ˌyellow ˌfever*	/P0 – 10/	/20 – 10/	/20 – 10/
2ème élément monosyllabique	*ˌsecond-ˈhand*	/P0 – 1/	/20 – 1/	/20 – 1/
2ème élément plurisyllabique	*ˌgold ˈmedal*	/P – 10/	/2 – 10/	/2 – 10/
1er élément pluriaccentuel	*ˌintroˈductory ˈoffer*	/R0P0 – 10/	/2010 – 10/	/2010 – 10/
2ème élément pluriaccentuel	*ˈcross-exˌaminˈation*	/P – 02010/	/2 – 02010/	/2 – 02010/

Dans la deuxième colonne, nous avons cité un exemple lexicalisé en indiquant le schéma accentuel sur la graphie. Toutefois, nous nous sommes heurtée au problème soulevé par Ladefoged (1993) concernant le nombre de symboles de l'IPA qui existe pour désigner les degrés d'accent. Les indications correspondent donc à la notation de LPD 1990. Pour indiquer les distinctions affinées que nous proposons, nous nous servons du codage de la colonne 3 où chaque lettre ou chiffre correspond à la valeur accentuelle de la syllabe.

Cette nomenclature introduit des appellations qui veulent rendre compte avec précision de la hiérarchie accentuelle des composés. Pourquoi ces changements par rapport aux usages ? Pour les composés à accentuation simple, la

seule modification que nous ayons voulu apporter concerne le statut accentuel du deuxième élément : accent /3/ au lieu d'accent /2/. Cette question sera reprise dans l'étude des éléments moins ou non accentués et de la valeur de la voyelle dans ces syllabes.

Nous avons vu que dans les composés à accentuation double, chaque élément reçoit un accent mais ces deux accents ne sont pas identiques. L'on ne peut pas non plus les considérer tout à fait comme les accents primaires et secondaires des autres mots du lexique qui, eux, suivent des contraintes rythmiques. Dans un composé, il ne s'agit pas vraiment de contrainte, et le schéma accentuel indique un rapport entre les deux éléments qui a des implications au niveau de la construction du sens (comme nous l'avons vu en 1.2 avec l'exemple *ˌbrick ˈfactory* qui désigne une « usine faite en brique »). En quoi ces deux accents sont-ils différents ? Pour répondre, nous avons tenu compte de l'intonation et nous avons utilisé le terme accent /1/ pour le deuxième terme, en conformité avec l'usage des dictionnaires et parce que la chute mélodique a lieu sur la syllabe avec un accent /1/ dans un mot non composé tout comme dans un composé. L'introduction de /P/ à la place de l'accent /2/ s'explique car cet accent n'est pas comparable avec un accent secondaire sur un mot non composé. L'accent est attribué en fonction des indications qu'il fournit sur les relations entre les deux termes, et non seulement par rapport aux contraintes rythmiques. La nomenclature doit rendre compte de ces faits.

Nous constatons que les nouvelles appellations concernent les composés à accentuation double (/P1/ au lieu de /21/ sur *ˌback ˈdoor*, par exemple, et

l'introduction, en plus, de /R/ dans *ˌintroˌductory ˈoffer*, /R0P00 10/) alors que les termes habituels couvrent (ou presque) les composés à accentuation simple. Ce constat montre que ce sont ces cas qui sont problématiques. Pountney (1999) les appelle «*false compounds*» (Pountney, 1999: 46) parce qu'ils fonctionnent comme des composés du point de vue lexical, mais pas du point de vue accentuel. Cependant, notons que le même terme est employé par Lilly & Viel (1993/1999: 39) pour désigner les composés à accentuation simple dont le deuxième élément subit une réduction vocalique. Dans les deux cas, le comportement du composé est considéré comme étant exceptionnel. Nous verrons dans notre étude quel statut on peut leur attribuer.

CHAPITRE 2

Chapitre 2 : La constitution du corpus

Introduction

Dans le premier chapitre nous avons étudié le phénomène de l'accentuation et nous avons tenté de comprendre ce que l'accentuation des composés a de particulier. Nous avons débouché sur une proposition de nomenclature qui reflète cette spécificité. Nous avons également parlé des différentes théories portant sur ce sujet, et de ce que chacune peut apporter à l'analyse. Pour notre part, nous avons décidé de nous situer dans une perspective de linguistique de corpus et c'est pourquoi, dans le présent chapitre, nous allons expliquer comment nous avons procédé afin de constituer le corpus d'analyse.

Nous allons mettre en œuvre des traitements automatiques qui nous permettront de procéder à une analyse approfondie. Pour effectuer de tels traitements, il faut d'abord constituer un corpus d'analyse adapté ; nous constituons donc un corpus sur support informatique qui est ainsi interrogeable par des outils de traitement automatique. L'avantage de cette méthode consiste en sa capacité à opérer sur un grand nombre de données, de manière rapide, et surtout fiable, même si les résultats obtenus sont à analyser d'un œil critique. Dans ce chapitre, nous expliquerons comment nous avons choisi nos sources et les méthodes et les critères que nous avons utilisés afin de constituer ce corpus. Dans le chapitre suivant, nous expliquerons quels sont les outils de traitement automatique utilisés pour analyser les données.

Nous devons disposer d'un corpus important, couvrant un large éventail de domaines, même s'il n'est pas totalement exhaustif. Pour qu'il soit exploitable, il

doit présenter les informations qui nous sont indispensables pour notre travail. Nous allons donc déterminer quelles sont ces informations. Notre analyse porte sur les schémas accentuels des mots composés en anglais britannique. Pour être à même de mener à bien ce travail, nous devons donc prévoir des informations portant sur un certain nombre d'éléments. Le corpus doit contenir des informations sur la graphie d'abord, puisqu'il s'agit de la donnée de base de toute recherche lexicographique. Ensuite, des indications syntaxiques permettront d'effectuer des recherches sur des catégories différentes et d'obtenir des résultats que l'on peut évaluer selon le type syntaxique. Puisque notre étude porte aussi sur la prononciation des mots, la phonie nous fournira des renseignements importants, mais surtout, bien entendu, il nous faut impérativement une indication du schéma accentuel.

Nous avons choisi deux types de sources différents : des sources écrites et des sources orales. Nous expliquerons en détail les raisons de nos choix pour chaque type de source dans les parties 2 et 3 de ce chapitre qui traitent de la constitution même du corpus d'analyse. En introduction, nous voulons simplement préciser que les sources écrites fournissent des indications provenant de dictionnaires lexico-phonétiques, donc sur l'accent lexical. En revanche, les sources orales nous renseignent sur la prononciation effective, en discours, de ces mêmes entrées ; elles fournissent également un complément d'information concernant des entrées non lexicalisées (des néologismes, par exemple). Ces sources sont ainsi complémentaires.

2.1 Le contenu du corpus

Le mot «corpus» implique des choix: d'abord des choix concernant les sources que nous utilisons pour le constituer; nous venons d'en parler brièvement. Mais ce mot implique aussi des choix en ce qui concerne les données que nous sélectionnons dans les sources et qui seront soumises à l'analyse. Nous devons donc déterminer quelles sont les données à analyser, ce qui nous amène à aborder la question fondamentale «qu'est-ce qu'un composé»?

2.1.1 Qu'est-ce qu'un composé?

Un des intérêts majeurs des composés consiste en leur grande productivité. Ce phénomène est souligné par Picoche (1992: 121-2):

> La langue est créatrice [...] elle est capable [...] à partir d'un nombre fini de morphèmes, d'engendrer des mots nouveaux en nombre théoriquement illimité, au moyen des mécanismes de la dérivation et de la composition.

Picoche parle ici du français, mais c'est tout aussi vrai de la langue anglaise, comme le souligne Tournier (1985: 137):

> La construction constitue le macromécanisme le plus productif de l'anglais contemporain et, parmi les processus relevant de la construction, c'est la composition par juxtaposition qui vient en tête, suivie de la suffixation, puis de la préfixation.

Il est vrai que la composition est fréquemment utilisée pour créer des néologismes désignant des concepts nouveaux. Ainsi, quand l'Homme a réussi à voyager dans l'espace, on a créé *spaceship*. Cette juxtaposition de deux noms, construction très répandue d'après Tournier (voir ci-dessus), fait bien ressortir l'idée nouvelle à partir d'idées connues, par le moyen de leur association.

Cependant, nous avons parfois le choix entre plusieurs formes. Là il ne s'agit plus de néologismes mais plutôt de concepts connus. Les agrégats nominaux

peuvent être des noms juxtaposés (N+N), des noms avec un génitif associé au premier (N's+N), ou encore des noms reliés par une préposition (N+prep.+N). Le choix d'une forme de composition plutôt qu'une autre n'est certes pas fortuit. Comme le dit Huart (1989: 47): «Il est extrêment rare qu'à une différence de forme ne corresponde pas une différence de valeur». Son analyse approfondie des différents agrégats nominaux, qu'elle compare entre eux, repose la question de la définition même d'un composé.

2.1.1.1 Les critères de définition d'un composé

2.1.1.1.1 Le critère sémantique

Il y a des critères multiples, ce qui tend à suggérer que la question est délicate, mais il y a un début de consensus autour du critère sémantique. Kingdon (1958b: viii) propose la définition suivante (souligné par nous):

> English-type compound. A semantic unit that has attained dictionary status, composed of two or more words written separately, joined by a hyphen, or as one word, and whose meaning may or may not be a straightforward combination of the meaning of its components.

Nous constatons que les indications graphiques n'apportent que peu d'aide puisque les trois manières d'écrire sont associées aux composés. Même au niveau sémantique, comme le montrent les mots soulignés, Kingdon nous laisse toute latitude. En effet, la définition reste très floue: elle pourrait convenir à des syntagmes autant qu'à des composés. Mossé (1945:131) est un peu plus précis (souligné par nous):

> On dit qu'il y a composition lorsque deux termes qui ont chacun leur sens propre sont réunis pour former un mot nouveau. Le sens du mot composé est souvent différent de la somme des sens des composants.

Cette fois-ci, nous constatons que la définition en termes sémantiques tend vers la nouveauté du sens du composé. Ce qui est à remarquer dans ces deux citations est l'utilisation du terme « mot » qui n'a jamais pu être défini de façon complètement satisfaisante. Le terme est d'ailleurs utilisé différemment par chaque auteur. Kingdon considère qu'un mot est graphiquement séparé, alors que Mossé considère qu'un composé (avec des variantes graphiques déjà citées par Kingdon) est un mot. Fudge (1984 : 134) arrive à cerner toute l'ambivalence des composés : « Are compounds one word or two ? The answer is that they are *both* one word *and* two ».

Le critère sémantique est à nouveau proposé par Deschamps qui, lui, évite de parler de « mots » (Deschamps et al., 2000 : 73), et sa définition rejoint celle de Mossé en grande partie : « compounds can be defined as semantically inseparable units ». Comme le dit Picoche (Picoche, 1992 : 17)

> Le meilleur critère de l'unité du mot composé est donc sémantique et résulte du fait qu'il est interprété comme le signifiant d'un concept unitaire et non de l'association d'un déterminé et d'un déterminant.

Cette définition est beaucoup plus précise que celle proposée par Kingdon, et elle reflète certainement la réalité psychologique du concept de composé tel qu'il est ressenti par un locuteur natif. Cependant, l'unicité du sens est difficilement mesurable et, bien que la définition soit plausible, elle ne suffit pas toujours comme critère de sélection. Jespersen (Jespersen, 1961-5: 134) a conscience de ces limites (voir l'adverbe que nous soulignons) : « A compound may perhaps be <u>provisionally</u> defined as a combination of two or more words so as to function as one word, as a unit ». Huart apporte une lumière nouvelle en

situant le composé par rapport à la théorie des opérations énonciatives. Pour elle, le composé est le reflet d'une relation primitive, il y a inclusion dans une classe (Huart, 1989: 45), ce qui permet d'écarter des agrégats irrecevables tels que *secretary school et *tourism arrival (pas de relation primitive), et *jazz origins ou encore *wall hole (qui ne nomment pas de classe)[1] :

> Les agrégats nomment des classes ; on ne s'en sert pas pour référer à une occasion particulière, ni pour opérer des repérages ne pouvant être que spécifiques [...] Il faut que les deux termes rentrent dans une relation primitive au niveau des notions.

Néanmoins, le critère du sens des composés n'est pas le seul qui soit cité. Nous avons voulu l'étudier plus longuement ici, car il s'agit d'un critère qui ne peut en aucun cas être utilisé pour extraire des données de façon automatique. En revanche, les autres critères qui sont souvent mentionnés peuvent, eux, servir à ce genre de traitement automatique. Il s'agit du critère syntaxique, du critere graphique, du critère phonétique, et enfin du critère accentuel. Nous allons étudier chacun de ces critères, puis, dans la deuxième partie, nous expliquerons comment nous les avons utilisés afin de constituer le corpus à partir des sources écrites.

2.1.1.1.2 Le critère syntaxique

Ce critère apporte un complément d'information par rapport au critère sémantique. Il s'appuie sur des tests de manipulation de suites de mots que l'on peut ensuite considérer ou non comme des composés. Un des tests possibles est celui de la non séparabilité d'un composé. S'il est possible d'intercaler un mot comme *and* entre les deux composants, on peut considérer qu'il ne s'agit pas d'un

---

[1] Le symbole * est utilisé ici pour désigner un agrégat inacceptable, agrammatical.

composé. De même, si le mot *very* peut qualifier l'adjectif (dans le cas d'une formation adjectif + nom, comme *red tape*), alors il ne s'agit pas d'un composé ; si la mise à la forme comparative de l'adjectif est possible, la séquence ne forme pas non plus un composé. D'autres manipulations comme la mise au pluriel pour les noms peuvent aussi servir de tests. Pour que l'on considère l'unité comme un ensemble, et donc comme un composé, il faudrait pouvoir la manipuler comme le mot individuel. Ces tests pourraient faire considérer comme des composés des expressions ou séquences qui n'en sont pas. Marchand nous met en garde (Marchand, 1969 : 125), en proposant des exemples tels que *bring to light* ou encore *give the sack*, la mise au pluriel du nom dans les deux cas n'est pas envisageable :

> [...] many motivated phrases are entirely degrammaticalized (lexicalized) i.e. any modification can only apply to the whole combination while the constituents are no longer susceptible of characterization.

Adams passe en revue tous ces tests (Adams, 1973 : 58) mais elle aussi attire notre attention sur leurs limites, et elle conclut :

> There is then no easy answer, covering all cases, to the question of 'what is a compound'. In this chapter I have been fairly liberal in including collocations which I felt were to some degree established, and therefore candidates for inclusion in this category. Some of my examples may not be compounds for all readers, but since any attributive-head noun group is potentially a compound and will come to be felt as one if it happens to be used often enough, it is as well not to be too arbitrary in deciding one way or the other. Compound-status can be seen as a matter of degree...

La fréquence d'emploi, cause du phénomène, est elle-même hautement affaire de degré. On peut donc parler de « degré de figement », comme Gross (1988 : 70)[2].

---

[2] Gross base son analyse sur la syntaxe et rejette les hypothèses favorisant une approche sémantique.

Les réserves qu'émet Adams concernent plus particulièrement des unités non lexicalisées que nous trouverons dans nos sources orales. Nous retenons ses remarques, et lors de la constitution du corpus d'analyse, nous suivrons sa décision d'inclure certaines unités qui n'ont pas un statut de composé lexicalisé. Retenons aussi sa remarque sur la fréquence d'usage, car effectivement c'est un phénomène puissant qui peut aboutir à la lexicalisation, parfois très rapidement. Comme le fait remarquer Tournier (Tournier, 1988 : 66), le critère fréquentiel doit être pris en compte. Il donne pour exemple la séquence *gold-rush* : « Ceci montre qu'une séquence fortuite peut évoluer vers la lexicalisation à mesure que s'accroît sa fréquence d'emploi ». Cependant, il est difficile d'appliquer ce critère sans avoir recours à un corpus spécifique[3], car il repose sur des impressions psychologiques et difficilement mesurables. Nous passons à présent à un autre critère qui, lui, est plus facile à attester : le critère graphique.

2.1.1.1.3 Le critère graphique

Nous avons cité Kingdon au début de cette section et sa définition d'un composé « de type anglais » faisait référence à la manière de l'écrire. Ces composés peuvent s'écrire de trois manières : en un seul mot (appelé composé compact[4] dorénavant), avec un trait d'union (composé à trait d'union) et en deux

---

[3] Il serait possible de mesurer des différenciels de fréquence à partir d'un corpus échantillonné de manière stable, et étudié diachroniquement (LOB et FLOB, par exemple), mais cela dépasse notre cadre de travail.

[4] Nous n'employons pas le terme « compact » au sens culiolien dans le cours de cette thèse. Deschamps et al. (2000 : 73) emploient le terme « closed compound ».

mots séparés par un espace (composé non compact). De plus, le même composé peut être orthographié de façon différente. Parfois, ceci correspond au degré de familiarité du terme, ce qui rejoint le critère de la fréquence cité plus haut. Fowler (Fowler, 1996: 370) fait remarquer qu'un mot composé spécialisé sera écrit en composé compact par un spécialiste du domaine alors qu'un non-spécialiste aura tendance à montrer dans le choix de l'orthographe qu'il est composé de deux éléments: « Specialized journals often print *seaurchin* as one word; in more general work it is printed as *sea urchin* or *sea-urchin*. The familiarity of a term is important ». Bauer va plus loin et démontre l'inconsistance de la graphie, en prenant l'exemple du terme *girlfriend*. Ce composé peut s'écrire des trois manières, et dans son enquête, un seul dictionnaire distingue le sens du composé compact de celui du composé non compact. Bauer rejette également l'hypothèse de l'évolution de la graphie qui tendrait à l'écriture en compact, au fur et à mesure que le temps passe[5]. Des néologismes comme *airside* peuvent parfaitement s'écrire en compact alors que le terme *college degree*, pourtant familier, s'écrit en non compact. Il conclut (Bauer, 1998: 69): « It seems that, to the extent that one spelling becomes relatively standardized, it is by convention rather than because of any linguistic principles at work ». De même, Tournier (1988 : 64) commente le flottement noté dans l'orthographe. D'après lui, il « contribue à rendre le critère

[5] Nous faisons une étude de l'évolution de la graphie en 4.3.3.2.1.

graphique relativement peu fiable». Adams (1973 : 59) aussi considère que ce critère n'est pas fiable[6] :

> Spelling is an unreliable criterion of compound status. Compounds may be written as one word, as two hyphenated words, or as two separate words, sometimes regardless of the degree of unity that may be felt between the elements.

De même, Paillard (2000 : 46) rappelle qu'un composé peut être signalé par l'écriture en compact (ce qu'il appelle «la fusion graphique») qui est une condition suffisante mais non nécessaire. Lorsque nous utiliserons ce critère pour extraire des données de nos sources écrites, et lorsque le corpus d'analyse sera constitué, nous serons en mesure de l'évaluer plus précisément (voir 4.3.2.2).

2.1.1.1.4 Le critère phonétique et phonologique

Une des règles fondamentales de Guierre tient compte de la corrélation entre le non-accent et la réduction vocalique, mais, nous l'avons vu au chapitre 1, les composés sont une exception majeure aux règles générales, puisqu'ils présentent une très forte tendance à ne pas subir de réduction vocalique sur la syllabe qui perd son accent du fait de la composition. Guierre (Guierre, 1979 : 350) formule cette règle de manière claire : «[...] l'absence de voyelle réduite est une condition nécessaire, mais non suffisante, pour qu'un mot [dissyllabique] soit composé»[7]. Nous sommes ainsi en possession d'un critère très utile qui nous permettra d'isoler toute une catégorie de composés, mais le terme «non suffisante» nous met en garde contre une trop grande généralisation.

---

[6] Voir aussi Lyons (1977 : 535) et Bauer & Renouf (2001 : 112) qui insistent sur le manque de cohérence du système graphique.

[7] Cette condition nous rappelle les propos de Paillard (2000 : 46) sur la graphie.

La non-réduction vocalique n'est pas la seule caractéristique phonologique que nous pouvons associer aux composés. Nous trouvons dans Troubetzkoy (Troubetzkoy, 1948) une étude des facteurs qui peuvent servir de limite de mot ou de morphème (ou les deux, un morphème pouvant être un mot). Nous pouvons de la même façon considérer les facteurs indiquant la frontière entre les deux composants d'un composé. D'après Troubetzkoy, ces facteurs peuvent être phonématiques (donc en relation avec des phonèmes) ou bien aphonématiques (par exemple, l'accent). L'association de certains phonèmes peut indiquer un placement particulier par rapport au mot. Ses exemples sont tirés de beaucoup de langues ; pour l'anglais, la qualité claire ou sombre du /l/ est un indicateur de la position dans le mot (Troubetzkoy, 1948 : 300). Nous pouvons étendre ce constat, et l'utiliser pour déterminer la présence d'un composé compact dans certains cas, en tenant compte des contraintes phonotactiques. Certaines séquences de phonèmes indiquent une limite de mot ; lorsque nous trouvons une de ces séquences à l'intérieur d'un mot, là où elle est illicite normalement, nous sommes en présence d'un composé : *mudguard*, *topcoat*, où l'on trouve les séquences /dg/ et /pk/ qui ne se trouvent qu'à la frontière d'un mot. De même, l'assimilation à l'intérieur d'un mot dépendra de son statut de composé ou de non-composé. Dans EPD14, l'exemple de la prononciation de *in* dans *income* et *incoming* est mentionné. La séquence de phonèmes /nk/ sera prononcée [nk] si le locuteur considère qu'il y a bien deux éléments (donc un composé, comme pour *incoming*), mais il y aura assimilation si le locuteur n'a pas conscience de deux éléments distincts et la séquence sera prononcée [ŋk] (comme dans *income*) : « Often assimilation within a word will depend on the extent to which the word is felt as a

compound» (Jones, 1977 : xxi). Il est intéressant de noter la formulation qui fait référence aux impressions intuitives du locuteur. Nous aborderons les problèmes que peuvent poser ces impressions dans la section suivante.

Quant à l'accent, il ne marque pas de syllabe particulière en anglais *a priori*, contrairement à d'autres langues où l'accent est fixe et indique un rang syllabique de façon très nette. Pourtant, le critère accentuel est souvent cité comme un indicateur assez fiable de la composition.

2.1.1.1.5 Le critère accentuel

D'après Adams, ce critère peut permettre de distinguer entre un syntagme nominal et un composé. Elle dit (Adams, 1973, 59) : « Accent is sometimes helpful in deciding whether a noun phrase is a compound. Many compounds are distinguished from noun phrases by having the nucleus on the first element ». Nous avons déjà fait référence à cette tendance à l'accentuation à gauche (accent simple) dans le premier chapitre. Elle explique pourquoi dans la grammaire générative il existe une règle appelée la *Compound Stress Rule* qui permet d'attribuer un accent sur le premier élément, et qui est à opposer à l'autre règle qui détermine le placement de l'accent, la *Nuclear Stress Rule*, qui, elle, concerne l'accent nucléaire dans la phrase qui tend à être à la fin. La grammaire générative part du principe que les seuls vrais composés obéissent à cette règle, mais nous voulons mettre ce point de vue à l'épreuve.

Tournier propose le même critère, mais sans adopter le point de vue des génératives, il reste convaincu que le schéma accentuel normal d'un composé

est à accentuation simple. Il compare le syntagme nominal et le composé (Tournier, 1988 : 66) :

> L'ensemble de la séquence porte deux accents principaux. Il n'en va pas de même pour un composé, qui se comporte généralement du point de vue accentuel comme un mot simple, en ce sens que, dans la grande majorité des cas, il n'a qu'un seul accent principal (placé, le plus souvent, sur le premier élément).

Il conclut (Tournier, 1988, 66) : « Il reste, bien entendu, que ce critère, plus fiable que les précédents, ne sert pas dans la langue écrite, et qu'il ne fonctionne pas dans certains cas (à vrai dire peu nombreux) ». C'est ce que nous verrons dans nos analyses.

2.1.1.2 Les critères de définition d'un composé : Quel critère appliquer ?

Nous avons passé en revue les différents critères pouvant déterminer s'il s'agit ou non d'un composé. Il nous reste à déterminer le ou lesquels nous allons utiliser pour constituer notre corpus. Aucun d'eux ne paraît tout à fait satisfaisant. Nous avons constaté les limites du critère graphique qui manque de consistance, ainsi que les tests syntaxiques qui peuvent donner les mêmes résultats à des non-composés. Les séquences phonologiques donnent des indications, mais il est clair que les séquences qui marquent la frontière entre les différents composants ne sont pas toutes illicites en milieu de mot. On peut très bien trouver une séquence voyelle + consonne dans n'importe quelle position ; ce n'est pas réservé à des composés tels que *jaywalk*, par exemple. Le critère accentuel est loin de faire l'unanimité. Si Marchand (1969 : 27) tient à ce critère pour écarter les syntagmes lexicaux, tels *iron curtain*, (qu'il appelle « lexical phrases »), le point de vue de Bauer change radicalement entre 1978/1983 et 1998. En 1978 et 1983, d'après lui, seul l'accent pouvait servir de critère pour distinguer entre le composé et le

syntagme : « I considered only stress as a possible distinguishing criterion between the two classes » (Bauer, 1998 : 71). Dans son article de 1998, il ne considère plus l'accent de cette manière. Jespersen (1961 : 135) critique aussi le critère accentuel en ces termes :

> If, therefore we stuck to the criterion of stress, we should have to refuse the name of compound to a large group of two-linked phrases that are generally called so, such as *headmaster* or *stone wall.*

Même si *stone wall* est peut-être inclus à tort dans cette catégorie, le statut de composé de *headmaster* est incontestable, et il conclut (Jespersen, 1961 : 137) que nous devons nous reposer sur le critère sémantique :

> As formal criteria thus fail us in English, we must fall back on semantics, and we may perhaps say that we have a compound if the meaning of the whole cannot be logically deduced from the meaning of the elements separately, see e.g. *bedroom, -clothes, -post, -time.*

Cette définition nous paraît trop restrictive. En appliquant ce critère à lui seul, nous excluons beaucoup d'unités qui peuvent être considérées comme des composés, par exemple *steam train*, dont le sens global est parfaitement compréhensible à partir des éléments.

Pour conclure cette discussion sur les différents critères de définition d'un composé, nous nous tournons vers Tournier (1988, 67) :

> Comme on a pu l'observer, aucun des critères proposés pour décider si une séquence est un mot composé n'est à lui seul déterminant dans tous les cas. Toutefois l'application conjointe des cinq critères peut, sinon aboutir à une certitude, du moins permettre de réunir un faisceau de présomptions. L'attitude la plus prudente pour un non-anglophone est de se fier à la décision de lexicographes anglophones de répertorier ou non telle séquence comme un composé, avec cette réserve que, tout dictionnaire étant par définition en retard sur l'évolution du lexique, il conviendra de toujours choisir un dictionnaire récent et de se rappeler qu'une séquence fréquente non-répertoriée a toutes chances de se trouver sur la voie de l'intégration.

Nous allons nous inspirer de ces paroles en ce qui concerne les sources écrites, mais les sources orales pourront justement nous fournir d'autres renseignements au niveau de ces séquences non répertoriées.

Nous allons reprendre tous ces critères en les utilisant afin de constituer le corpus d'analyse à partir des sources écrites, ce qui permettra de les évaluer, mais avant de faire cette démarche, nous voulons aborder deux autres questions concernant le contenu du corpus d'analyse. Premièrement, nous devons poursuivre la discussion sur ce qu est un composé en étudiant les cas limites que nous avons trouvés dans nos sources écrites, puis déterminer sur quels critères nous incluons ou nous excluons ces cas. Deuxièmement, nous devons préciser l'étendue du corpus d'analyse et mettre certaines limites en ce qui concerne le type d'entrée à inclure.

2.1.2 Qu'est-ce qu'un composé ? Les cas limites

2.1.2.1 La dérivation et la composition

La dérivation a déjà été citée en tant que processus de formation lexicale, tout comme la composition. La distinction entre ces deux types de formation n'est pas aussi facile à faire qu'il n'y paraît au premier abord, surtout en ce qui concerne la préfixation, ce qui explique pourquoi elle a souvent été traitée avec la composition (voir les remarques de Paillard, 2000 : 47). Si nous reprenons les critères de définition d'un composé (voir 2.1.1.1), les mots dérivés sont construits à partir de morphèmes liés alors que la composition fait intervenir des éléments

libres[8]. Nous citons les propos d'Adams (1973 : 30) qui compare les deux types de construction :

> A 'compound word' is usually understood to be the result of the (fixed) combination of two free forms, or words that have an otherwise independent existence, as in *frostbite, tape-measure, grass-green.* These items, though clearly composed of two elements, have the identifying characteristics of single words: their constituents may not be separated by other forms, and their order is fixed. In a 'derived word', at least one element, the affix, is a bound form, with no independent existence and, characteristically, the more general meaning that one would expect a 'grammatical' element to have.

Adams fait référence à deux critères pour différencier ces processus : le premier concerne l'indépendance (ou la dépendance) des morphèmes ; le second fait intervenir la compositionnalité sémantique[9].

Le concept de morphème lié paraît clair. Il s'agit d'un morphème qui est toujours associé à un radical (ou à une forme de base), qui n'apparaît jamais isolément ; il est donc dépendant. Dans le cas de l'affixation, ces morphèmes liés ont plusieurs fonctions. Ils servent à créer des catégories syntaxiques (tel affixe génère des noms, par exemple *-hood*, tel autre génère des adverbes, par exemple *-ly*, tel autre des verbes, par exemple *en-*) ; ils sont du reste productifs. Pourtant, il faut dire que le statut de dépendance de certains affixes n'est pas toujours clair, et certains d'entre eux ont une sorte de double statut (comme *fore* qui est à la fois morphème lié et morphème libre, voir 2.1.2.1.2).

---

[8] Bauer emploie le terme « morph » : il en existe deux sortes : « bound morphs » et « potentially free morphs ». Les composés se forment à partir de la deuxième. Voir Bauer, 1988 : 11.

[9] Voir aussi les propos de Corbin (1992 : 3) qui parle du « pouvoir référentiel » des constituants d'un mot composé alors qu'un affixe n'a pas ce pouvoir.

Certains affixes peuvent avoir une fonction sémantique et leur adjonction indique qu'il faut interpréter le mot d'une certaine façon : comme la négation de la forme de base, pour prendre l'exemple des préfixes et les suffixes négatifs (*un-*, *in-*, *im-*, *il-*, ... *-less*) ; c'est ce qu'Adams appelle le sens général. Cependant, il nous arrive d'hésiter sur le statut de certains mots : sont-ils des dérivés ou bien des composés ? Ces hésitations sont plus fréquentes dans les mots préfixés que dans les mots suffixés en raison de la forte teneur sémantique de certains préfixes, malgré la définition d'Adams qui tente de différencier du point de vue de la sémantique les affixes des dérivés (dont le sens est général) et les composés (dont le sens serait plus précis).

Plusieurs autres facteurs interviennent lorsque nous devons décider du statut de mot composé ou de mot dérivé. D'abord, nous hésitons parfois parce que notre appréciation est influencée par la graphie. Un mot à trait d'union est susceptible d'être interprété comme un composé parce que l'indépendance des éléments est suggérée par cette orthographe. Nous pouvons aussi être influencée par le comportement accentuel du mot. Il faut préciser que cette dernière remarque ne concerne que les préfixes. Les mots qui portent un accent primaire sur l'élément que nous pouvons considérer en tant que préfixe peuvent être interprétés de deux manières : ou bien le préfixe n'est plus ressenti en tant que tel (comme dans *recap* /10/, forme tronquée de *recapitulate* /20100/), et il s'agit donc d'un mot sans structure interne, ou bien le préfixe fonctionne comme un élément indépendant qui rapproche le mot des vrais composés à accentuation simple (*supermarket* /1030/). Par ailleurs, les préfixes portent très souvent un accent secondaire,

l'accent primaire étant sur la forme de base (qui peut aussi être considérée comme étant le deuxième élément d'un composé). Ce comportement accentuel est intéressant à deux titres : d'abord parce qu'il les rapproche des composés du fait qu'il implique la même violation de règle accentuelle et rythmique, puisque deux accents forts ne doivent pas se suivre (voir la règle n° 2 de Guierre citée en 1.5.3.1.2), par exemple, *revalue* /210/ ; ensuite, parce qu'il reflète une identité forte sur le plan de la sémantique qui fait également penser à un élément d'un composé. Fournier, comme Guierre, distingue deux types de préfixes – séparable et inséparable (*re-cover* signifiant « couvrir à nouveau » et *recover* signifiant « se remettre ») – et tire les conclusions suivantes de son analyse (Fournier, 1992 : 83) :

> On voit que le concept prend ici valeur de 'séparabilité sémantique'. Conformément à cette observation, les préfixés à préfixe séparable se distinguent du point de vue de leur prononciation en ce qu'ils se comportent comme deux unités quasi-indépendantes : en ce qui concerne l'accentuation, le préfixe reçoit ainsi un accent /2/ et le radical un accent /1/ : prepaid, unjust, recover.

Il va même jusqu'à nous proposer une nouvelle définition du mot (Fournier, 1992, 89) :

> Cette analyse, corrélée par le comportement accentuel, nous conduit à proposer une définition du mot en anglais : un mot est une unité sémantique autonome, ou 'unité sémantique' s'interprète comme 'non séparable en unités sémantiques'. Un préfixe, parce qu'il s'agit bien d'une unité sémantique mais qu'elle n'est pas autonome, aurait alors un statut intermédiaire, statut intermédiaire manifesté par le degré de son accent : un accent secondaire et non pas un accent principal, soit, en quelque sorte, un demi-mot.

Cette définition montre bien les difficultés que nous pouvons rencontrer en sélectionnant les entrées, mais avant de procéder à l'étude systématique des entrées à sélectionner ou non, nous voudrions ajouter un autre type d'élément qui est potentiellement problématique. En effet, jusque-là, nous avons parlé de morphème lié et de morphème libre en faisant référence aux préfixes et aux

suffixes ; pourtant, il faut également mentionner une catégorie intermédiaire qui a été appelée « quasi-morphème » par Guierre, et « quasi-lexème » par Paillard (2000 : 48), qui est le terme que nous utiliserons ; Wells les appelle *combining forms*. Cette dernière appellation est celle utilisée par Chambers 1996, alors que Chambers 1988 l'ignore. En principe, il s'agit d'éléments qui se situent à mi-chemin entre l'affixe et l'élément de composition. Ces éléments sont distingués pour deux raisons : d'abord par la manière dont ils fonctionnent en construction car ils sont à la fois capables de prendre la première ou la dernière place dans le mot (avec parfois quelques modifications), ils peuvent eux-mêmes être affixés, ce qui ressemble au fonctionnement des éléments des composés ; ensuite, parce qu'ils ont une identité sémantique très forte et sont donc très utilisés pour former des néologismes, particulièrement dans des domaines techniques et scientifiques[10]. Néanmoins, ces éléments n'apparaissent pas (sauf exception) en tant que morphèmes libres. Ainsi, les mots formés à partir de quasi-lexèmes posent les mêmes problèmes de sélection que certains préfixés et suffixés. Ils figurent dans l'étude qui suit.

Nous avons distingué plusieurs raisons pour lesquelles nous hésitons entre le statut de dérivé ou bien celui de composé. La graphie, le comportement accentuel, la sémantique, la fonction et le degré de dépendance des éléments ont été cités. Nous avons également proposé des définitions qui devraient nous permettre d'arriver à sélectionner les entrées. A présent, nous allons procéder comme suit :

---

[10] Voir Trevian (2000).

d'abord nous distinguerons plusieurs catégories correspondant aux éléments qui nous ont fait hésiter ; puis nous démontrerons où se situent les problèmes ; enfin, nous expliquerons comment nous sommes arrivée à prendre une décision les concernant. Nous fixerons ainsi nos critères de sélection ou d'exclusion.

Notre démarche a été la suivante. Dans un premier temps, nous avons extrait de nos sources écrites des fichiers susceptibles de contenir une proportion importante de composés[11]. Les explications et les résultats concernant ces fichiers feront l'objet de la partie suivante. Ensuite, nous les avons parcourus afin de marquer les entrées et arriver à ne retenir que les composés en nous servant essentiellement de notre intuition d'anglophone. C'est là qu'il nous a fallu nous poser des questions. Afin d'y répondre, nous avons eu recours à deux éditions consécutives du dictionnaire Chambers (1988 et 1996), et, le cas échéant, à la version en ligne du dictionnaire étymologique Oxford. Les deux versions du Chambers sont complémentaires dans la mesure où elles présentent (presque) les mêmes données de manière différente, car leurs objectifs sont différents. La version 1988 fait état systématiquement de l'étymologie des entrées, mais pour gagner de la place, elle utilise une structure arborescente (appelée « nesting » dans l'introduction de Chambers, 1988 : vi) où les entrées composées sont présentées sous l'entrée principale (le premier élément). En revanche, la version 1996 vise une plus grande clarté et donc cette structure est remplacée par des entrées

---

[11] Ce traitement est expliqué en 2.2.1. La source est la première édition du *Longman Pronouncing Dictionary*, version électronique, qui est due à Guierre et disponible au sein de l'équipe FORELL.

individuelles ; on précise : « Compounds are given as separate entries from their base words » (Chambers, 1996 : ix). Par contre, il n'y a plus de référence systématique à l'étymologie.

Nous nous intéressons également aux changements dans la nomenclature utilisée. En effet, lorsque la version 1988 indique une catégorie susceptible d'être employée en composition, elle marque l'expression *in composition*, alors que la version 1996 emploie le terme *combining form*, absent de la version 1988. Le dictionnaire 1996 en donne la définition suivante : « Word-forming element that occurs in combinations or compounds » (Chambers, 1996 : 273). Qui plus est, il n'y a pas un simple échange dans la nomenclature car un certain nombre d'entrées n'ont plus la même étiquette dans la version la plus récente (des entrées auxquelles la version 1988 attribue la catégorie de préfixe deviennent des *combining forms*, et inversement, des entrées susceptibles d'entrer dans une composition sont étiquetées comme des préfixes).

Nous avons regroupé les éléments à étudier selon les catégories suivantes :

1/ Eléments ayant un statut simple de morphème lié (ou quasi-lexème)

2/ Eléments qui ont un statut double (affixe et morphème libre)

3/ Eléments qui se ramènent à des éléments tronqués

Nous tenterons à chaque fois de comprendre pour quelles raisons l'intuition de l'anglophone fait hésiter, et comment résoudre ces flottements.

2.1.2.1.1 Eléments à statut simple de morphème lié

Cette catégorie inclut les éléments qui correspondent au critère de la dépendance. En effet, ils ne fonctionnent jamais comme des morphèmes libres et on peut même s'étonner que l'on puisse penser les inclure en tant qu'éléments de composition. Pourtant, c'est parfois le cas. Nous allons tenter de comprendre pour quelles raisons nous pouvons être tentée d'inclure certaines de ces entrées dans notre corpus d'analyse, ce qui nous apportera quelques éclaircissements en ce qui concerne la perception de ce qu'est un composé. Nous étudierons d'abord les préfixes puis les autres éléments (en général des quasi-lexèmes) qui s'ajoutent en début de mot. Nous n'aborderons la question des suffixes que sous la rubrique suivante, car tous les suffixes qui nous ont posé des problèmes sont également des morphèmes libres.

2.1.2.1.1.1 Les préfixes

Sous cette rubrique, nous serons amenée à différencier, comme le fait Fournier, après d'autres, les préfixes séparables et les préfixes inséparables. Ce sont les préfixes séparables, ceux justement qui ont une identité forte, qui nous intéressent plus particulièrement. Nous allons en étudier quelques-uns, parmi ceux que les deux versions du Chambers classent comme des préfixes, en commençant par deux préfixes ayant un sens fortement ressenti. Ces deux préfixes sont pris à titre d'exemple, et représentent toute la classe des préfixes à une syllabe qui ont un statut simple de morphème lié. Ensuite, nous étudierons un cas de préfixe à deux syllabes, qui présente donc des particularités accentuelles; enfin, nous

étudierons un cas, *co-*, qui s'apparente aux préfixes à une syllabe mais dont le traitement par le dictionnaire est différent.

2.1.2.1.1.1.1 Les cas de *re-* et *de-*

Est-il vraiment tenable de considérer que le sens de *re-* signifiant *à nouveau* est un sens général, comme le suggère Adams (Adams, 1973 : 30) dans sa définition de la dérivation ? Son statut de préfixe n'est sans doute pas assuré par cette interprétation, comme en témoigne aussi sa prononciation. C'est plutôt la caractéristique « inséparable » qui est mise en évidence. Comme Fournier (1992), nous pensons que l'identité sémantique de ce type de préfixes est reflétée par leur comportement accentuel (qui ressemble à celui des composés). Dans le cas de *re-*, Wells (Wells, 1990 : 584) fait remarquer que le préfixe productif *re-* est toujours marqué d'un accent ; lorsque *re-* ne correspond pas au préfixe productif, l'accent se perd et une réduction vocalique s'ensuit. Il donne les indications suivantes : « As a productive prefix meaning 'again' (sometimes spelt with a hyphen), ˈˌriː (*refill* v ˌriːˈfɪl, n ˈriːfɪl). Otherwise, with a vaguer meaning, rɪ (*react* rɪˈækt».

De même, nous constatons que pour *de-* ces deux prononciations sont possibles (sans accent avec réduction pour des mots comme *defer*, mais en tant que préfixe séparable, sans réduction et avec un accent sur le préfixe). Le sens privatif du préfixe séparable *de-* modifie la teneur de la forme de base. Nous distinguons entre différentes occurrences de *de-*, préfixe séparable: par exemple, *decapitation*, *deforest*, *demilitarize*. Il y a des différences dans la prononciation du préfixe : pour le premier mot, il n'y a pas d'accent et il y a réduction vocalique, pour le second, la prononciation principale donne un préfixe accentué mais les

variantes montrent que nous pouvons très bien trouver le préfixe sans accent et avec réduction vocalique; ce n'est que pour le dernier mot qu'aucune réduction vocalique n'est indiquée, et le préfixe accentué est proposé (entre parenthèses), suivi d'un accent primaire, les variantes ne concernant que la forme de base. Nous proposons les explications suivantes pour ces trois cas de figure. Le mot *decapitation* signifie un processus qui n'a pas de mode inverse, même si *capitation* existe en morphème libre; le sens du préfixe est donc moins fort, ce qui est reflété par l'absence d'accent et la présence d'une réduction vocalique. Le mot *deforest* est sans doute ressenti comme désignant une action inverse (l'inverse de «boiser»), mais *forest* n'est pas un verbe, d'où un affaiblissement de l'identité sémantique du préfixe; nous avons donc le choix d'accentuer ou non le préfixe. En revanche, le dernier exemple, *demilitarize*, démontre clairement un processus inverse d'où l'obligation de ne pas réduire la voyelle, et la probabilité d'accentuer le préfixe.

Nous nous sommes intéressée à ces cas à titre d'illustration. Nous avons également étudié des préfixes tels que *pre-*, *post-*, *dis-*, *mis-*, *non-*, *trans-*, et les mêmes remarques sont valables: le prefixe séparable se distingue du préfixe inséparable sur le plan de la phonie et de l'accentuation; seuls, certains cas avec le préfixe séparable correspondent à une identité sémantique forte qui entraîne un schéma accentuel rappelant celui des composés. Nous voudrions néanmoins étudier deux autres cas qui apporteront des éclairages à notre argumentation: les préfixes *inter-* et *co-*.

2.1.2.1.1.1.2 Le cas de *inter-*

Nous allons prendre les exemples suivants: *intervene*, *interpose*, *interviewer*, *internationalize*, *interplanetary* et *interweave*. Tous sont des mots simples, sans trait d'union, donc la graphie ne donne aucune indication sur l'identité des composants. Les deux premiers mots n'utilisent pas le préfixe de la même manière: *intervene* est constitué de *inter* + *vene* et ce deuxième élément n'est pas un morphème libre, même si on le retrouve dans *convene*; nous avons donc un préfixe inséparable. L'accent secondaire sur la première syllabe de *inter* est un accent rythmique, obligatoire car l'élément *vene* porte un accent primaire et nous ne devons pas laisser les deux premières syllabes du mot sans accent aucun. Cet accent ne doit pas être confondu avec l'accent porté par *re-* ou *de-* qui reflète une identité forte sur le plan sémantique. Ici, c'est un accent rythmique qui est attribué en fonction des règles citées au premier chapitre (voir 1.5.3.1.2). En fait, ce mot nous est venu du latin, et le préfixe apparent n'est pas considéré comme tel actuellement. Pour le second exemple, *interpose*, bien qu'il existe un morphème *pose*, son sens en tant que morphème libre est très éloigné du sens de l'ensemble *interpose* et nous avons donc à nouveau un exemple de préfixe inséparable, accentué de la même façon que *intervene*, et pour les mêmes raisons.

Le troisième exemple est différent des autres sur le plan accentuel puisque c'est la première syllabe du préfixe qui porte l'accent primaire. Ce schéma rappelle celui des composés à accentuation simple, mais ici il s'agit de tout autre chose. Le mot *interview* nous vient du français et il s'accentue selon les règles

habituelles de l'anglais où l'accent primaire tend à être en début de mot. Le préfixe *inter* n'est tout simplement pas considéré en tant que préfixe.

Les trois derniers exemples, *internationalize*, *interplanetary* et *interweave* sont tous réellement des préfixés. *Inter-* se rattache dans chaque cas à un morphème libre; cependant, il nous semble que l'unité sémantique n'a pas le même poids pour tous. Peut-être est-ce le résultat de la grande familiarité du terme *internationalize*, ou peut-être est-ce le léger décalage de sens entre le morphème libre et le préfixé, mais nous avons l'impression que le locuteur n'a pas conscience des éléments distincts de ce mot. Il est clair que ce mot comporte le sens «entre nations», mais l'indépendance des éléments n'est pas fortement ressentie, alors que pour les deux derniers exemples (un adjectif et un verbe), nous pouvons quasiment visualiser les planètes et les fils à tisser. Ces impressions démontrent aussi à quel point il est difficile d'arriver à des conclusions nettes quand on essaie de départager les entrées en utilisant des critères sémantiques; et pourtant, c'est le seul qui reste en dernier lieu.

2.1.2.1.1.1.3 Le cas de *co-*

Wells (1990: 138) fait la remarque suivante :

> New compounds in *co-* vary between early and late stress, with a preference for late (*ˌco-arˈranger*) except when the second element has only one syllable (*ˈco-heir*). In established words, note however *ˈcoˌpilot*, *ˈcosine*.

Cette référence aux composés en *co-* est troublante. Il s'agit d'un morphème lié, mais peut-être Wells est-il influencé comme nous par la forte identité sémantique de ce préfixe qui est très productif. Nous pourrions ajouter que la présence du trait d'union renforce l'impression d'indépendance des éléments, et la remarque

concernant l'accentuation des dissyllabes en *co-*, qui ne s'accentuent que sur la première syllabe, va aussi dans le sens d'une interprétation comme composé à accentuation simple (dont seul le premier élément porte l'accent). C'est en tout cas pour ces raisons que nous avons hésité sur des entrées comme *co-star*. Au niveau de la nomenclature, il faut également émettre quelques réserves en ce qui concerne l'utilisation du terme *préfixe* par Wells, car il utilise le mot *préfixe* indifféremment pour parler de *ex-* dans *ex-actor* et dans *except* (préfixes dont le statut est différent, comme nous l'avons vu plus haut). Nous ne voyons pas très bien comment *ex-* pourrait être un préfixe alors que *co-* fabriquerait des composés puisque *ex-* existe en morphème libre (avec le sens très restreint d'ancien compagnon, il est vrai) alors que *co-* ne l'est jamais, et leurs identités sémantiques nous paraissent aussi fortes l'une que l'autre.

2.1.2.1.1.2 Les critères d'identification des préfixes

Dans tous ces exemples, il n'y a pas d'hésitation sur l'indépendance relative du morphème, car il est clairement lié (même *ex-* car le morphème libre a un sens restreint uniquement). Cependant, la présence éventuelle d'un trait d'union peut entraîner une interprétation des éléments comme étant indépendants, malgré le fait que le préfixe n'a jamais une fonction de morphème libre. Néanmoins, la cause principale de nos hésitations réside dans l'existence d'une identité forte du préfixe sur le plan sémantique. Dans nos explications du fonctionnement de *re-* et *de-*, il est entendu que le préfixe modifie le sens de la forme de base. Peut-être est-ce là que nous pourrons faire la différence entre un préfixe et un élément de composition. Le sens d'un préfixe peut être général, il peut modifier le sens de la

forme de base, alors qu'un élément de composition ne va pas forcément modifier le sens de l'autre élément ; il y apportera plutôt des précisions. Pour l'instant, nous prenons la décision de ne pas inclure ce type de préfixe parce qu'il ne fonctionne jamais comme un morphème libre, même si sur le plan accentuel et parfois le plan graphique son comportement fait penser aux composés. Une identité sémantiquement forte n'est pas une raison suffisante pour considérer ce type d'élément comme étant un élément de composition.

2.1.2.1.1.3 Autres éléments en début de mot : les quasi-lexèmes

Ces éléments correspondent surtout à ceux qui sont marqués *in composition* dans Chambers 1988 et *combining form* dans l'édition de 1996. Quelques-uns ont changé de catégorie entre les deux éditions, étant marqué *pfx* en 1988 puis *combining form* en 1996, ou (le plus souvent) *in composition* en 1988 et *pfx* en 1996. Cette tendance reflète sans doute un souci de précision dans la dernière édition, et montre, si besoin est, que la frontière entre la dérivation et la composition n'est pas définie.

Beaucoup d'éléments étudiés sont des quasi-lexèmes dont le statut semi-indépendant a été mentionné. Parmi ceux-ci, nous retrouvons sensiblement les mêmes phénomènes que pour les préfixes, à savoir : quelques éléments inséparables (*circumvent*, par exemple, à opposer à *circumnavigate* où le sens de *circum* correspond bien à l'idée « autour », et où il est associé à un morphème libre) ; une graphie avec un trait d'union (dont la fréquence augmente avec la longueur du mot) ; enfin, un comportement accentuel qui ressemble à celui des composés à accentuation simple dans certains cas (*'tele,banking*, *'demigod*). C'est

vraiment dans ce dernier cas que nous nous sommes posée la question de savoir s'il faut inclure ces entrées dans le corpus d'analyse et il est intéressant en soi que l'hésitation soit réelle et influencée par le type d'accentuation. Pourtant, nous constituons un corpus qui est destiné à être analysé au niveau des schémas accentuels des entrées ; il n'est donc pas tenable de sélectionner les entrées de ce corpus selon un critère accentuel uniquement. Nous devrons ou inclure tous les mots avec *tele-* et *demi-* (et il y en a d'autres) ou n'en inclure aucun. Nous reviendrons à cette question dans la conclusion de cette partie sur le contenu du corpus.

Nous pouvons ajouter quelques remarques concernant la prononciation de ces éléments qui est révélatrice en ce qui concerne les facteurs qui influencent la perception du locuteur natif. Wells donne la prononciation /ˈbaɪəʊ/ pour le quasi-lexème *bio-*, mais il indique que cette prononciation est sujette au phénomène de la réduction (Wells, 1990: 75) :

> In some well-known words, the weakening of the second diphthong to ə has become thoroughly established, with the consequence in RP that SMOOTHING from ˌbaɪˌə to ˌbaə is also heard. But in some other cases, where the separateness of the prefix is strongly felt, əu remains strong

C'est ce dernier cas qui nous intéresse, car le locuteur va indiquer que l'élément a une certaine indépendance en gardant la voyelle pleine, ce qui implique que la perception sera également influencée par la non-réduction et l'élément sera interprété en fonction de ce phénomène. Une première remarque s impose : nous voyons à nouveau l'interaction de la production et de la perception (et donc l'interprétation). Notons aussi le manque de cohérence dans la nomenclature utilisée par Wells qui indique la catégorie *comb. form*, puis parle de préfixe.

Nous avons étudié des quasi-lexèmes évidents tels que *ecto-*, *immuno-*, *proto-*, *ortho-*, *thermo-* (*pfx* dans Chambers 1996, comme *neo-* et *multi-*), sans réelle hésitation sur leur statut, sauf pour les quelques éléments cités plus haut. Les vraies difficultés apparaissent quand l'élément est aussi un morphème libre. Nous allons à présent étudier les éléments qui ont ce statut mixte.

2.1.2.1.2 Eléments à statut double : élément en construction et morphème libre

Sous cette rubrique, nous traitons les cas d'éléments qui fonctionnent en construction (pour reprendre le terme de Tournier cité plus haut (1985 : 137)) et qui ont par ailleurs un statut de morphème libre. Nous serons amenée à distinguer trois types de morphèmes libres : ceux dont le sens est en rupture avec le sens de l'élément en construction, ceux dont le sens est restreint (comme l'exemple de *ex-* plus haut qui ne s'applique qu'à un seul cas de figure, signifiant « ancien compagnon »), et enfin ceux dont le sens correspond à celui de l'élément en construction. Pour l'instant, nous préférons parler d'élément en construction plutôt que d'élément de composition ou bien d'affixe, puisqu'il faut justement que nous prenions une décision concernant leur statut. Nous étudierons d'abord les éléments en construction placés en seconde position, puis ceux qui se placent au début du mot.

2.1.2.1.2.1 Les éléments en construction placés en seconde position

2.1.2.1.2.1.1 Les cas de *-like*, *-most*, *-ship* et *-scape*

Ces quatre éléments illustrent très bien la diversité des cas que nous avons à traiter. *Like* et *most* sont incontestablement des morphèmes libres, ayant un sens

par eux-mêmes proche de celui que nous trouvons dans la construction. Ainsi, *antlike* signifie « ressemblant à une fourmi », et *southernmost* signifie « la direction le plus en direction du sud ». En dehors de cette correspondance sémantique ces deux éléments partagent la caractéristique de formation de classes de mots : des adjectifs pour *-like*, et des adjectifs ou des adverbes pour *-most*. Cette caractéristique est également une fonction que l'on associe aux affixes. C'est pour cette raison que nous avons pris la décision d'exclure de notre corpus d'analyse toutes les entrées formées à partir de ces éléments, considérant qu'il s'agit de suffixés. Toutefois, beaucoup d'éléments de composés forment des familles de mots ayant la même catégorie, mais il ne s'agit pas d'une création de classe entière. La question n'est pas résolue (voir Adams, 1973, 30 pour qui *-like* peut être considéré comme un élément de composition).

Si le sens de *like* et *most* correspond aux éléments *-like* et *-most*, le sens du morphème *ship* est en rupture totale avec celui de l'élément en construction. En fait, historiquement, l'élément correspond au morphème *shape* (avec un sens de « forme » très abstrait), et il était considéré comme un élément de composition. Pourtant, aujourd'hui, nous ne pouvons pas le considérer de la même façon. L'opacité de la forme, et sa non-productivité dans l'anglais contemporain, fait que cette terminaison n'est plus du tout ressentie comme étant un élément de construction par le locuteur natif ; nous avons donc décidé d'écarter les entrées en *-ship*.

L'élément *-scape* nous fournit un exemple du phénomène contraire. Le morphème *scape* existe parce qu'il a été créé à partir du mot *landscape* (donc par

« rétroformation », ou « backformation » en anglais), qui nous vient du hollandais. L'élément est apparu suffisamment intégré dans la langue pour avoir donné un certain nombre d'autres formes, telles que *cityscape*, *townscape, seascape* ou encore *Netscape*. Notre hésitation porte d'abord sur l'existence de *-scape* en tant que morphème libre, ce qui n'est pas une évidence ; pourtant, du point de vue de la phonie, il s'intègre parfaitement dans le système phonologique de l'anglais. De plus, il existe toute une famille de mots construits avec cet élément, mais il s'agit systématiquement de noms (le verbe *escape* n'a évidemment pas la même origine et ne serait en aucun cas considéré comme un composé). Nous souhaitons ne pas être trop restrictive, et dans la mesure où le morphème libre existe, où il est en relation avec l'élément en construction, où il existe une famille de mots (que nous ne considérons pas comme une classe, car elle est assez restreinte) formée à partir de cet élément, nous incluons *-scape* dans le corpus d'analyse. Ce raisonnement sera appliqué aussi à des éléments comme *-burger* qui présentent les mêmes caractéristiques.

2.1.2.1.2.1.2 Les cas de *-way*, *-ways* et *-wise*

Ces trois éléments sont traités ensemble car ils ont des sens très voisins. En effet, le morphème *way* a pour pluriel *ways*, et le sens de *-ways* se rapproche de celui de *-wise*. Si ces éléments sont associés, nous ne sommes pas arrivée aux mêmes conclusions dans les trois cas. D'après les critères que nous avons formulés, les mots en *-way* correpondent bien à notre définition d'un composé puisque l'élément *way* existe comme morphème libre avec le même sens. Par ailleurs, il n'y a pas de fonction particulière de création de catégorie ; si les mots

formés à partir de *-way* sont des noms (*highway*, par exemple), il s'agit plutôt d'une famille (c'est-à-dire un petit nombre de mots) que d'une classe (ce qui sous-entend plusieurs centaines). Nous suivrons le même raisonnement pour des mots en *-proof*, par exemple[12] (suivant ainsi Adams et Tournier, et non Marchand). En revanche, l'élément *-ways* crée des adverbes, et le sens du morphème libre est un peu décalé. Nous laisserons donc ces entrées de côté, comme nous laisserons celles en *-wise* pour les mêmes raisons (dans le cas de *-wise*, le sens du morphème est carrément en rupture car le morphème libre est un adjectif signifiant « sage » et il n'a aucun rapport avec la direction).

2.1.2.1.2.1.3 Le cas de *-meter*

Cet élément pose des problèmes particuliers. Il s'agit d'un quasi-lexème qui peut être associé à d'autres quasi-lexèmes ou à des éléments du lexique de base (morphèmes indépendants). *Meter* au sens d'un instrument de mesure[13], correspond au sens de l'élément que ce soit dans *gas meter* ou *speedometer* où on voit l'influence de la construction néoclassique avec la présence du « o », et le déplacement de l'accent sur la deuxième syllabe. Citons Fowler (1952 : 242), très critique des mélanges de la langue indigène et les formations étrangères :

> The word-makers have missed an opportunity with *meter*; we have an English METER that we use in *gas-meter* and *water-meter*; why could they not have given us *flood-meter* and *speed-meter* [...] instead of our present monstrosities?

---

[12] Ce critère de distinction entre un élément de composition et un élément de dérivation est difficile à appliquer, comme nous le voyons d'après la diversité des avis parmi les lexicologues.

[13] Selon Wells (1990), deux prononciations existent pour cette acceptation de *meter*, correspondant à nos deux exemples.

Le vrai dilemme est de savoir jusqu'à quel point nous pouvons utiliser la phonie en tant que critère de sélection, dans la mesure où nous voulons constituer un corpus afin d'analyser l'accentuation des entrées. Il est clair que *gas-meter* est un composé car il correspond aux critères qui ont été donnés (deux éléments indépendants dont le sens n'est pas en rupture avec le morphème libre). Par contre, la prononciation de *speedometer*, avec l'accent primaire sur la deuxième syllabe, résulte en l'opacification du deuxième élément qui n'est plus du tout reconnaissable car il y a une réduction vocalique qui lui fait perdre son identité. Dans les cas où un élément n'est plus reconnaissable par suite d'un processus phonologique, nous nous permettons d'exclure les entrées. Ce point sera discuté dans la rubrique suivante sous le sous-titre *les formes contractées*.

2.1.2.1.2.1.4 Les cas de *-by* et *-out*

Ces deux éléments sont à l'origine des particules de verbes qui se retrouvent dans une construction, donc des entrées comme *standby* (du verbe *stand by*) et *blowout* (du verbe *blow out*) sont incluses dans le corpus puisque le morphème libre correspond au sens de l'élément. De même, des entrées comme *blow-up* seront incluses selon les mêmes critères : ce mot est déverbal. Ce même principe sera appliqué aux éléments en construction en première position.

2.1.2.1.2.2 Eléments en construction en première position

2.1.2.1.2.2.1 Le cas de *in-*

Cet élément est étiqueté « préfixe » dans les deux éditions de Chambers mais l'existence aussi d'un morphème libre *in* est une évidence. Pourtant, la sélection des données où *in-* fonctionne comme un élément de composition n'est pas

simple. Trois cas de figure se présentent qui sont illustrés par les exemples suivants : *include*, *indoor*, *in-tray* et *inflow*. Dans le premier exemple, *in-* ne fonctionne pas comme un préfixe séparable ; le mot n'a pas de structure interne. En revanche, dans le second, *indoor*, *in-* est un préfixe signifiant « à l'intérieur de ». Pour *in-tray*, *in-* est un élément séparable avec le sens du morphème libre, et un fonctionnement similaire (ce qui est renforcé par le trait d'union). Il peut donc être considéré comme un élément de composition (pour cet exemple uniquement, Chambers, le signale ainsi, et non comme préfixe). De même, le dernier exemple, *inflow*, sera sélectionné puisque les critères proposés plus haut s'appliquent ; à savoir qu'il est déverbal (de *flow in*). Il est clair que le fait de sélectionner les entrées à inclure dans le corpus nécessite un travail d'analyse à chaque fois recommencé et une connaissance de leur étymologie et de leur relation au reste du système lexical.

2.1.2.1.2.2.2 Les cas de *pro-* et *anti-*

Ces deux éléments peuvent fonctionner comme des morphèmes libres avec le même sens que celui de la construction. Il faut écarter des entrées telles que *antipathetic* (dont le sens de l'ensemble est en décalage par rapport au sens de pathetic, signifiant « pathétique » ou « minable ») et *procreative* (où le sens de l'élément n'est pas très précis). Nous sélectionnons donc les entrées dont le sens correspond bien à *pour* et *contre*. Lorsque *pro-* signifie *à la place de* (pronom), nous le considérons comme un préfixe.

2.1.2.1.2.2.3 Le cas de *fore-*

Ce cas problématique est soulevé, entre autres, par Adams (1973 : 114) qui le traite comme un élément de composition pour les raisons suivantes : « *Fore* is not often found as an independent word; it is so used, however, in (for instance) nautical contexts [...] and so I have treated it as a particle rather than a prefix ». Nous arrivons à une conclusion un peu différente parce que nous distinguons deux sens de *fore-*, qui peut signifier « devant » (*foreleg*), et « avant » (*forewarn*). Le sens du morphème libre correspond à la situation spatiale et non temporelle, donc nous sélectionnons les entrées avec le sens spatial de *fore-*, mais non celles où *fore-*, un sens temporel (contrairement à Adams qui inclut *foreboding*, par exemple).

2.1.2.1.2.2.4 Les cas de de *out-* et *under-*

Le cas de *out* ressemble beaucoup à celui de *in* que nous avons étudié plus haut. En fait, il est étiqueté *pfx* dans la dernière édition de Chambers, alors que l'édition 1988 donne l'étiquette *in composition* (comme *over* et *up* qui seront traités de la même manière). Nous illustrons notre argumentation à l'aide des trois exemples suivants : *outrage*, *outline* et *outburst*. Dans le premier cas, nous n'avons pas deux éléments mais un mot qui vient du français et qui n'a que l'apparence d'une construction. Nous l'écartons bien entendu car *out* est un élément inséparable ; il n'y a pas de structure interne. Quant à *outline*, le lien entre le morphème et l'élément n'est pas clair. Il n'y a pas d'identité forte, donc nous l'écartons également. En revanche, le dernier exemple, *outburst*, est déverbal : donc nous le sélectionnons.

Contrairement à *out-*, *under-* est toujours étiqueté *in composition* (ou *combining form*) par Chambers. Il a plusieurs sens semblables («sous» au sens spatial et au sens qualitatif) qui peuvent se trouver dans l'élément de composition (*underlay*, *undergraduate*, *underdeveloped*). Ainsi, nous sélectionnons toutes les entrées formées à partir de *under-*, à l'exception de *understand*.

2.1.2.1.2.2.5 Les cas de de *counter-* et *by-*

Pour chacun de ces cas, nous constatons que l'élément fonctionne de différentes façons. En ce qui concerne *counter-*, le morphème libre est un verbe ayant le même sens que l'élément en construction dans beaucoup de cas (*counter-attack* par exemple), mais, comme le fait remarquer Adams (1973), il y en a d'autres où il est devenu opaque (*counterpart*) et encore d'autres où il n'est pas séparable (*countermand*). La sélection se fera donc selon le critère de la correspondance entre le sens de l'élément et celui du morphème libre, et nous écartons les entrées opaques ainsi que celles où *counter-* est un élément inséparable.

Quant à *by-*, l'élément en construction a deux sens bien distincts. Les deux ont en commun la signification «à côté de», mais l'un correspond plus précisément au sens du morphème libre et signifie proximité (*bypass*, par exemple, qui est également déverbal). Les entrées ainsi formées sont incluses dans notre corpus d'analyse. L'élément *by-* est aussi utilisé avec la signification «secondaire» et il existe une famille de mots formée à partir de ce modèle, tels que *byproduct*; nous ne sélectionnons pas ces entrées car le sens de l'élément est décalé par rapport à celui du morphème.

2.1.2.1.2.2.6 Les cas de *semi-* et *sub-*

Quelques autres éléments correspondent à des morphèmes libres, mais leur sens est restreint, comme *semi* et *sub* qui ont une signification restreinte : *semi* désigne un type de maison, et vient de *semi-detached,* alors que *sub* signifie *submarine* (ou *substitute player*)[14]. Le préfixe *semi-* (d'origine latine) a une identité forte ; nous remarquons qu'il est en compétition avec deux autres éléments : *demi-* (d'origine française) et le morphème indépendant *half.* Tournier fait remarquer que la concurrence entre matrices lexicogéniques est fréquente, et cite cet exemple de vacillement entre la dérivation et la composition (Tournier, 1985 : 162) qui reflète le contenu sémantiquement important des éléments. Cependant, nous avons déjà suggéré que la seule identité forte d'un élément ne suffit pas pour qu'on le considère comme un élément de composition.

Dans le cas de *sub-*, les hésitations que l'on peut avoir à son égard concerne plutôt des entrées très précises comme *subsoil* qui est manifestement en opposition avec *topsoil* qui, lui, correspond aux critères que nous avons définis pour des composés. Dans ces cas, nous nous permettons d'inclure le terme afin de ne pas passer à côté de certains rapprochements.

---

[14] Voir aussi la rubrique 2.1.2.1.2.3.2 « Eléments qui se ramènent à des éléments tronqués ».

2.1.2.1.2.3 Eléments en construction en première ou seconde position

2.1.2.1.2.3.1 Les quasi-lexèmes

Peu de quasi-lexèmes sont aussi des morphèmes libres, car ils se définissent comme n'étant pas autonomes, malgré leur identité sémantique forte. Ils ont aussi la particularité de pouvoir se placer en première ou en deuxième position. Les quasi-lexèmes qui fonctionnent aussi comme des morphèmes libres sont en général des unités tronquées que nous étudierons sous la rubrique suivante. Nous pouvons mentionner *radio* et *audio*, mais les deux ont un sens restreint par rapport à leur usage en construction. Nous sélectionnons les entrées dont le sens de l'élément correspond au morphème libre, et qui sont reconnaissables sur le plan phonologique (*audio-typist* /ˌɔːdiəʊˈtaɪpɪst/, qui retient son identité phonétique sur les deux éléments, mais non *audiometer* /ˌɔːdiˈɒmɪtə/ où cette identité est perdue). Il nous semble que le sens du morphème libre *super* (comme *auto*, *micro* ou encore *hyper* et *hypo*) est en décalage par rapport à son sens en construction ; nous ne sélectionnons pas ces entrées.

2.1.2.1.2.3.2 Eléments qui se ramènent à des éléments tronqués

Quelques éléments qui paraissent être des quasi-lexèmes sont en réalité des troncations et représentent non pas l'élément mais une construction. Adams a longuement étudié ce type de construction qu'elle appelle *blends* (des amalgames, d'après Tournier), et les éléments tronqués sont appelés *splinters* (Adams, 1973 : 152)[15]. Nous citons quatre exemples qui nous intéressent pour le corpus

[15] Voir aussi Paillard (2000) qui les appelle « fractomorphèmes ».

d'analyse : *cine-*, *eco-*, *para-* et *-phone-*. Lorsqu'il y a construction à partir de ces éléments en tant que fragments, il nous semble que nous pourrions les inclure car il y a correspondance entre le sens du morphème libre et l'élément. Le fragment *cine-* représente *cinema* (d'où *cinecamera*), *eco-* représente *ecology* (d'où toute la famille de néologismes ainsi formés), *para-* représente *parachute* (qui explique *paragliding*), et *phone* représente *telephone* (ce qui permet *phone card*, mais *phone* a atteint le statut de morphème libre de toute façon).

Cependant, ce type d'entrées n'est pas sans poser de problème, justement au niveau de la prononciation. Adams (1973: 152) fait état de ces difficultés :

> The greatest difficulty in pronouncing blends is in deciding which syllable should take the primary accent. My informants found this kind of difficulty with well over half the examples I asked them to pronounce.

Ce type de construction comprend également des deuxièmes éléments tronqués, et c'est peut-être là que la difficulté est la plus ressentie (ou bien lorsqu'il y a troncation des deux). Nous pensons à *breathalyzer*, par exemple. Nous nous proposons d'exclure ce type d'entrée lorsque les éléments ne sont pas pleinement reconnaissables, mais dans les cas des quatre exemples cités plus haut, nous pensons que l'identité du fragment est suffisamment forte pour représenter sans ambiguïté l'élément, et que nous pouvons donc justifier leur inclusion dans le corpus. Nous reviendrons brièvement à la question des amalgames dans la discussion de l'étendue du corpus.

2.1.2.2 Les syntagmes et les composés

Cette question épineuse a été abordée dans la discussion des définitions de ce qu'est un composé. Bauer (1988 : 102) établit la ressemblance entre les

composés et les dérivés par opposition aux syntagmes: « like derivatives, compounds provide names for entities or actions. This is opposed to providing descriptions which is the function of syntax ». Toutefois, la difficulté de déterminer le statut de certaines séquences est soulevée par Lyons (1977 : 535) qui distingue les catégories suivantes : « word-compound », « syntactic compound » et « compound lexeme ». Il utilise le critère accentuel pour classer les unités dans la première catégorie : si elles sont à accentuation simple (*greenhouse* ou *red coat*, par exemple), alors ce sont des « word-compounds »[16] ; mais il reconnaît que la distinction entre les deux autres catégories n'est pas claire : « It is hard to draw a distinction between regularly constructed, but institutionalized, syntactic compounds and petrified compound lexemes »[17]. Il classe l'unité *country house* dans la dernière catégorie d'après un critère sémantique, car le sens de cette unité est plus restreint que ne laissent entendre les composants ; en revanche, l'unité *stone wall* serait considérée comme un « syntactic compound ».

Le critère accentuel est celui qui est le plus souvent cité pour départager ces deux types de formations. Les génératives se fient à ce seul critère, mais leur approche est souvent critiquée, par Ladd entre autres. Pour prendre son exemple (Ladd, 1984 : 253), il fait état de la différence dans le schéma accentuel entre

---

[16] Lyons (1977 : 525) oppose « word-compounds » à « phrasal compounds » qui ont une accentuation double.

[17] Lyons fait la distinction entre « institutionalization » et « lexicalization » (qui est associé à ce qu'il appelle « petrification », qui correspond au « figement » de Gross (1988). Voir aussi les exemples de Lutstorf (1963 : 145) qui démontrent toute l'échelle de figement à partir du lexème simple *man* jusqu'au syntagme *a military man*, en passant par les étapes intermédiaires : *woman, gentleman, businessman*.

'green 'house (séquence adjectif + nom, où *green* qualifie *house*) et *'greenhouse* (unité composée où l'ensemble désigne un abri en verre où on fait pousser des fleurs et des légumes), et rejette l'hypothèse soutenue par Chomsky & Halle selon laquelle ce genre de différence peut s'expliquer en termes de syntaxe. Il propose d'autres explications, en faisant référence à la valeur de la désaccentuation du deuxième élément qui indique la présence d'une certaine relation entre les éléments. L'hypothèse selon laquelle la désaccentuation du deuxième élément indique une relation préconstruite entre les éléments est également soutenue par Huart (1989 : 80) :

> L'accent unique sur le premier terme fonctionne donc comme la marque en surface qui signale à l'interlocuteur qu'il est censé être en mesure de reconstituer le repère qui manque.

Nous reviendrons dans le chapitre 4 (4.2.2.2.1) à cette discussion, mais dans l'immédiat, il s'agit de déterminer des critères afin de sélectionner les unités que nous considérons comme étant des composés, et il en existe qui sont souvent considérés comme tels et qui ont une accentuation double (voir plus haut les réserves qu'a émises Jespersen vis à vis du critère accentuel qui est trop restrictif (Jespersen, 1961), même s'il est fiable).

Nous ne pouvons pas appliquer le critère accentuel afin de déterminer le statut d'une unité, car nous constituons un corpus afin de l'analyser de ce point de vue. En revanche, il existe d'autres critères, comme le critère syntaxique. Nous précisons que le problème se pose pour les entrées écrites avec un espace graphique entre les éléments, et plus particulièrement des séquences adjectif + nom, comme *red tape*, par exemple, à laquelle nous avons déjà fait référence (en

2.1.1.1.2). En effet, dans ce cas, nous ne pourrons pas insérer un autre adjectif entre *red* et *tape* car il s'agit bien d'une unité et non d'un adjectif suivi d'un nom ; nous ne pourrons pas non plus mettre l'adjectif en position d'attribut. Cependant, ce type de test a des limites.

En ce qui concerne notre sélection, nous avons décidé d'adopter une approche non restrictive afin de ne pas passer à côté de rapprochements intéressants. Ainsi, nous incluons les entrées des sources écrites (dictionnaires lexico-phonétiques) correspondant à cette graphie, qui sont à considérer comme des composés non compacts potentiels. Leur présence dans ces sources montre qu'ils ont un statut – au moins en termes de fréquence, si ce n'est au niveau de la lexicalisation[18] – qui leur permet d'y figurer. En ce qui concerne les sources orales, nous ne disposons pas de critère graphique de sélection, donc nous avons retenu des séquences qui nous semblent désigner un concept unitaire (voir la définition de Picoche plus haut (Picoche, 1993 : 17)).

2.1.2.3 Les éléments non reconnus

Sous cette rubrique, nous allons étudier un problème qui a déjà été soulevé dans la discussion concernant la dérivation et la composition : le cas où un élément n'est pas reconnu en tant que morphème libre. Nous nous sommes souvent fiée au critère de l'existence d'un morphème libre pour distinguer les dérivés des composés, ce qui nous paraît justifié dans ce cas. Mais cette approche

---

[18] Contrairement à Lyons (1977), nous ne faisons pas de distinction entre « institutionnalisation » et « lexicalisation ». Dans nos travaux, le terme « lexicalisation » indique qu'une unité a atteint le statut lexical de l'entrée dans le dictionnaire.

se justifie-t-elle de la même manière lorsque nous essayons de départager les lexèmes simples des unités complexes? Cette question a été soulevée dans trois contextes différents : lorsque le sens global de l'entrée est inconnu ; lorsque le sens de l'entrée est connu, mais un des éléments n'est pas reconnu ; enfin, lorsque nous avons à faire à ce qui était historiquement un composé.

2.1.2.3.1 Entrées inconnues

En premier lieu, là où nous avons rencontré des entrées inconnues, nous avons dû chercher leur origine, puis décider de leur statut. Dans ces cas, nous avons fait confiance aux explications qui nous ont été fournies par le dictionnaire, et nous avons effectivement utilisé ce critère. Toutefois, il est intéressant de noter que l'intuition de l'anglophone permet de prédire l'origine d'un nombre important d'entrées, ou par une reconnaissance partielle des éléments, ou par les associations phonologiques dont nous avons parlé dans la discussion des critères phonologiques pour définir un composé. C'est ainsi que nous pensions que des entrées comme *pushpin* (un jeu d'enfants où il faut pousser des petites quilles) aurait un statut de composé : c'est effectivement le cas. Pourtant, il y a parfois des surprises, et ce dans les deux sens (des entrées que nous pensions avoir un statut de composé et qui ne le sont pas, et inversement). Les erreurs dans les prédictions étaient souvent dues au fait que les mots étaient à l'origine des noms propres, comme *billycock* qui désigne un type de chapeau ; l'origine du mot serait de William Coke (d'une famille noble anglaise) ou encore William Cock (fabricant de chapeaux en Cornouaille). D'autres, malgré leur aspect parfaitement anglais, et une prononciation tout à fait intégrée, sont aussi des lexèmes simples d'origine

étrangère (*cockroach* qui vient de l'espagnol *cucaracha*). En revanche, le terme *sheldrake* nous a surpris dans la mesure où il s'agit d'un composé ; il est à opposer à *shelduck* et l'élément *shel* serait de l'anglais dialectal *sheld*. Pourtant, le premier élément ne nous a pas paru s'intégrer dans le système graphique de l'anglais (nous nous attendons à deux *l*). Il existe d'ailleurs des variantes graphiques : *shieldrake*, dont la graphie du premier élément est aussi éloignée du système anglais, ou *shelldrake* qui, lui, s'intègre parfaitement.

2.1.2.3.2 Entrées connues

L'autre contexte où nous devons poser cette question est lorsque nous rencontrons des entrées connues qui, d'après notre intuition d'anglophone, sont ressenties comme étant des composés, mais sans que nous puissions reconnaître les deux éléments. Parfois, la non-reconnaissance de l'élément est le fait de notre manque de connaissance, comme pour *copestone* qui est une pierre que l'on met en haut d'un mur (du verbe *cope*, qui désigne l'action de mettre une pierre dans cette position) ou bien *saltlick* (qui désigne un endroit où vont les animaux pour avoir du sel). Néanmoins, il y a d'autres cas bien connus où l'élément n'existe vraiment pas. Il s'agit de mots comme *bilberry* (et *cranberry*[19] ou *raspberry*, pour prendre des exemples d'une même famille) ou *woodruff*. Cette question a été longuement discutée par Marchand (Marchand 1969 : 11) qui prend le point de vue très restrictif de ne considérer comme composés que les entrées ayant deux

[19] Bauer (1988) intègre ces cas en leur attribuant un statut particulier. L'élément non reconnu est considéré comme « unique morph », ce qui implique que son existence n'est attestée qu'une seule fois. Ce type d'unité a aussi été appelé « cranberry morph ».

éléments indépendants. Nous pensons, comme Boisson, que cette approche est trop contraignante. Ce dernier critique le point de vue de Marchand (Boisson, 1980 : 256-7) :

> Cela me paraît trop exigeant, au moins au regard du sentiment linguistique du locuteur. Pourquoi s'interdire en effet de considérer "cranberry" comme composé sous prétexte que "cran-" est un morphème lié : il suffit – du moins me semble-t-il – que le second morphème soit libre.

Nous suivrons cette suggestion, ce qui fait que nous incluons des entrées telles que *bilberry*, mais nous excluons *woodruff*, faute de précisions sur le second élément. Il nous paraît important de tenir compte de l'intuition de l'anglophone.

2.1.2.3.3 Les composés historiques

Pour terminer cette discussion des critères à utiliser pour déterminer le statut de composé, et donc l'inclusion ou non dans le corpus d'analyse, nous passons à l'étude des problèmes que nous pose une classe d'entrées que nous allons appeler des composés historiques. Nous avons souligné que nous avons eu recours au dictionnaire pour des précisions étymologiques. Parfois, ces précisions nous ont confirmé le statut de composé des entrées, mais pouvons-nous justifier l'inclusion de ce type d'entrée dans le corpus si nous avons besoin d'indications du dictionnaire pour déterminer le statut ?

Nous distinguons deux cas de figure : les entrées où il y a opacité des deux éléments et celles où un des éléments est reconnu. Le mot *tadpole* est un exemple du premier cas. L'origine du premier élément est *tāde* signifiant *toad* en vieil anglais (qui devient *tood* en moyen anglais), suivi de *poll* qui signifie *head*. Nesfield (1939 : 213) cite cet exemple et d'autres :

> In words of two or more syllables the accent is usually thrown back on the first syllable, and in many cases this has the effect of altering the spelling :- *bon-fire* from *bone-fire. Hus-band* from *house-band* (lit. house-dweller) ... *Tad-pole* from *toad-poll. Fort-night* from *four-teen-night. Suf-folk* from *South-folk. Fur-long* from *furrow-long.*

Nous notons avec intérêt ses remarques concernant l'effet de l'accentuation sur la graphie, mais une étude de ce genre de phénomène n'a pas sa place dans notre analyse que nous situons au niveau de l'anglais contemporain. Dans tous ces exemples, il y a un certain degré d'opacification. Nous nous intéressons d'abord au cas de *tadpole* : le premier élément est inconnu aujourd'hui et le second, à cause du changement graphique, n'est pas reconnaissable (et ce sens de *poll* n'est pas non plus très courant, même si on le retrouve dans *pollard*). De même, *husband* est totalement opaque. Dans la mesure où les éléments ne sont pas ressentis comme étant des éléments de construction, nous pouvons écarter ce type d'entrée de notre corpus. Nous restons fidèle au sentiment linguistique dont parle Boisson (1980 : 257).

Dans les autres exemples, *bonfire*, *fortnight*, *Suffolk* et *furlong*, le second élément reste identifiable, mais le premier est rendu opaque par l'abrègement phonétique[20] et le changement orthographique, ce qui fait que l'ensemble n'est plus ressenti comme un composé. De même, dans le cas de *cowslip* et de *midriff*, nous reconnaissons le premier élément mais non le second : l'ensemble *cowslip* vient du vieil anglais *cūslippe* où *cū* signifie *cow* et *slippe* signifie *slime*, et *midriff* a la même origine (*mid*, "middle" + *hrif*, "belly"). Citons les propos de Benveniste

---

[20] Les changements phonétiques sont aussi relevés par Thun (1963: 230), Marchand (1969: 5) et Deschamps et al. (2000: 73) comme étant indicatifs d'une perte de motivation. Ils impliquent que le mot n'a plus de structure interne et n'est donc pas à considérer comme un composé.

(1974 : 171) qui nous aideront à prendre une décision concernant ce type d'entrées :

> Il y a composition quand deux termes identifiables pour le locuteur se conjoignent en une unité nouvelle à signifié unique et constant [...] Il faut en tout cas que le locuteur puisse isoler et identifier les deux termes [...] Il faut aussi que le sens des membres pris isolément ait un rapport à peu près intelligible avec celui du composé.

Nous ne reviendrons pas sur les deux premiers points dont nous avons déjà discuté (la fonction de création de classes génériques et l'indépendance des morphèmes), mais le dernier point est significatif. En fait, cela explique pourquoi les mots cités ci-dessus ne sont plus ressentis comme des composés, car l'opacification d'un des éléments a fait disparaître le lien sémantique avec l'ensemble. Ceci est vrai dans tous les cas, que ce soit le premier ou le second élément qui ne soit pas reconnaissable (contrairement à des entrées comme *bilberry*).

Enfin, nous allons voir deux autres exemples d'entrées où nous ne reconnaissons pas un des éléments du fait qu'il a perdu sa transparence depuis le stade du vieil- ou du moyen-anglais : *nickname* (du moyen anglais *necke-name* de *eke-name* avec le *n* de l'article indéfini, *eke* signifie addition) et *walnut* (du vieil anglais *walhnutu*, de *w(e)alh*, *foreigner*, et *hnutu*, *nut*). Dans ces deux cas, nous reconnaissons le second élément, et nous avons le sentiment qu'il s'agit de composés, car il y a une relation évidente entre le sens du terme et l'élément reconnu. Ces entrées sont donc incluses dans le corpus d'analyse.

Nous proposons d'adopter le même principe de sélection en ce qui concerne la famille des *-wort*. Il existe un nombre important d'entrées, par exemple, *bladderwort*, désignant des types de plantes qui finissent par cet élément (il est attesté par les dictionnaires). Il est vrai qu'il n'est plus très utilisé en morphème

libre à l'heure actuelle, mais il n'est pas non plus devenu opaque, et c'est pour cette raison que nous incluons ces entrées dans le corpus.

2.1.3 Les critères de sélection des composés pour le corpus d'analyse

Nous avons étudié les cas limites où nous hésitions sur le statut de composé de certaines entrées. La frontière entre la dérivation et la composition a fait l'objet d'une étude approfondie et nous sommes arrivée à la conclusion que nous pouvons généralement nous fier au critère de l'identité forte des éléments, et de l'existence d'un morphème libre, au moins pour le second élément, si cet élément a le même sens que ce morphème afin de departager les composés et les dérivés. Notons que notre interprétation du critère sémantique est beaucoup plus restrictive ici qu'elle ne l'est par la suite. Nous utilisons quelquefois des critères phonologiques pour déterminer la reconnaissance de ces morphèmes. Ensuite, nous avons fait état de deux autres difficultés: la différenciation entre un syntagme et un composé et l'identification des entrées composées lorsqu'il y a méconnaissance des éléments. Sous ces deux rubriques, nous sommes partie des mêmes critères que précédemment, mais nous avons été amenée à les modifier. Le critère de l'indépendance du morphème est appliqué au second élément uniquement; le critère sémantique devient non plus une correspondance stricte entre le sens d'un élément et le morphème, mais une simple relation. Cette définition des critères est impossible à appliquer pour départager des dérivés et des composés, car elle aboutirait à une sélection quasiment sans limites. En revanche, elle est utile quand on traite d'autres types d'entrée pour permettre

l'inclusion dans le corpus d'un certain nombre d'entrées ressenties comme ayant un statut de composé.

A présent, nous allons passer à une discussion plus générale qui concerne également le contenu du corpus d'analyse. Ayant déterminé quelles sont les entrées que nous pouvons considérer comme étant des composés, nous allons voir quelle devra être l'étendue du domaine que nous souhaitons soumettre à l'analyse. Ce sera la conclusion de cette partie.

2.1.4 Quelle sera l'étendue du corpus d'analyse ?

Nous avons d'abord tenté de répondre à la question: qu'est-ce qu'un composé ? Nous passons à présent à une autre question : que voulons-nous analyser ? Le lexique contenu dans nos sources écrites comporte différentes catégories : elles ne présentent pas toutes les mêmes caractéristiques, ni le même intérêt. Afin de nous donner les moyens de faire une analyse de l'accentuation dans les mots composés de l'anglais contemporain, notre corpus doit contenir des données cohérentes. Nous avons donc été amenée à en écarter certaines, ou bien à en limiter la sélection. Nous allons présenter les domaines où les entrées sont susceptibles d'être en marge de nos objectifs, et nous fixerons les limites de l'étendue des entrées à étudier.

2.1.4.1 Composés étrangers

Cette catégorie d'entrées présente certaines des caractéristiques des composés avec des éléments non reconnus que nous avons étudiés plus haut, particulièrement les composés historiques. En effet, nous avons souvent recours à l'étymologie pour comprendre leur statut de composé, même si nous les

reconnaissons parfois, à condition d'avoir une petite connaissance de la langue d'origine : ainsi, si nous avons fait un peu d'allemand, nous comprenons *bratwurst* qui désigne une espèce de saucisse. Néanmoins, le vrai problème des mots étrangers, y compris les composés, est leur non-intégration dans le lexique de base de l'anglais, et donc leur non-représentativité. Cela est évident sur le plan graphique puisque nous trouvons des séquences de lettres que nous ne trouvons pas dans le lexique indigène (un *k* en fin de mot par exemple), mais cette non-intégration est surtout importante sur le plan phonologique. Le comportement accentuel d'un emprunt ne reflète pas forcément les tendances du lexique indigène. Ainsi, ce type d'entrées ne peut pas nous apporter des renseignements fiables pour notre analyse. Il y a un grand nombre de mots et d'expressions de ce genre qui sont d'origine française, latine, allemande, hollandaise et bien d'autres, que nous écartons de l'analyse parce qu'ils risquent d'être atypiques.

Certains composés étrangers ne le sont que pour partie (mot étranger + mot indigène ou inversement). Le statut de composé de telles entrées est facilement reconnaissable, et nous avons des exemples comme *iceberg*, *faldstool* ou *faltboat*. Il est intéressant de constater qu'ils comportent une des caractéristiques des composés, à savoir la non-réduction de l'élément inaccentué. Cependant, nous pouvons également considérer qu'il s'agit d'une marque de non-intégration dans le lexique indigène. Même si l'ensemble est facile à comprendre, l'élément étranger reste un élément non reconnu du lexique de base, et il est prudent d'écarter également ce type de formation semi-étrangère, car nous n'avons pas de garantie d'un fonctionnement représentatif.

Le dernier exemple, *faltboat*, est un peu exceptionnel. Il a une variante graphique *foldboat* qui comporte deux éléments tout à fait reconnaissables du lexique de base. Notons toutefois que la formation base verbale + nom n'est pas celle que l'on attend; la formation verbe en -ing + nom serait plus conforme à l'usage, mais il n'en reste pas moins que cette entrée est en voie d'intégration. Nous nous proposons de la retenir à ce titre.

2.1.4.2 Quasi-lexèmes

Nous revenons à la question des quasi-lexèmes qui a déjà été discutée plus haut, parce que ces entrées forment une catégorie importante et sont souvent associées aux composés. Nous devons donc préciser l'étendue de notre champ de recherche. C'est une question à laquelle il n'a pas été facile de répondre; nous sommes passée par plusieurs étapes intermédiaires et si nous avions vite écarté l'idée de les étudier tous en raison de leur fonctionnement spécifique, et du nombre d'entrées que cela représente, nous avons été longtemps tentée d'étudier les composés formés d'un quasi-lexème et d'un morphème libre. L'appellation *QLM* pourrait même être proposée (quasi-lexème + morphème libre). Finalement, nous avons décidé de restreindre le corpus au niveau des entrées quasi-lexématiques aux quelques cas où nous pouvons considérer qu'il s'agit d'un morphème libre (*mania* dans *beatlemania*), ou de la troncation d'un composé quasi-lexématique (*para* dans *paragliding*).

2.1.4.3 Noms propres

Ce problème est apparu très vite lorsque nous marquions les fichiers susceptibles de contenir des composés, tout d'abord en raison du nombre très

important d'entrées de ce type. Comme nos fichiers sont extraits de dictionnaires lexico-phonétiques, il est évidemment normal d'y trouver des indications de ce genre. Nous allons étudier les différents cas de figure et décider si nous soumettons ces entrées à l'analyse.

2.1.4.3.1 Noms propres étrangers

Il est clair que les noms propres étrangers doivent être traités de la même manière que les autres entrées d'origine étrangère. Des entrées telles que *Schwarzkopf* peuvent être reconnues comme ayant un statut de composé si on connaît l'allemand, mais nous n'aurons pas plus de garanties de la représentativité sur le plan accentuel de ce type d'entrées que pour des noms communs et autres entrées du lexique qui sont d'origine étrangère. Il faut donc les écarter du corpus d'analyse.

2.1.4.3.2 Noms propres indigènes

Voyons d'abord toutes les entrées avec une majuscule qui désignent un nom de lieu ou de personne. Ces entrées présentent souvent des caractéristiques très spécifiques qui rappellent celles des composés historiques. En effet, si certains de ces noms propres sont à l'évidence des composés, comme *Northwood* (qui devait être à l'origine un lieu boisé au nord, ou bien au nord d'un bois – il n'est pas facile de les interpréter avec exactitude), d'autres, comme *Northwich*, sont moins transparents. Voici un domaine d'étude passionnant, mais qui demande une approche diachronique, et une recherche dont l'étendue dépasse très largement les limites de notre analyse. C'est pourquoi nous avons décidé d'écarter ce type d'entrées du corpus d'analyse, à une exception près. Il s'agit des noms de rue qui

présentent un des problèmes au niveau de l'accentuation des composés dont on a le plus parlé : pourquoi dit-on ˌ*Wood* ˈ*Road* avec une double accentuation, mais ˈ*Wood Street* avec une accentuation simple ? Nous ne pouvons pas négliger ces données.

De même, il y a beaucoup de noms propres désignant des nationalités, et quand il y a double nationalité, il y a une forme de construction, comme dans *Anglo-Irish*. Ces entrées semblent toutes fonctionner de la même façon du point de vue accentuel, c'est-à-dire avec une accentuation double. Le fait d'adopter une construction quasi-lexématique repose la question de cette catégorie d'entrées que nous avons fini par écarter de l'analyse. C'est pourquoi nous écartons tous les noms de nationalité double.

En revanche, lorsqu'il y a une combinaison d'un nom propre et d'un nom commun, ce qui arrive fréquemment dans les composés non compacts pour les noms de maladie, par exemple (*Down's syndrome*), nous avons décidé de les inclure dans le corpus d'analyse. Ainsi, les composés formés à partir du nom de l'inventeur d'un objet (ou d'une loi, ou d'autres processus) associé à son invention sont soumis à l'analyse. Nous avons également inclus les noms propres désignant des occasions ou événements tels que *Easter Day*[21].

---

[21] Ces entrées sont analysées en 4.6.2, « Etude des entrées ayant une majuscule ».

2.1.4.3.3 Noms propres tdmk

Il y a un certain nombre d'exemples d'entrées auxquelles on a attribué l'abréviation *tdmk* pour désigner les noms de marque. Ce ne sont pas des entrées tout à fait comme les autres et beaucoup ne sont que des noms propres simples, comme *Librium* ou *Levi's* (qui est en fait une troncation que nous ne pouvons pas considérer en tant que composé car il n'y a plus qu'un seul élément). Néanmoins, certains d'entre eux sont des créations significatives, ce qui représente une source intéressante de néologismes. Les deux exemples suivants serviront d'illustration : *Blu-Tack*, qui est une sorte de pâte bleue utilisée pour coller des affiches sans laisser de trace (de *blue* et *tack*) ; et *Bostik*, qui est une marque de colle. Ici, seul le deuxième élément *stick* est reconnaissable, avec une orthographe qui se veut moderne pour attirer l'attention, mais nous avons déjà justifié la décision de considérer les entrées dont le second élément est reconnu. Nous avons donc inclus ce type d'entrée dans le corpus d'analyse.

2.1.4.3.4 Acronymes

Nous avons davantage hésité au niveau des acronymes. A partir du moment où une lettre représente quelque chose, et que la suite a une signification, nous pouvons considérer ces entrées comme des composés. Leur schéma accentuel s'apparente plutôt à un schéma de l'accent en discours, ce qui nous paraît intéressant. Toutefois, ces entrées sont réellement en marge, car du point de vue sémantique, nous ne pouvons pas faire de rapprochement au niveau des relations des éléments entre eux. C'est pourquoi nous les écartons du corpus d'analyse. En

revanche, nous étudions les acronymes qui entrent dans une construction, telle *ENT specialist* ou *A road* (en 4.6.2).

2.1.4.4 Formes contractées

Une forme contractée est par définition un composé car il s'agit de deux éléments écrits en un seul mot graphique (sans espace). Au niveau de la forme, nous pourrions argumenter que certaines formes contractées (en *n't*) ressemblent à des préfixes ou à des suffixes, ayant une sorte de fonction négative, mais sur le plan de la sémantique, l'élément contracté a une forte identité. Sa particularité reste que l'un de ces éléments n'est pas complètement reconnaissable sur le plan graphique, puisqu'une apostrophe remplace une (ou plusieurs) des lettres de sa graphie, et cette caractéristique reflète ce qui s'est passé au niveau phonologique : il y a eu réduction. Nous pouvons ajouter la suggestion de Marchand (Marchand, 1969,: 3) à propos des éléments de composition *-land*, *-man* et *-berry* qui, suggère-t-il, ont un caractère semi-suffixal (ils sont souvent réduits, justement), mais nous procèderons plutôt de la manière inverse : ce qui est souvent considéré comme un préfixe ou un suffixe est inclus dans le corpus d'analyse, et les entrées en *-land*, *-man* et *-berry* y ont sûrement leur place. Nous avons utilisé le phénomène de la réduction pour écarter certaines entrées lorsque nous avons considéré que les modifications phonologiques (et accentuelles) rendaient impossible la reconnaissance d'un élément *(-meter*, par exemple), ce qui nous permet d'écarter les entrées qui sont des formes contractées, mais la question de leur statut n'est pas résolue.

2.1.4.5 Amalgames

Nous terminons cette partie avec une catégorie de mots à laquelle nous avons fait référence dans le cadre de l'argumentation concernant les quasi-lexèmes. Cela a permis l'inclusion dans le corpus de certaines entrées quasi-lexématiques parce que nous avons considéré que l'élément était le résultat d'une troncation d'un composé quasi-lexématique. L'ensemble qui en résulte est une sorte d'amalgame (l'exemple donné est *paragliding*).

Cette catégorie de mots ne se limite pas à ce type de formation. Nous avons déjà indiqué que l'un ou l'autre élément peut être tronqué, ou bien les deux. Ces entrées sont fascinantes, d'abord parce qu'elles sont souvent des créations nouvelles et permettent d'étudier les néologismes, ensuite parce qu'elles sont assez imprévisibles. Cette dernière caractéristique trouve son reflet dans les difficultés de prononciation dont nous avons parlé. Il est tentant d'inclure une catégorie de mots qui comporte des entrées tels que *permafrost* (*permanent frost*) et *kissogram* (formé de *kiss* + *o* épenthétique des formations quasi-lexématiques + *gram*, de *telegram*). Pourtant, il est peut-être plus prudent de les écarter de notre corpus d'analyse car ils ont des caractéristiques multiples peu représentatives, qui entraînent des questions au niveau phonologiques et la composition exclut la voyelle épenthétique.

2.2 Sélection des données : sources écrites

Dans la première partie de ce chapitre, nous avons déterminé quels sont les critères de sélection pour le corpus d'analyse. Nous allons à présent expliquer les

méthodes que nous avons adoptées afin d'extraire les données à inclure, d'abord en ce qui concerne les sources écrites, puis les sources orales.

2.2.1 Choix des sources

Nous avons choisi d'utiliser des dictionnaires électroniques car ils fournissent une source d'informations pertinentes pour notre analyse, et leur forme électronique nous permet d'extraire les données qui nous intéressent de façon rapide et surtout sans erreur, grâce à des traitements automatiques. Nous disposons d'un certain nombre de ces sources : *Oxford English Dictionary* (OED) sur CD-ROM, *Longman Pronouncing Dictionary*, édition 1990 (LPD90) au laboratoire FORELL en liaison avec Guierre, *New Pronouncing Dictionary* de Guierre (NPD), *Cambridge International Dictionary of Contemporary English*, édition 2000 (*CIDE*) et le fichier des composés extrait par Guierre de la 12e édition du *English Pronouncing Dictionary* (EPD12).

2.2.1.1 Les informations

Nous avons besoin tout d'abord de la graphie. En fait, nous constatons que beaucoup de ces dictionnaires indiquent non seulement la graphie mais aussi la graphie inverse. Cette information est précieuse dans un travail sur les composés car elle permet d'accéder aisément aux données graphiques par le début du mot (donc le premier terme) ou par la fin (le deuxième terme) sans préjuger de l'orthographe avec ou sans trait d'union. Elle permet également les rapprochements par rapport à des suffixes ou les désinences (deuxième terme se terminant par *-ing* ou *-ed* par exemple).

Nous nous intéressons également à la qualité des informations syntaxiques. Bien entendu, peu de fichiers indiquent qu'une entrée a un statut de composé, et les informations portent sur l'entrée dans son ensemble (verbe, nom, adjectif, etc.). Aucun ne donne des informations concernant chacun des termes, du moins pas dans l'entrée du composé. Pour obtenir cette information, il faut se reporter à l'entrée du terme en tant que mot isolé, mais il y a des cas d'ambiguïté quand il s'agit d'éléments plurifonctionnels. Le mot *lighthouse*, par exemple, est-il un endroit où il y a un éclairage (dans ce cas *light* est un nom), ou bien un endroit pour éclairer (où *light* serait un verbe)? De même, pour *payday*, s'agit-il du jour de la paie (où *pay* est un nom) ou du jour que l'on vous paie (où *pay* est un verbe)?

Certaines sources indiquent la phonie qui fournit des informations au niveau de la coupe syllabique et surtout au niveau de la prononciation des voyelles des syllabes autres que celle qui est accentuée. En effet, lorsque l'on s'intéresse à l'accentuation, les autres segments sont affectés et une analyse des syllabes non ou moins accentuées bénéficie de l'apport de ces informations. Les éventuelles variantes phoniques sont également intéressantes puisqu'elles apportent des informations complémentaires qui seront importantes à prendre en compte pour tester les hypothèses qui seront dégagées par l'analyse de la prononciation principale.

Quant à l'accent, pour optimiser l'étude des schémas accentuels, il est impératif de pouvoir consulter des indications claires, de préférence numériques (précisant la valeur du schéma accentuel sous forme de chiffres); c'est donc une

information essentielle pour notre analyse. Certaines sources indiquent, de plus, le nombre de syllabes du mot, ce qui peut faciliter certaines recherches et permettre certains rapprochements.

2.2.1.2 Les sources non retenues

Nous avons procédé à des tests sur un échantillonnage de différents dictionnaires électroniques pour déterminer lesquels correspondent à nos critères et dans quelle mesure nous pouvons les utiliser afin de constituer notre corpus. Nous avons écarté le *Oxford English Dictionary* qui présentent des incohérences informationnelles qui rendront impossible un traitement automatique des données. Zumstein (2000 : 10) fait une étude approfondie de la question, et arrive aux mêmes conclusions.

> Il faut en conclure ici que l'exhaustivité lexicale des fichiers fondés sur 'OED' ne peut être un critère déterminant dans le choix du corpus. Les nombreux problèmes que posent les déficiences informationnelles et variationnelles des fichiers nous ont conduit à écarter ces fichiers en tant que corpus de référence.

Nous avons également décidé d'écarter le *Cambridge International Dictionary of Contemporary English*, bien qu'il marque les composés. L'inconvénient de cette source est qu'elle ne permet pas d'extraire et donc d'exploiter les données qui nous intéressent ; nous ne pouvons que les consulter. Les entrées correspondant à une sélection selon le critère *composé* ne s'écrivent qu'avec trait d'union ou bien en deux mots, ce qui laisse supposer qu'il en manque beaucoup. Par ailleurs, ce dictionnaire n'indique que la graphie et la phonie, sans indiquer le schéma accentuel en clair ni la graphie inverse.

2.2.1.3 Les sources retenues

Nous avons retenu les trois autres dictionnaires: *Longman Pronouncing Dictionary*, édition 1990 (LPD90), *New Pronouncing Dictionary* de Guierre (NPD), et le fichier des composés extrait par Guierre de la 12ème édition du *English Pronouncing Dictionary* (EPD12). Nous allons faire un bref descriptif des caractéristiques de la version électronique de ces trois dictionnaires.

2.2.1.3.1 *English Pronouncing Dictionary*, 12ème édition (EPD12)

Deux versions électroniques de ce fichier existent. La version originale, appelée *Rhymes8*[22], inclut la graphie, l'indication syntaxique du composé, la phonie (dans un code particulier, avant SAMPA[23], utilisant exclusivement des majuscules qui sont d'ailleurs réservées aux seules indications phoniques) et les schémas accentuels sous forme numérique. La deuxième version, *Rhymes9*, a été phonétisée par J.-M. Fournier qui a utilisé IPA Times (qui est proche de SAMPA mais non identique) pour indiquer la phonie. Dans les deux versions, le système de hiérarchie accentuelle est également celui que nous employons actuellement, c'est-à-dire, /1/ indique l'accent primaire, /2/ l'accent secondaire et /0/ indique une syllabe non accentuée, mais les composés avec une accentuation double sont bien sûr données selon la notation de l'époque (voir 1.7.2.1, « Notation de la hiérarchie accentuelle dans les dictionnaires de prononciation »). Ainsi *able-bodied* est indiqué avec le schéma /1010/, et non /2010/. Le fichier des composés extraits de

---

[22] http://www.mshs.univ-poitiers.fr/Forell/PHONDICT/PHONDICT.htm

[23] Voir Wells, 1992 & 1995.

EPD12 correspond à la première version (donc la phonie n'est pas en SAMPA). Ce fichier existe également classé par graphie inverse. Il comporte 4 510 entrées.

Dans la mesure où ce fichier comporte des entrées susceptibles d'être des composés, d'après le critère graphique utilisé par Guierre, il est évident qu'il sera le point de départ de notre corpus. De plus, il contient la plupart des informations que nous souhaitons incorporer pour notre analyse. Ces entrées ont été isolées par Guierre parce qu'il voulait écarter les mots composés de son analyse de l'accentuation de l'anglais (Guierre, 1979) dans la mesure où ils semblent correspondre à des règles d'accentuation différentes du reste du lexique. Il a donc mis à notre disposition un corpus constitué des entrées qu'il a retirées de son fichier principal.

Toutefois, il faut noter que Guierre ne disposait pas à l'époque des outils que nous avons actuellement, ce qui a une incidence sur le codage notamment. Les majuscules étant réservées à la phonie (codage pré-SAMPA), il n'y a pas de majuscules sur les noms propres de la graphie. La reprise de ce fichier aujourd'hui sous-entend que nous utilisions les notations actuelles. Ainsi, afin de mettre ce fichier à jour, nous proposons de réintroduire les majuscules à partir de la version papier de EPD12[24].

Le critère que Guierre utilise pour extraire les composés est un critère purement graphique : la présence d'un trait d'union. Il est clair aujourd'hui (et peut-

---

[24] A terme, nous voudrions également transformer la phonie en code SAMPA, comme l'a fait Fournier pour Rhymes8 (qui ne comporte pas les composés de EPD12).

être déjà à l'époque de EPD12, mais dans une moindre mesure), que cette définition est restrictive. Nous essayerons donc de compléter ce fichier en apportant des données plus récentes, quelle que soit leur forme graphique.

2.2.1.3.2 *New Pronouncing Dictionary* de Guierre (NPD)

Ce dictionnaire compilé par Guierre dans les années 1990 contient 41 297 entrées. Les indications sont plus complètes que celles du fichier de EPD12, puisqu'elles donnent la graphie inverse, les variantes et le nombre de syllabes, en plus des informations déjà indiquées dans EPD12 sur la graphie, la catégorie syntaxique du composé et la phonie (en code SAMPA). Ce dictionnaire est parmi les rares sources qui indiquent la catégorie « composé » (codée {Cp}) dans la catégorie syntaxique. Il est donc utile parce qu'il fournit des données assez complètes, et aussi parce que cette indication « Cp » nous permet d'isoler les composés très facilement, et donc d'étoffer notre corpus.

2.2.1.3.3 *Longman Pronouncing Dictionary*, édition 1990 (LPD90)

La version électronique de ce dictionnaire est utilisée dans le cadre de l'équipe d'accueil FORELL qui a obtenu un droit d'utilisation sous licence exclusive. Il contient 60 705 mots[25]. L'indication d'un statut « composé » n'est pas présente En revanche, dans le fichier de travail utilisé au sein de l'équipe d'accueil, les entrées graphiques sont régulièrement reliées aux informations

---

[25] Cette version correspond aux entrées principales de la version papier du *Longman Pronouncing Dictionary*. De plus, Guierre a inclus des balises qui permettent d'isoler certains types de données, et il y a ajouté des informations (la catégorie syntaxique, le nombre de syllabes, le schéma accentuel sous forme numérique et la graphie inverse).

suivantes, dans l'ordre : catégorie syntaxique (avec parfois d'autres informations concernant la variation), phonie, variantes éventuelles et commentaires, nombre de syllabes, schéma accentuel sous forme numérique, et enfin graphie inverse. Ces informations sont séparées par des balises ({£} qui précède le nombre de syllabes, {√} devant la graphie inverse, par exemple) qui sont relativement fiables. Ce fichier sera une des principales sources pour notre corpus, compte tenu de la qualité des indications. Elles sont d'une part complètes et correspondent le mieux possible à nos exigences. D'autre part, elles sont cohérentes, ce qui permettra un traitement automatisé fiable. Dans la section suivante, nous allons étudier ces traitements.

2.2.2 Comment extraire les données ?

Dans la première partie de ce chapitre, nous avons étudié les critères qui sont habituellement associés aux composés. Ce sont ces critères que nous avons pris comme point de départ, afin de créer des sous-fichiers contenant des entrées qui y correspondent, et qui sont donc susceptibles de contenir des composés. Comme nous l'avons expliqué, il a fallu ensuite les reprendre afin de vérifier leur statut de composé, et d'écarter toutes les entrées qui ne correspondaient pas aux critères. Les résultats de ce travail nous ont permis de tester les hypothèses qui ont servi à la création des sous-fichiers, et nous récapitulons plus loin ces résultats en termes de taux de réussite (le nombre d'entrées qui ont effectivement un statut de composé, selon les critères que nous avons définis). Mais d'abord, nous allons expliquer comment nous avons procédé pour extraire les données.

2.2.2.1 MPW et l'extraction des données

En dehors du fichier de EPD12 qui ne contient *a priori* que des composés au départ, nous avons recours à des traitements automatiques qui nous permettent d'extraire les données. Nous nous servons d'un logiciel de programmation appelé MPW (Macintosh Programmer's Workshop), sous le système MacOS d'Apple™ qui permet d'opérer des tris commodément sur des fichiers. Il faut lancer une requête en respectant la syntaxe de commande qui aboutira à la création d'un sous-fichier correspondant aux eléments de la requête. Il permet également de récupérer dans un autre sous-fichier les entrées complémentaires (celles qui, dans le fichier source, ne correspondent pas à la requête). A partir de ce principe simple, nous avons à notre disposition un outil puissant de création de sous-fichiers contenant des données très ciblées.

C'est grâce à cet outil que nous avons pu engager la constitution d'une partie du corpus d'analyse : celle qui correspond aux sources écrites électroniques. Nous avons formulé des requêtes correspondant aux critères suivants : le critère syntaxique, le critère graphique, le critère phonologique / phonétique, et enfin le critère accentuel. A présent, nous allons présenter les enjeux techniques, et nous essayerons d'évaluer l'efficacité du traitement, critère par critère.

2.2.2.2 Extraction des composés selon le critère syntaxique

Il faut d'abord préciser qu'il n'est pas possible d'utiliser les manipulations syntaxiques dont nous avons parlé quand nous avons présenté ce critère. Le traitement automatique n'est possible que dans le cas où il existe un marquage

explicite du statut de composé. Nous avons vu que seul NPD permet d'isoler les composés grâce à cette information.

La requête (voir Annexe A) porte sur la suite graphique {Cp}[26] contenue dans la partie de l'entrée balisée comme catégorie syntaxique (entre {*f*} qui signale la fin de la graphie, et {[}qui marque le début de la phonie). Le fichier ainsi obtenu comporte 4 210 termes. Il est intéressant de noter qu'aucun terme ne s'écrit avec un trait d'union ; ils s'écrivent tous en un seul mot.

L'application de ce critère à NPD permet de créer un sous-fichier qui ne contient effectivement que des composés. Toutefois, l'unicité de la graphie implique que ce fichier n'est pas complet, car beaucoup de composés s'écrivent autrement. Nous isolons seulement des composés compacts. Peut-être s'agit-il de mots ayant des variantes graphiques : des mots pouvant s'écrire en un seul mot, comme dans le NPD, mais ayant une variante avec un trait d'union, par exemple, que l'on trouvera dans d'autres sources. Le fichier des composés de Guierre (EPD12) confirme qu'il y a en effet quelques entrées semblables (écrites avec un trait d'union dans EPD12), mais les entrées extraites de NPD sont souvent absentes du fichier EPD12[27]. Ces fichiers (EPD12 et NPD) élaborés tous deux par Guierre à des moments différents se complètent, mais ils sont sans doute encore

---

[26] Les séquences recherchées par une programmation MPW ou sous Access (voir chapitre 3) seront indiquées entre parenthèses{}. Le caractère espace est donc indiqué : { }.

[27] Voir 4.3.2.2 3.1, « Le contenu des tables », où nous faisons une étude comparative de la graphie des différents dictionnaires et de leur contenu.

insuffisants à eux deux. Aucun composé non compact ne figure ni dans l'un, ni dans l'autre. Par ailleurs, nous ne sommes pas en mesure d'extraire des données de LPD90 en utilisant le critère syntaxique, car aucune indication ne le permet, et ce dictionnaire est sensiblement plus grand que NPD.

2.2.2.3 Extraction des composés selon des critères graphiques

Nous avons déjà fait référence à ce critère. En effet, les mots composés peuvent se trouver sous trois formes graphiques : écrits en seul mot (les composés compacts), écrits avec un trait d'union, et écrits en deux mots séparés par un espace (les composés non compacts). Il faudrait certainement rajouter certains mots écrits avec une apostrophe, comme *cat's eye*, souvent appelés génitifs génériques lexicalisés, donc formant des classes comme les composés.

Sur le plan technique, le critère graphique est relativement simple à appliquer pour la recherche de données. Nous avons donc extrait les mots écrits avec un trait d'union dans les sources qui nous donnent les informations dont nous avons besoin (voir Annexe A). A l'aide de MPW nous avons recherché la séquence {-} dans la partie balisée pour la graphie (entre {•} indiquant le début de la ligne et {*f*} indiquant la fin de la graphie pour LPD). Nous avons lancé cette requête sur les fichiers LPD90 et NPD (à titre de vérification, puisque nous en avions déjà extrait les entrées composées, normalement).

Nous n'avons trouvé aucune entrée à trait d'union dans NPD, ce qui explique que les entrées marquées {Cp}, extraites par la première requête, s'écrivent toutes en compact. En revanche, en ce qui concerne LPD90, le sous-fichier des entrées avec un trait d'union comporte 2 365 termes. Si le trait d'union

comme seul critère est trop restrictif pour constituer un fichier complet de composés, il reste un critère valable et fiable et le fichier des entrées avec un trait d'union comportera certainement une très forte proportion de composés.

La présence d'une apostrophe peut également être la marque d'un composé. D'ailleurs, certains composés qui ne s'écrivent plus avec l'apostrophe et ne forment plus qu'un seul mot ont été historiquement générés par *'s*. Nous avons lancé une recherche sur la présence de la suite {≈'≈} (où le symbole {≈} correspond au symbole {*} généralement utilisé dans les expressions régulières, c'est-à-dire « 0 ou n caractères » dans le champ balisé comme celui de la gr hie. Pour NPD, nous trouvons 102 termes, et pour LPD 293 termes (voir Annexe A).

Le fichier des mots à apostrophes de NPD révèle la présence d'un certain nombre de lexèmes simples comme *e'er*, dont l'orthographe avec une apostrophe reflète la prononciation et indique un changement de graphie (l'apostrophe remplace une ou p' -s lettres, comme le « v » dans *e'er*). Il y a également des formes contractées dont nous avons parlé. Le fichier de LPD90 contient, en plus de ces termes, des composés. Toutefois, il est de toute évidence impossible d'extraire les composés générés par *'s* historiquement, mais où l'apostrophe ne paraît pas (*craftsman*, par exemple)[28]. L'application de ce critère aura une certaine réussite, mais qui est loin d'être totale. Nous évaluerons la réussite des différents

[28] Cet exemple est de Marchand (1969 : 65) qui fait remarquer qu'il est difficile de distinguer entre une formation en *'s* et une avec le pluriel.

critères d'extraction à la fin de cette section, en 2.2.3, « Résultats des extractions par traitement automatique ».

L'autre type de composé graphique qui semble adapté à ce genre de traitement est le composé non compact. Nous avons donc voulu isoler ces entrées à l'aide d'une requête de recherche du caractère correspondant à l'espace, { }, dans la graphie. Cette requête n'a pas abouti parce qu'il est nécessaire de préciser la présence avant et après l'espace de caractères alphabétiques, et les requêtes formulées n'étaient pas suffisamment sélectives. Cependant, à l'issue de certains tests, il apparaît clairement que la version électronique du dictionnaire dont nous disposons ne comporte que très peu d'entrées correspondant aux critères graphiques d'un composé non compact : ces entrées sont quasiment toutes des expressions étrangères telles que *post meridiem* ou *de rigueur*, qui sont associées à une phonie et qui figurent comme entrée principale dans le dictionnaire, version papier. Les très nombreuses sous-entrées de la version papier, telle que *fire engine*, ne figurent pas, sans que l'on puisse s'expliquer s'ils ont été ou non maintenus dans le fichier fourni par Longman.

Il est évident que le critère graphique ne permet pas d'isoler tous les composés, loin de là. Nous venons d'en constater les limites en ce qui concerne les composés non compacts qui sont très nombreux dans la version papier (qui donne des indications concernant leur schéma accentuel sur la graphie). Ce dernier constat réduit considérablement la portée des informations que nous pouvons obtenir automatiquement et pose la question d'une saisie manuelle pour une partie du corpus au moins. Par ailleurs, il faut ajouter que le critère graphique ne peut

pas non plus être utilisé pour extraire les composés compacts, qui n'ont pas de signe distinctif graphique pour les distinguer. Il faudra utiliser d'autres moyens pour extraire ces entrées.

Il est clair à présent que le critere graphique est une indication intéressante, mais non suffisante pour constituer un corpus de composés. Rappelons que Guierre l'a utilisé dans son fichier extrait de EPD12, où il inclut tous les termes avec un trait d'union. Pourtant, il a lui-même adopté d'autres critères de sélection dans NPD puisque les entrées marquées comme étant des composés sont écrites en un seul mot. Ce changement peut tenir à la prise en compte d'autres indications, mais il peut également refléter dans une certaine mesure la notation qui s'est imposée dans les dictionnaires récents. A l'époque de EPD12, ce dictionnaire de prononciation ne comportait que des entrées écrites en un seul mot ou bien avec un trait d'union, à quelques rares exceptions près. Le composé non compact n'y avait pas sa place. Aujourd'hui, les usages ont changé, comme le montrent les explications de LPD (éditions 1990 et 2000 : ix) :

> The pronunciation of compound words can often be derived from their component parts. Stress patterns, however, are not always easy to predict and so important compounds and their patterns are listed after the main entry.

Dans EPD12, rares sont les sous-entrées qui indiquent un composé, et ces quelques indications n'ont dans tous les cas rien de systématique. En revanche, ce système est complètement adopté par les deux dictionnaires de prononciation de l'anglais britannique, LPD et EPD dans leur version actuelle, même si ces indications ne figurent pas dans la version électronique de LPD dont nous disposons.

Les réserves que nous avons émises au sujet de l'efficacité du critère graphique nous obligent à poursuivre notre recherche sur les critères d'extraction possibles afin de constituer un corpus complet de composés. A présent, nous abordons le critère phonétique et phonologique.

2.2.2.4 Extraction des composés selon les règles phonologiques et phonétiques

Jusque-là, nous avons utilisé des critères objectivement codés dans les dictionnaires comme tels : la catégorie syntaxique « composé », puis la graphie. Il s'agit dorénavant de rechercher des critères plus subtils afin de poursuivre la constitution de notre corpus. Toutefois, nous ne cherchons pas à extraire d'autres entrées du fichier NPD, malgré le fait que ce fichier présente les qualités de balisage requis, puisque nous avons déjà extrait les composés marqués syntaxiquement, et ce fichier est plus petit que LPD (30% de mots en moins). Nous travaillons donc exclusivement à partir du LPD car c'est le seul autre fichier qui permet ce traitement, vu la régularité du balisage de. ntrées.

Nous devons regarder attentivement les caractéristiques mêmes d'un composé par rapport aux autres mots de la langue. Chaque fois que le composé constitue une exception aux règles générales, nous avons à notre disposition un critère permettant d'isoler les composés des autres mots.

2.2.2.4.1 La non-réduction de la voyelle du deuxième élément

Nous avons déjà signalé le fait que, dans le cas des composés à accentuation simple, il n'y a pas de réduction du second élément ; sa voyelle garde sa qualité

pleine (voir 2.1.1.1.4)[29]. Avant de pouvoir appliquer ce critère de façon efficace, il faut d'abord déterminer quelle en est sa portée. Il ne peut fonctionner que sur un second composant monosyllabique car, dans le cas d'éléments plurisyllabiques, il y a forcément un accent (secondaire ou tertiaire) sur une des syllabes, donc pas de réduction. Si nous étudions les dissyllabes à accentuation simple, ils représentent une catégorie d'entrées susceptible d'inclure des composés correspondant à ce critère, car, dans ce cas, nous aurons forcément une syllabe accentuée correspondant au premier élément, et une syllabe non accentuée correspondant au second. Compte tenu de ces précisions, nous sommes en mesure d'appliquer ce critère en lançant une requête sur la qualité vocalique de la syllabe non accentuée dans les entrées dissyllabiques (voir Annexe A).

2.2.2.4.1.1 Méthode d'extraction

La méthode d'extraction des données correspondant à ce critère a été la suivante. Nous isolons d'abord les dissyllabes sans trait d'union (puisque les entrées à trait d'union ont déjà été extraites), à l'aide de la recherche de la séquence {£2} (deux syllabes) ; nous trouvons 24 472 termes, soit plus de 40% du total des entrées. Puis nous recherchons les séquences accentuées /10/ ; le fichier résultant comporte 20 156 termes. Nous nous intéressons à toutes les entrées ayant une voyelle pleine dans la deuxième syllabe, mais vu sa taille, nous avons préféré scinder le fichier, voyelle par voyelle.

---

[29] Cette règle est évaluée en 4.3.2.3, « Etude de la phonie ».

Les requêtes étaient simples pour la majorité des voyelles pures où on a recherché la suite {•≈*f*≈...≈&≈Ÿ} où la séquence {•≈*f*≈} signale qu'il y a d'abord la graphie avec les marqueurs de début et fin de graphie suivis d'un nombre indéterminé de caractères, puis la séquence {...} qui indique le début de la dernière syllabe, {&} qui correspond en SAMPA à /æ/ et {Ÿ} qui signale la fin de la phonie. Cependant, trois voyelles pures posent des problèmes de sélection car leur symbole intervient également dans les diphtongues : /ɪ/, /ʊ/ et /ə/. Pour ces trois voyelles, nous avons dû procéder par étapes en créant un « ensemble /ɪ/ », par exemple, comprenant à la fois la voyelle pure /ɪ/ et les diphtongues /ɪə/ /eɪ/ et /ɔɪ/ ; ensuite, nous avons extrait une des voyelles à partir de « l'ensemble /ɪ/ » tout en créant le fichier complémentaire « ensemble /ɪ/ moins /ɪə/ » en même temps ; à partir de celui-ci (« ensemble /ɪ/ moins /ɪə/ »), nous avons extrait la voyelle suivante, en créant un fichier complémentaire et ainsi de suite. De la sorte, nous avons pu isoler les quatre voyelles (voir le détail de la programmation en Annexe A).

2.2.2.4.1.2 Remarques

Nous devons émettre une réserve au niveau de l'efficacité de cette requête : elle se limite aux dissyllabes. Il n'est pas possible de lancer cette requête sur des mots de trois syllabes (ou davantage), car nous ne disposons pas de borne[30]

[30] Plus généralement, la borne elle-même servirait de marqueur du processus de composition, permettant ainsi d'isoler tous les composés quel que soit leur schéma accentuel.

marquant la frontière morphologique (#) entre les éléments du composé, ce qui suppose que nous ne savons pas *a priori* si un des éléments est monosyllabique. Sans cette information, il n'est pas possible d'isoler d'autres composés à accentuation simple avec un deuxième élément monosyllabique. En revanche, ceux qui ont un deuxième élément plurisyllabique pourront être isolés par le critère de l'accent tertiaire (voir 2.2.2.5, « Extraction des composés selon un critère accentuel »).

Ce critère pose un deuxième problème. Nous avons écarté les entrées en /10/ ayant une voyelle réduite en deuxième syllabe ; mais si schwa (seul) représente toujours une voyelle réduite, /ɪ/ l'est fréquemment aussi (/ʊ/ est très minoritaire). Nous pouvons donc effectivement écarter le fichier des schwas qui ne correspond pas à notre critère de sélection, mais le fichier des /ɪ/ peut comporter des mots avec une voyelle réduite (donc ne correspondant pas non plus au critère) ou bien sans voyelle réduite (et dans ce cas, les entrées y correspondent) ; la phonie ne permet pas de différencier les deux cas. Ainsi, ce sous-fichier aura sûrement une forte proportion de non-composés, car il y a beaucoup d'entrées dont la seconde syllabe est effectivement réduite.

En effet, nous devons tenir compte de la mise en garde de Guierre (1979 : 564) :

> Rappelons que, selon notre hypothèse de travail, la corrélation entre [-Acc] et [+red] est orientée dans un seul sens : la réduction vocalique présuppose la non-accentuation ou la désaccentuation, mais la non-accentuation ou la désaccentuation n'entraînent pas nécessairement la réduction.

Cela suppose donc qu'un nombre peut-être important de ces dissyllabes ne sont pas des composés, non seulement dans le sous-fichier des /ɪ/, mais aussi dans tous les autres sous-fichiers. La non-réduction est une condition pour être un composé, mais ce critère à lui seul ne suffit pas pour qu'un mot soit considéré comme tel.

Notre critère d'extraction repose sur la règle spécifique concernant les composés selon laquelle ils ne doivent pas subir de réduction sur la syllabe non accentuée. Le nombre de composés ainsi trouvé servira d'indication de la validité du critère d'extraction qui est utilisé. Ces résultats figurent dans le récapitulatif à la fin de cette section (en 2.2.3), mais cela ne suffit pas pour évaluer complètement ce critère, car il faudra aussi tenir compte des composés de plus de deux syllabes qui pourraient correspondre à cette règle : c'est-à-dire ceux dont le deuxième élément est monosyllabique.

2.2.2.4.2 Les séquences phonétiques indiquant une frontière morphologique

Dans cette section, nous explorons les possibilités que nous offre un recours à la phonie, et plus précisément aux contraintes phonotactiques. Les propos de Guierre (1979 : 263) nous inspirent un certain optimisme, malgré la petite réserve :

> En bref, il semble que les contraintes phonotactiques et graphotactiques permettent à elles seules la reconnaissance des "composés", (shamrock) ; mais quand ces contraintes ne jouent pas, la forme et le sens des composés doivent être pris en considération.

Dans l'exemple, *shamrock*, la séquence phonétique /mr/ ainsi que la séquence graphique «mr» sont illicites : nous pourrions en conclure qu'il s'agit d'un

composé[31]. Cette manière de « reconnaître » les composés correspond justement à notre démarche : nous utilisons d'abord des critères qui leur sont associés pour extraire des entrées susceptibles d'être des composés, et ensuite, nous vérifions le statut de composé de ces entrées en faisant appel à des critères sémantiques, entre autres.

Les contraintes graphotactiques ont déjà été utilisées. Nous avons fait référence aux contraintes phonotactiques dans la discussion du critère phonétique / phonologique de détermination d'un composé (voir 2.1.1.1.4), et nous venons d'utiliser le phénomène de la non-réduction, mais il y a aussi des séquences de sons autorisées et des séquences illicites qui peuvent servir de marqueur d'une borne morphologique, en anglais comme dans d'autres langues. Ces considérations sont très importantes dans une théorie de la structure syllabique où elles servent pour déterminer la coupe syllabique. Dans un composé, la coupe se fait toujours à la frontière morphologique (de l'élément, #). C'est ce que fait Wells dans la première édition du LPD (1990 : xx) : « A syllable boundary is found wherever there is a word boundary, and also coincides with the morphological boundary between elements in a compound » et dans la deuxième (2000 : xix) : « LPD assumes that there is a syllable boundary wherever there is a boundary between the elements of a compound ». Nous constatons à nouveau l'intérêt de marquer les composés de cette manière. Cette frontière se trouve souvent au milieu d'une

[31] Toutefois, dans cet exemple, malgré ces indications, le fait qu'il n'y ait pas de lien sémantique entre les éléments et le mot nous fait penser que ce n'est pas un composé.

séquence de sons qui n'est pas autorisée dans le reste du lexique ; ce constat sert donc parfois pour distinguer un composé et il pourrait ainsi nous être utile pour extraire un certain nombre de termes afin de compléter le corpus.

Plusieurs questions se posent. La suite utilisée comme un critère doit être suffisamment longue (et/ou rare) pour que l'extraction soit pertinente. Il n'est pas utile d'opérer un tri qui génère un fichier comportant la moitié du dictionnaire. Il faut donc être sélectif et certainement y associer d'autres critères. Nous ne nous intéressons qu'aux mots de trois syllabes ou davantage, puisque le critère de la non-réduction a permis de traiter les entrées dissyllabiques, sauf exception. Il est vrai qu'une extraction a déjà été effectuée sur les entrées ayant certains critères graphiques, mais ces entrées peuvent facilement être écartées, comme nous l'avons fait pour les dissyllabes.

Cependant, nous émettons des réserves en ce qui concerne une extraction faite à partir de ce type de requête. Tous les composés dont le premier élément se termine par une voyelle seront forcément laissés de côté (*keyring*, par exemple) parce qu'ils ne comportent pas de séquence indiquant une frontière morphologique. De même, bien d'autres séquences se trouvent à l'intérieur d'un mot monomorphématique : /kl/ dans *forklift*, par exemple, ne se distingue pas du /kl/ de *cleft*. Même s'il était éventuellement possible d'utiliser la coupe syllabique indiquée par LPD, il nous semble que l'application de ce critère laisserait trop d'entrées de côté ; nous ne l'avons pas utilisé comme moyen d'extraction. Toutefois, il serait intéressant de le tester à partir du corpus d'analyse dans son ensemble, lorsque nous saurions que les entrées sont toutes des composés, afin

d'évaluer sa portée. Ce traitement pourrait également servir pour déterminer la borne # qui nous offrirait de cette manière des moyens de traitement de ces données qui ne sont pas actuellement disponibles (recherches morpho-syntaxiques en isolant les deux éléments, notamment).

2.2.2.5 Extraction des composés selon un critère accentuel

Lorsque nous avons discuté de ce critère dans la première partie de ce chapitre, nous avons constaté que l'accentuation du premier élément est souvent citée comme étant la marque d'un composé. Dans ce cas, il s'agit de ce que nous appelons l'accentuation simple. Les critères phonologiques/ phonétiques ont permis d'extraire ce type de composés en ce qui concerne les dissyllabes de nos sources, mais qu'en est-il des autres entrées? Nous allons d'abord utiliser une spécificité de la notation de LPD afin d'extraire d'autres composés à accentuation simple.

2.2.2.5.1 Les composés à accentuation simple

2.2.2.5.1.1 La notation de l'accent /3/

Revenons sur la manière de noter l'accent qui a été longuement discutée dans le chapitre 1 (en 1.7). LPD90 prévoit la notation d'un accent tertiaire (représenté par /3/ sous forme numérique) et qui est utilisée dans trois cas de figure.

Nous le retrouvons tout d'abord dans des mots relativement longs où il s'intercale entre l'accent /2/ et l'accent /1/. Duchiron nous énonce comme une règle (Duchiron, 2000 : 5) : « Le préfixe *contra-* implique l'ajout d'un troisième accent puisqu'il y aura quatre syllabes inaccentuées avant l'accent primaire ». Elle donne

comme exemple le mot *contraindication* dont le schéma accentuel est /203010/. En fait, il n'y aurait que trois syllabes inaccentuées avant l'accent primaire mais nous retenons le principe que dans les dérivés formés à partir d'un préfixe (mono ou plurisyllabique) et d'un mot avec un accent secondaire, le dérivé portera un accent secondaire sur le préfixe, et l'accent secondaire de la base se transformera en accent tertiaire (educational, /20100/, co-educational, /230100/).

Nous le retrouvons également dans des mots avec le suffixe *-ism*. Ainsi, LPD90 donne la prononciation suivante pour *pacifism* : /ˈpæsɪˌfɪzəm/, qui a le schéma accentuel /1030/. Dans ces mots, l'accent /3/ se trouve après l'accent primaire.

Enfin, nous le trouvons dans les composés dont le premier terme porte l'accent /1/ (accentuation simple, appelée *early stress* par Wells) où le deuxième élément est plurisyllabique (donc doit comporter un accent). L'introduction est, là-dessus, explicite (Wells, 1990 : xviii) : « In an early stressed compound, the mark appears in the second element only when it consists of more than one syllable ».

Afin d'extraire les entrées susceptibles d'être des composés, il suffit donc de lancer une requête pour rechercher la présence d'un accent /3/ après l'accent /1/ dans le champ de la notation du schéma accentuel numérique (donc la séquence {£≈.≈1≈3≈}, où {£} désigne le champ du nombre de syllabes et {.} le début du schéma accentuel). Ensuite, nous devons écarter les mots en *-ism* puisqu'ils correspondent aussi au critère de la recherche (accent /3/ après accent /1/) ; ils sont

au nombre de 376. Il reste 1281 entrées susceptibles d'être des composés (voir Annexe A).

Cette requête nous a permis d'extraire (sauf cas exceptionnel) tous les composés à accentuation simple avec un deuxième élément plurisyllabique du LPD. Nous nous attendons à une forte proportion de composés dans ce sous-fichier, d'après la définition de la notation. Cependant, il est clair que nous n'avons pas encore réussi à extraire tous les composés à accentuation simple.

2.2.2.5.1.2 Autres requêtes

Dans un premier temps, récapitulons les moyens que nous avons utilisés pour extraire ce type de composé. En dehors du marquage syntaxique, qui n'est possible que sur NPD et donc de portée insuffisante, nous avons utilisé des critères graphiques, phonétique/phonologique et la notation de l'accent /3/. Nous n'avons pas pu extraire les composés non compacts de LPD (absents de la version électronique), ni certains composés compacts (écrits en un seul mot) : ceux de plus de deux syllabes, dont le deuxième élément est monosyllabique. Nous ne pouvons pas extraire des entrées qui sont absentes de la version électronique ; par contre, les composés compacts sont bien présents dans LPD et nous devons trouver un moyen de les extraire.

Que savons-nous du schéma accentuel d'un tel composé ? Puisqu'il s'agit de composés à accentuation simple, cela implique que le premier élément est accentué alors que le second ne l'est pas. Nous ne savons pas *a priori* sur quelle syllabe tombera l accent primaire, mais en revanche nous savons que la dernière syllabe aura la valeur /0/, puisqu'elle correspond au deuxième élément inaccentué

monosyllabique du composé. Nous sommes en mesure de lancer une requête pour rechercher la séquence {≈1≈0}, ce qui signifie que nous extrayons les entrées correspondant à un schéma accentuel d'un nombre indéterminé de syllabes ayant une dernière syllabe inaccentuée. La requête est limitée aux entrées de trois syllabes ou plus, n'ayant pas de trait d'union, pour ne pas extraire des entrées déjà prises (voir Annexe A). En raison de la taille du sous-fichier, nous avons l'avons scindé selon le nombre de syllabes (trois, quatre, et ainsi de suite).

La réserve évidente que nous pouvons formuler à l'égard de cette requête est justement le fait qu'elle va générer un sous-fichier d'une taille très importante (en effet, il y a 17255 entrées). Pourtant, c'est le seul moyen d'extraire les entrées manquantes. Nous constatons les limites du traitement automatique des données, puisque la seconde étape de vérification du statut de composé des entrées sera très longue, mais le récapitulatif des résultats à la fin de cette section nous permettra d'évaluer à quel point il était nécessaire de procéder ainsi.

2.2.2.5.2 Les composés à accentuation double

Il ne reste plus qu'à extraire les composés à accentuation double qui nous manquent. Jusque-là, en dehors du marquage {Cp} de NPD, seuls les critères graphiques ont permis d'extraire ces composés. En termes de notation, ils comportent un accent secondaire sur le premier élément, et un accent primaire sur le second, quel que soit le nombre de syllabes des éléments. Nous pouvons donc rechercher la séquence {≈1≈}. Le symbole {≈} à la fin indique que la valeur de la dernière syllabe est inconnue, contrairement au cas précédent où il s'agissait d'un second élément monosyllabique inaccentué. Ici, nous ne savons pas *a priori*

le nombre de syllabes ni du premier ni du second élément. Comme précédemment, nous avons écarté les entrées avec un trait d'union, et scindé le sous-fichier selon le nombre de syllabes.

Nous émettons la même réserve en ce qui concerne les limites du traitement automatique parce que la taille du sous-fichier ainsi généré est de 9387 entrées. Nous pouvons prévoir qu'il y aura une forte proportion d'entrées portant un accent secondaire rythmique (nous avions distingué la valeur de cet accent par rapport à celui qui est porté par un composé dans le premier chapitre) ; cette hypothèse se vérifiera surtout pour celles de quatre syllabes ou plus, en raison des contraintes accentuelles de certains suffixes. Malgré la longueur de la deuxième étape de vérification du statut composé ou non de ces entrées, il est encore nécessaire afin de s'assurer que le corpus d'analyse ne comporte pas d'omissions.

Nous remarquons au passage que la notation en vigueur à l'époque de EPD12 aurait dû nous permettre d'extraire tous les composés à accentuation double sans exception puisque ce type de composés portaient deux accents primaires, un sur chaque élément. Malheureusement, la recherche de ce schéma dans Rhymes8 (la version électronique de EPD12) n'aboutit qu'à un sous-fichier où il y a principalement des préfixés. Les composés de ce dictionnaire qui comportent tous un trait d'union avaient déjà été écartés de cette version.

2.2.2.6 Extraction des composés selon une recherche sur les éléments

Dans cette section, nous terminons notre discussion sur les moyens utilisables pour l'extraction des composés en opérant des traitements automatiques à partir de sources électroniques. Jusqu'alors, nous avons utilisé des critères que

l'on associe aux composés afin de les extraire du fichier principal, mais peut-être pourrions-nous utiliser les éléments de composition eux-mêmes.

Nous n'avons pas encore de marquage # des éléments de composition; cependant, ces éléments sont connus. D'ailleurs, en faisant des recherches sur le trait d'union, nous avions isolé des quasi-lexèmes, préfixes et suffixes, présentés en tant qu'éléments (sans radical ou autre morphème de rattachement, comme *hyper-*), dont certains peuvent être considérés comme le premier ou le deuxième élément dans un composé (comme *one-* ou *half-*). LPD90 présente quelque 410 suffixes et 412 préfixes (ou quasi-lexèmes) de cette manière.

Nous pouvons envisager un traitement, élément par élément, pour rechercher les occurrences. Par exemple, nous recherchons toutes les entrées qui commencent par la séquence graphique *fire*, ce qui permettra de rassembler les mots composés avec *fire* pour premier élément. En prenant cet élément, nous aboutissons effectivement à une liste de 36 entrées (dont *fire* tout seul) ne comportant que des composés écrits en un seul mot ou avec trait d'union.

Toutefois, nous formulons deux réserves. D'abord, cette recherche sera probablement fastidieuse (ce qui n'est pas l'objectif d'un traitement automatique) dans la mesure où chaque élément sera pris séparément. Nous pouvons envisager la saisie des éléments comme des variables qui permettraient de lancer le traitement de manière itérative sur les éléments les uns après les autres, mais cela suppose que la liste des éléments soit exhaustive pour que le traitement soit fiable.

La deuxième réserve a déjà été émise au sujet des contraintes phonotactiques. Il faut, pour que le traitement automatique donne des résultats pertinents, que l'élément présente des caractéristiques suffisamment distinctives pour ne pas générer une extraction comportant beaucoup d'entrées non composées, ce qui peut être appelé « bruit ». A titre d'exemple, la recherche du mot *age* qui sert d'élément de composition générera également tous les suffixés et les terminaisons en *-age*, si la requête n'est pas associée à d'autres critères (phoniques, notamment). Il est clair que plus l'élément en question est long et différent des suites de caractères qui composent le lexique en général, plus la recherche sera fructueuse pour les composés véritables.

2.2.2.7 Les exceptions

Les sous-fichiers qui ont été créés selon les critères que nous avons étudiés recouvrent apparemment l'ensemble des composés. Toutefois, il reste quelques cas exceptionnels qui échappent à ces critères. D'abord, nous avons constaté que certains éléments plurisyllabiques qui sont placés en deuxième position ne sont pas marqués avec l'accent /3/. C'est le cas de certaines entrées en *-berry*, et aussi des entrées dont le premier élément est *back-*. S'agit-il d'une erreur du fichier ? ou d'une volonté délibérée ? Dans le cas de *-berry*, son statut possible comme suffixe a été soulevé. En fait, ces entrées ont été remarquées lors de l'étape de vérification

du statut des entrées du sous-fichier des entrées avec le schéma /*1*0/[32], donc elles sont incluses dans le corpus d'analyse.

En revanche, d'autres entrées généralement considérées comme des composés à accentuation simple n'ont pas pu être extraites. Il y a quelques rares exceptions à la règle de la non-réduction d'un élément de composé inaccentué. Certains éléments monosyllabiques subissent une réduction (les mots en *-man*, par exemple). Si l'entrée elle-même est dissyllabique, elle a été écartée des traitements automatiques d'extraction, puisque nous n'avons pas retenu les entrées avec schwa en deuxième syllabe. Le débat est ouvert en ce qui concerne leur statut de composé. Nous analyserons ces cas à partir de la comparaison avec les entrées de NPD et EPD en 4.3.2.3.

2.2.3 Résultats des extractions par traitement automatique

Nous allons rendre compte des résultats de l'étape de vérification des sous-fichiers de LPD90 qui ont été extraits selon les critères associés aux composés. Un récapitulatif succinct est donné dans le tableau suivant.

---

[32] Nous employons le symbole {*} qui indique « 0 ou n caractères », même si la programmation MPW impose le caractère {≈}. Voir Annexe A.

Figure 2.1 : Récapitulatif des résultats par critère d'extraction

Critères	Nb total d'entrées	*Nb d'entrées exclues*[33]	Nb de composés	Composés en % du nb total d'entrées	Composés en % du nb d'entrées prises en considération
Graphiques					
Trait d'union	2365	*203*	2085	88.16%	96.44%
Apostrophe	259	*207*	8	3.09%	15.38%
Phonologiques					
Non-réduction des 2e voyelles des schémas /10/	9387	*3257*	2836	30.12%	46.05%
Accentuels					
ACCENT /3/	1034	*164*	724	70.02%	83.22%
Schéma /*1*0/	17255		837	4.85%	
SCHEMA /*2*1*/	9387		665	7.07%	
NOMBRE DE COMPOSES EXTRAITS	39735		7155	18.01%	

Nous précisons que les chiffres de la colonne 2 (le nombre total des entrées) correspondent au nombre d'entrées extraites du fichier principal en appliquant un traitement automatique, chaque requête étant conçue à partir d'un critère bien précis, comme nous l'avons expliqué dans la section précédente. Les sous-fichiers sont classés par rapport à ces critères dans le tableau ci-dessus: critères graphiques, phonologiques/ phonétiques et accentuels. Nous avons vérifié toutes ces entrées, et nous avons retenu celles qui correspondent aux critères que nous avons définis dans la première partie de ce chapitre (2.1.1.1, « Les critères de définition d'un composé » Ainsi, en plus des composés communs, le chiffre de la colonne 4 comprend certains noms propres, désignant des noms de marques (tel

[33] A noter que les entrées exclues correspondent à celles qui sont écartées de la cible du corpus, non les lexèmes simples (voir Annexe B du chapitre 2).

BluTack, cité en 2.1.4.3.3), ainsi que des noms de rues et d'événements. Des précisions sur la proportion de composés noms propres retenus pour le corpus d'analyse par rapport aux composés communs sont données en Annexe B. Le tableau en annexe donne également le détail, sous-fichier par sous-fichier, des différentes catégories,[34] ce qui permet de montrer un aperçu du lexique dans son ensemble, et de voir comment il est décomposé (noms propres et autres entrées, composés et lexèmes simples).

Dans la colonne 3 du tableau ci-dessus, nous précisons également le nombre d'entrées qui sont exclues de notre champ d'étude. Nous l'avons expliqué dans la section 2.1.4, « Quelle sera l'étendue du corpus d'analyse ? » Nous écartons donc la plupart des noms propres (noms de personnes et noms de lieux), les acronymes, les formes contractées, ainsi que les composés étrangers et les composés historiques devenus opaques.

Grâce à cette distinction entre le nombre total d'entrées du sous-fichier, toutes catégories confondues, et le nombre d'entrées susceptibles d'être soumis à l'analyse (c'est-à-dire, à l'exception des catégories citées ci-dessus comme les noms propres, amalgames ...), nous obtenons des résultats qui ont été calculés selon deux méthodes différentes. Le premier pourcentage (colonne 5) représente le

---

[34] Les catégories proposées sont d'abord celles qui sont incluses dans le corpus d'analyse, à savoir : les composés communs et les composés noms propres ; puis, les catégories exclues du corpus : les noms propres (avec deux sous-catégories : ceux qui sont sans doute des composés, comme *Northwood*, et ceux qui forment des lexèmes simples comme *Wantage*) ; ensuite, nous avons regroupé les acronymes, formes tronquées, amalgames, composés historiques et composés étrangers ; et enfin, les lexèmes simples.

taux de composés par rapport à l'ensemble des données; le second (colonne 6) représente le taux de composés par rapport au nombre d'entrées susceptibles d'être incluses dans le corpus d'analyse. C'est le deuxième chiffre qui est le plus pertinent et qui sera commenté dans la section suivante, mais il est intéressant de noter l'écart parfois important entre les résultats selon le mode de calcul. Nous précisons toutefois que nous n'avons pas fait cette distinction pour les deux derniers sous-fichiers parce qu'ils ont été créés afin de compléter le corpus des composés et de nous permettre d'inclure les entrées qui n'avaient pas été ciblées par les critères de recherche précédents.

2.2.3.1 Conclusions

Pour conclure cette section, nous allons évaluer le rendement des critères d'extraction qui ont été utilisés afin de constituer le corpus d'analyse à partir des dictionnaires électroniques. Nous commenterons les résultats du tableau d'abord en ce qui concerne le critère graphique, puis le critère phonologique / phonétique, et enfin le critère accentuel.

2.2.3.1.1 Le critère graphique

C'est le critère graphique de la présence d'un trait d'union qui a généré les pourcentages de « rendement » les plus élevés, ce que nous avions d'ailleurs prévu. Les quelques entrées qui ne sont pas des composés sont d'origine étrangère, comme *hula-hoop* (Hawaïen) ou *chin-chin* (un toast chinois!). Plus de 96% d'entrées avec un trait d'union sont donc des composés, mais nous rappelons la réserve en ce qui concerne ce critère. Si la présence d'un trait d'union indique de façon très sûre qu'il s'agit d'un composé, ce critère n'est orienté que dans un sens.

C'est-à-dire que l'absence de trait d'union n'est pas une indication d'un statut de non-composé, ce qui est évident lorsque nous reprenons les résultats du tableau. Il y a environ deux mille composés à trait d'union sur un total de plus de sept mille, sans compter les composés non compacts qui ne figurent pas dans la version électronique de LPD90. Cela représente moins de 30% du total. En tenant compte des composés non compacts, les entrées à trait d'union de LPD90 ne représentent qu'environ 15%.

Quant à l'apostrophe, nous avions prévu une efficacité faible dans la mesure où ce signe est employé dans plusieurs contextes qui ne sont pas forcément en rapport avec la composition. En effet, l'apostrophe peut représenter l'omission d'une lettre comme dans la forme abrégée de *hallo* : *'allo,* où elle reflète la prononciation. Elle peut aussi correspondre à une lettre omise dans une variante (phonique et graphique) d'un mot comme *ever*, (*e'er*). Elle indique également le génitif qui peut parfois être considéré comme entrant dans un composé, mais il faut préciser qu'il y a plusieurs cas de génitifs tronqués, sans deuxième élément, ce qui ne nous permet pas de les considérer ainsi : *Levi's*, par exemple, qui est la forme tronquée de *Levi's jeans*. Si nous écartons les noms propres à ne pas inclure, et surtout les formes contractées (il y en a près de 100), nous n'arrivons qu'à un taux de 15% environ. Toutefois, notre critère de sélection a laissé de côté les entrées où l'apostrophe est indiquée comme une variante graphique (donc, après la graphie principale).

2.2.3.1.2 Le critère phonologique / phonétique

Nous avons extrait des sous-fichiers comportant des entrées dissyllabiques en /10/ n'ayant pas de voyelle réduite dans la syllabe inaccentuée. Dans le tableau qui suit, nous avons repris les mêmes rubriques que dans le tableau précédent. Par ailleurs, nous avons classé les sous-fichiers selon le type de voyelle, c'est-à-dire, en rassemblant les diphtongues, puis les voyelles tendues, et enfin les voyelles brèves. Nous avons classé la voyelle /ɪ/ séparément, car son fonctionnement pose des problèmes spécifiques.

Figure 2.2 : Récapitulatif des résultats : sous-fichiers des dissyllabes

Voyelle (par type)	Nb total d'entrées	*Nb d'entrées exclues (lexèmes simples)*	Nb composés	Composés en % du nb total d'entrées	**Composés en % du nb d'entrées prises en considération**
/eɪ/	700	*229*	287	*41,00%*	**55,62%**
/əu/	963	*385*	204	*21,18%*	**34,63%**
/aɪ/	560	*216*	238	*42,50%*	**52,42%**
/au/	238	*17*	168	*70,59%*	**90,81%**
/ɔɪ/	139	*78*	36	*25,90%*	**31,58%**
/ɪə/	33	*7*	7	*21,21%*	**50,00%**
/eə/	66	*10*	39	*59,09%*	**79,59%**
/uə/	17	*6*	0	*0,00%*	**0,00%**
Total diphtongues	2716	*948*	979	*36,05%*	**50,80%**
/i:/	447	*156*	138	*30,87%*	**50,36%**
/ɑ:/	364	*77*	149	*40,93%*	**65,93%**
/ɔ:/	464	*54*	207	*44,61%*	**79,31%**
/u:/	272	*105*	112	*41,18%*	**51,61%**
/ɜ:/	281	*37*	136	*48,40%*	**78,61%**
Total voyelles longues	1828	*409*	742	*40,59%*	**64,52%**
/e/	465	*128*	163	*35,05%*	**56,01%**
/æ/	567	*118*	280	*49,38%*	**70,35%**
/ɒ/	621	*162*	220	*35,43%*	**57,59%**
/ʊ/	194	*40*	67	*34,54%*	**62,62%**
/ʌ/	180	*35*	122	*67,78%*	**77,71%**
Total voyelles brèves (sauf /ɪ/)	2027	*483*	852	*42,03%*	**63,82%**
/ɪ/	2844	*1483*	263	*9,25%*	**15,06%**
TOTAL	9415	*3257*	2836	*30,12%*	**46,05%**

Ces chiffres méritent d'être commentés par rapport aux règles accentuelles que nous avons citées dans le premier chapitre. Si nous rappelons les règles des générativistes, l'accent est attribué sur les mots dissyllabiques en fonction de la

qualité vocalique de la deuxième syllabe. Si celle-ci est longue (ou diphtonguée), elle porte l'accent, sinon – dans le cas où elle est brève – c'est la première syllabe qui le portera. Nos statistiques portent sur toutes les entrées dissyllabiques ayant un schéma accentuel en /10/ et une voyelle pleine en deuxième syllabe. Selon la règle générativiste, si un mot dissyllabique comporte une voyelle longue en deuxième syllabe, il devrait avoir un schéma accentuel /01/; en suivant la même logique, les entrées en /10/ devraient comporter une voyelle brève dans la deuxième syllabe. Tout autre comportement est exceptionnel et doit être expliqué.

Or, une des caractéristiques spécifiques des composés à accentuation simple est justement la non-réduction du deuxième élément monosyllabique, quelle que soit la valeur de la voyelle. Notre fichier de dissyllabes comporte un nombre important d'entrées en /10/ ayant une voyelle tendue en deuxième syllabe; cependant, la qualité de la voyelle aurait dû attirer l'accent. Pourquoi ne porte-t-elle pas l'accent tout en étant longue? Ce comportement exceptionnel où l'accent reste sur la première syllabe peut peut-être s'expliquer parce que ces entrées sont des composés.

Le tableau dément cette hypothèse. Nous constatons que les sous-fichiers contenant les voyelles dites pures (voyelles longues et voyelles brèves) en dehors de /ɛ/ ont des pourcentages de composés élevés, certes, mais assez similaires, avoisinant les 64%, soit près des deux tiers des entrées prises en considération. Il n'y a apparemment aucune corrélation entre la longueur de la voyelle et la proportion de composés qui est semblable quelle que soit la valeur de la voyelle de la deuxième syllabe.

Quant aux diphtongues, nous remarquons que le taux de composés par rapport à l'ensemble des fichiers est moins élevé ; il arrive néanmoins à plus de la moitié des entrées prises en considération, avec des écarts sensibles selon les voyelles. Trois d'entre elles présentent des taux faibles : /ʊə/, /ɔɪ/ et /əʊ/. La première de ces voyelles ne comporte aucun composé, mais elle ne représente que 17 entrées (sur plus de neuf mille), et elle n'est pas du tout représentative (pas plus d'ailleurs que /eə/ qui a un taux de presque 80% avec seulement 66 entrées). Le cas de /ɔɪ/ est plus intéressant, puisqu'il y a quand même 139 entrées et un taux de moins du tiers avec un statut de composé. Mais c'est la voyelle /əʊ/ qui représente le cas le plus atypique puisqu'elle représente un taux de 32% avec 963 entrées, de loin le nombre d'entrées le plus important en dehors de la voyelle /ɪ/.

Ce chiffre est intéressant lorsque nous considérons à nouveau les règles qui nous permettent d'attribuer l'accent sur les dissyllabes. Roach fait de cette voyelle un cas spécial dans les règles qu'il a élaborées (Roach, 1991). Si les génératives considèrent que la première syllabe portera l'accent si la deuxième comporte une voyelle brève, Roach ajoute à cette condition la voyelle /əʊ/ qui fait figure d'exception donc, puisqu'elle est longue. Nous allons tenter de proposer une explication, en nous appuyant sur l'hypothèse que défend Guierre. Cette hypothèse repose sur un raisonnement qui fonctionne dans le sens inverse des règles déjà citées, à savoir que c'est l'accent qui détermine la qualité vocalique et non le contraire. Ainsi, puisque la première syllabe est accentuée, la voyelle de la seconde sera réduite (donc, brève par définition). Peut-on considérer que le

nombre élevé d'entrées ayant la voyelle /əʊ/ dans la deuxième syllabe inaccentuée peut s'expliquer par le fait que cette voyelle ne se réduit pas? Cette hypothèse répond à la fois à son comportement exceptionnel au vu du nombre d'entrées et explique également le faible taux de composés comparativement aux autres voyelles.

Cette hypothèse se vérifie également lorsque nous étudions les dissyllabes en /ɪ/. En effet, nous avons déjà expliqué que /ɪ/ a une double valeur : celle d'une voyelle pleine et celle d'une voyelle réduite ; toutefois, nous n'avons écarté que schwa des fichiers de dissyllabes[35]. Le fichier des /ɪ/ comportera forcément un grand nombre d'entrées dont la voyelle inaccentuée est réduite ; par conséquent, ces entrées ne correspondent pas à notre critère d'extraction, mais il est impossible de les écarter de manière automatique. Ce fait explique pourquoi, sur les 9415 dissyllabes étudiés, près de trois mille contiennent cette voyelle en deuxième syllabe. Nous pourrions faire d'autres calculs à partir des entrées avec et sans réduction vocalique mais cette entreprise serait périlleuse parce qu'il n'est pas aisé de distinguer entre les valeurs pleines et réduites de la voyelle. L'orthographe peut servir d'indication parfois, comme dans *hatchet*, où le « e » graphique est prononcé /ɪ/ et où la réduction est évidente, mais elle ne l'est pas pour *rabbit*, parce nous

[35] La très forte proportion de schwas dans la syllabe inaccentuée des dissyllabes en /ɪə/ (près de dix mille entrées sur vingt mille, à savoir une entrée sur deux) semble donner raison à Guierre. La corrélation entre l'accent et la réduction vocalique est évidente.

avons un « i » graphique. Nous nous limitons à une présentation séparée de cette voyelle, et à une analyse des statistiques qui tient compte de cette mise en garde.

Il reste à faire remarquer que, parmi les dissyllabes, un nombre non négligeable de lexèmes simples sont ou d'origine étrangère (*bouchée*, *kudu*) ou correspondent à des terminaisons et suffixes tels que *-oïd*, *-ine* ou *-bok*.

Nous rappelons que les résultats des sous-fichiers de dissyllabes, comme pour ceux qui ont été générés selon un critère accentuel ne comprennent pas les entrées dont la graphie comporte un trait d'union. Nous avons choisi d'écarter ces entrées pour éviter la redondance, mais cela peut fausser certaines statistiques. Dans tous les cas, il faudra réviser la proportion de composés à la hausse.

2.2.3.1.3 Le critère accentuel

2.2.3.1.3.1 L'accent /3/

Tout d'abord, d'après les résultats du tableau, le sous-fichier correspondant à la recherche de l'accent /3/ semble beaucoup moins efficace que prévu, même si nous arrivons à un taux de 83%. Nous rappelons que nous en avions déjà écarté les mots en *-ism* ainsi que les mots dérivés longs ayant un accent /3/ qui s'intercale entre l'accent /2/ et l'accent /1/. Cependant, si l'on considère les entrées trisyllabiques, nous arrivons à un taux très important (94% pour 735 entrées, ce qui représente environ les deux tiers du sous-fichier). Le taux beaucoup plus bas (environ 60% pour 305 entrées) pour les entrées de quatre syllabes ou plus s'explique facilement par la présence d'un nombre important de quasi-lexèmes. Il est effectivement intéressant de constater que la notation utilisée plus particulièrement pour les composés s'applique aussi souvent aux entrées formées

à partir de quasi-lexèmes, même si nous les avons exclus de la présente étude pour les raisons expliquées plus haut (2.1.4.2).

2.2.3.1.3.2 Les autres sous-fichiers

Contrairement aux autres sous-fichiers, ils ont été générés afin d'éviter toute omission et ne correspondent pas à un critère bien défini qui est susceptible de créer un sous-fichier comportant un nombre important de composés. Nous ne disposions pas de requête très affinée, ce qui explique pourquoi ils sont de taille aussi importante. Nous n'avons pas fait de marquage selon la catégorie de l'entrée (nom propre, lexème simple, acronyme ...) comme nous l'avons fait pour les autres sous-fichiers; nous avons simplement marqué les entrées correspondant à nos critères de définition d'un composé à inclure dans le corpus d'analyse.

En ce qui concerne le fichier des entrées susceptibles d'être des composés à accentuation simple ({*1*0}), la vaste majorité des entrées ayant ce statut ont trois syllabes, et quasiment toutes sont en /100/. Au-delà de trois syllabes, nous rencontrons surtout des lexèmes simples, (ou dérivés) et quelques quasi-lexèmes. Quant au sous-fichier des entrées susceptibles d'être des composés à accentuation double ({*2*1*}), nous avons beaucoup hésité sur le statut de composé d'un certain nombre de quasi-lexèmes. Comme prévu, ce sous-fichier contient beaucoup de dérivés, de quasi-lexèmes, et donc la présence de l'accent rythmique secondaire est une obligation (surtout dans les entrées de quatre syllabes ou plus).

Ces deux sous-fichiers nous ont obligée à poser des questions concernant le statut de certains quasi-lexèmes, et ils nous ont permis d'inclure dans le corpus d'analyse quelque 1502 entrées avec un statut de composé. Il était donc nécessaire

de procéder à ces vérifications, puisque nous complétons le corpus en l'augmentant de façon très sensible (près d'un tiers d'entrées en plus).

2.2.4 Saisie des données complémentaires : LPD version papier

Nous avons détaillé les différents moyens utilisés pour extraire les composés des fichiers électroniques, surtout depuis notre source principale, la version électronique de LPD90, et nous avons commenté les résultats. Pourtant, nous avons fait remarquer que cette source présente une insuffisance au niveau informationnel ; en effet, il manque dans cette version la quasi-totalité des composés non compacts. Ces entrées ne peuvent pas être exclues de l'analyse et la cohérence du corpus dépend également de la consistance des sources. De ce fait, il est nécessaire de procéder à une saisie manuelle de ces entrées.

Dans la version papier de LPD90, des composés (ainsi qu'un certain nombre de syntagmes) écrits avec des mots séparés figurent non pas dans l'entrée mais dans la sous-entrée ; ils comportent des indications accentuelles sur leur graphie, mais pas de phonie. L'entrée principale (« head word » en anglais) correspond toujours au premier élément du composé (ou au premier mot du syntagme). En ce qui concerne les syntagmes, ces indications servent surtout pour exprimer ou des expressions toutes faites : *apple-pie order*, par exemple (mais est-ce un composé ?) ou bien des cas de rétraction de l'accent (*stress shift*) : *all-day meeting* où le schéma accentuel de *all-day* /21/ en isolement devient /20/ devant *meeting* /10/ qui prend l'accent primaire. Wells (1990: xi) explique :

> LPD also includes a good selection of compounds and phrases, showing their stress patterns. Some of these illustrate the effect of the highly productive principle of stress shift which affects many longer English words.

Comme il s'agit d'une saisie manuelle, nous avons d'abord décidé quelles étaient les informations que nous pouvions – et que nous devions – inclure. Un traitement manuel permet d'inclure les éléments que nous considérons utiles; nous ne sommes plus tributaire d'un mauvais balayage des données, ou d'une incohérence dans la notation. Le marquage de certains éléments manquants dans le fichier électronique a été inclus pour ces composés. La graphie de l'entrée est donc saisie, non comme un ensemble, mais élément par élément, puis nous précisons le nombre de syllabes, le schéma accentuel (en y ajoutant la borne # pour indiquer la frontière morphologique), les éventuelles variantes sous forme de schéma accentuel, puis nous avons inclus un champ commentaire afin de mettre à profit cette étape fastidieuse qu'est la saisie manuelle. En effet, le concepteur du corpus remarque au fur et à mesure certains rapprochements et régularités (ou irrégularités) en réalisant ce travail. Nous n'incluons pas la phonie, ni la graphie inverse, pour des raisons pratiques: ce sont des éléments longs à saisir manuellement. Nous reviendrons à ces considérations dans la deuxième partie du chapitre 3 (3.2.1) quand nous parlerons de la structure des informations du corpus d'analyse dans la base de données qui permettra de les exploiter.

Ainsi, nous avons étoffé notre corpus d'analyse en y ajoutant plus de 6 000 entrées. Les noms propres qui dépassent le cadre de notre analyse ont quand même fait l'objet d'une saisie, mais non les entrées d'origine étrangère. Ces entrées sont mises de côté. Le corpus d'analyse comprend quelques 14 000 entrées provenant de LPD90, environ 4 500 de FPD12, et 4 200 de NPD.

Nous passons à présent à l'autre partie du corpus : les sources orales.

2.3 Sélection des données : sources orales

2.3.1 Objectifs

Nous avons voulu analyser non seulement des données provenant de sources écrites mais aussi de sources orales, et ce pour plusieurs raisons. Tout d'abord, les entrées des sources écrites font état de l'accent lexical, c'est-à-dire l'accent de mot sans contexte, lorsqu'il est prononcé isolément. Même si nous avons également accès à des variantes possibles, il est souhaitable de tester les prévisions des sources écrites, lexicographiques, dans un contexte oral, et donc de discours. Nous pourrons mieux évaluer l'influence de certains contextes, notamment intonatifs, en faisant une analyse des composés de cette manière. Ainsi, le premier objectif de cette partie du corpus est-il de mettre à l'épreuve les résultats provenant des sources écrites en étudiant le comportement d'unités lexicales dans tel ou tel contexte.

Par ailleurs, nos sources écrites ne contiennent par définition que des entrées lexicalisées, puisqu'il s'agit de dictionnaires de prononciation. Si nous nous limitons aux seules entrées lexicalisées, nous passerons à côté d'un certain nombre d'occurrences qui peuvent être considérées comme des néologismes[36] en voie de lexicalisation ou bien des unités lexicales ressenties comme formant des ensembles que l'on peut comparer à des composés. Cette remarque est particulièrement importante dans un domaine tel que la composition, domaine de

---

[36] Le terme « néologisme » est employé ici pour désigner toute création lexicale non répertoriée dans un dictionnaire.

créativité linguistique par excellence. Le deuxième objectif de la partie orale du corpus est donc d'inclure des données qui sont exclues des dictionnaires lexico-phonétiques. Cette démarche devrait nous permettre de faire des prévisions en ce qui concerne l'accentuation de nouvelles unités lexicales.

Enfin, les sources orales nous permettent de comparer le comportement de composés par rapport à des séquences de structure apparemment similaire mais qui ne sont pas des composés, surtout des séquences de type adjectif + nom telles que *dark cupboard* que nous pourrons confronter aux différentes occurrences de *dark* lorsque le mot entre dans une logique de composition (*dark room*, *dark magic*).

Ainsi, les sources orales servent à compléter le corpus, en y apportant des informations complémentaires concernant les entrées existantes et en y ajoutant de nouvelles entrées. Nous allons à présent expliquer comment nous avons fait le choix de ces sources.

2.3.2 Choix des sources

Toute source orale comportera des exemples d'unités lexicalisées qui pourront servir pour tester les hypothèses élaborées à partir de l'analyse des données des dictionnaires lexico-phonétiques que nous avons utilisés pour constituer la première partie du corpus. Nous nous sommes donc intéressée aux sources orales qui seraient susceptibles de comporter des unités lexicales qui ne figurent pas dans ces dictionnaires.

Les médias fournissent une source souvent riche de ce point de vue surtout lorsqu'il s'agit de l'actualité politique, sociale et technologique, c'est-à-dire de domaines en évolution perpétuelle où l'on va fabriquer des unités lexicales parce que l'idée exprimee est nouvelle, ou bien parce que nous pouvons exprimer le concept de façon plus concise en utilisant le processus de la composition. Dans ces cas, le concept en question n'est pas suffisamment précis, ou bien est trop nouveau, pour avoir été incorporé dans un dictionnaire. En fait, cette démarche est similaire à celle qu'empruntent les lexicologues lorsqu'ils font un travail sur corpus, comme Bauer & Renouf (Bauer & Renouf, 2001). Ils font appel à des écrits, souvent des périodiques, pour constituer un corpus où ils étudient la formation des unités composées, très souvent non lexicalisées, qui s'y trouvent. Si le type de source est similaire, la méthode diffère de la nôtre parce que nous nous limitons à un travail à partir d'enregistrements, et nous nous intéressons donc exclusivement à la radio et à la télévision, et non à la presse écrite.

A titre d'exemple, nous trouvons dans le secteur technologique des mots correspondant à des objets nouveaux, comme cette invention russe : *talking vodka bottle top*. Au cours du reportage, le terme réapparaît sous plusieurs formes : *vodka bottle top*, *talking bottle top*, qu'il faudra comparer. Dans un autre secteur, nous trouvons le terme *distressed debt*, que l'on peut considérer comme un terme économique au même titre que *bad debt* (une dette qui n'a aucune chance d'être remboursée) ou *national debt* (l'argent emprunté par un gouvernement et non encore remboursé). Ces deux derniers exemples figurent dans Chambers 1988 (dans la sous-entrée de *debt* et de *bad / national*) mais nous ne trouvons que

ˌnational ˈdebt (indiqué avec ce schéma dans la sous-entrée de *national*) dans LPD90.

L'inclusion ou non dans le dictionnaire de ce genre de terme est un sujet lexicologique très vaste que nous ne pouvons pas développer dans cette étude. La place dans un dictionnaire est forcément limitée ; l'éditeur doit faire des choix. Il n'en reste pas moins que la langue nous permet de former de telles unités, et que le locuteur natif n'a aucun mal à les comprendre et à les prononcer. Il est intéressant de noter que, pour l'exemple de *distressed debt*, ce terme est repris au cours du reportage et débouche sur un jeu de mots avec *distressed country*, qui réutilise l'adjectif en tant qu'adjectif qualificatif et non comme un élément de composition. Nous ne manquerons pas de faire des comparaisons de ce genre.

Si les médias couvrent un large éventail de secteurs où l'on rencontre beaucoup d'unités formées pour l'occasion, nous relevons peu de véritables néologismes susceptibles d'être réutilisés et éventuellement lexicalisés par la suite. C'est l'œuvre littéraire qui peut nous fournir le plus d'exemples de ce type de phénomène puisque les mots inventés entrent dans une logique narrative et seront repris au cours de l'histoire. Ainsi, nous avons choisi une œuvre qui est riche sur le plan de la création lexicale, et qui a fait l'objet d'un enregistrement. Il s'agit donc d'une source à la fois écrite et orale qui nous permettra d'étoffer le corpus en y incluant des informations complémentaires pour certaines entrées (celles correspondant aux composés lexicalisés déjà relevés pour le corpus) ainsi que de nouvelles entrées. Par ailleurs, comme cette source est d'abord écrite, elle nous aidera à remonter le mécanisme de la prédiction de l'accent par le locuteur

natif, puisqu'il rencontre le mot d'abord à l'écrit, et ensuite le prononce alors qu'il ne l'a jamais rencontré auparavant.

Nous avons retenu deux types de sources orales très différents pour constituer la deuxième partie du corpus pour les raisons que nous venons d'expliquer. Nous nous référons à la source documentaire pour la partie concernant les médias, et à la source littéraire pour l'autre. Nous allons à présent expliquer comment nous avons sélectionné les enregistrements, puis nous commenterons les spécificités et les limites de chaque type de sources. Nous commençons notre étude par la source documentaire.

2.3.2.1 Source documentaire

Les médias fournissent un grand éventail d'émissions sur l'actualité politique, sociale et technologique. Nous avons fait notre choix en fonction du contenu des émissions et parfois aussi pour des raisons techniques. Nous avons retenu une émission de télévision de BBC World, *Reporters*, ainsi qu'un ensemble de reportages, essentiellement de la radio, que nous appelons *documentaires divers*. Le fait d'avoir des reportages provenant d'émissions différentes permet de comparer des sources, des locuteurs, et des styles. Nous allons présenter les spécificités de ces deux types de sources documentaires.

2.3.2.1.1 Reporters

Chaque émission dure 30 minutes et propose de courts documentaires sur l'actualité, allant de trois à dix minutes environ. Il y a un certain nombre de locuteurs différents, puisqu'il y a plusieurs journalistes et souvent des interviews qui font intervenir encore d'autres locuteurs. Le fait d'avoir des locuteurs variés

présente des avantages puisqu'il permet de comparer le discours de personnes différentes qui parlent du même sujet, à l'intérieur d'un reportage mais aussi entre des reportages différents. L'intérêt de cette démarche est incontestable, même si, d'un point de vue technique, il n'est pas toujours aisé de faire cette comparaison.

Cependant, nous devons émettre deux réserves en ce qui concerne certains des intervenants. Si nous étudions une source orale, c'est aussi pour tenter de prendre la langue sur le vif, dans un contexte réel ; mais comme il s'agit d'une émission de BBC World où les reportages se font à travers le monde entier, il y a souvent des locuteurs non-natifs. Ils ne sont donc pas représentatifs et nous avons décidé de ne pas tenir compte de leur discours, sauf exception, dans la mesure où leur prononciation ne correspond pas forcément à l'anglais d'un locuteur natif. Nous remarquons également la mauvaise qualité du son de ces interviews, sans doute justement parce que le locuteur n'utilise pas une intonation et une accentuation anglaise (sans parler de la prononciation des sons à proprement parler).

La deuxième réserve concerne le discours peu naturel de certains journalistes qui articulent très explicitement, donnant la nette impression de lire un texte. Peut-être cela s'explique-t-il encore par le fait qu'il s'agit d'une émission de BBC World - chaîne télévisée dont les auditeurs ne sont pas tous anglophones – et que les journalistes articulent donc d'une manière un peu artificielle afin de se faire comprendre très clairement. Nous constatons notamment des mots prononcés avec une voyelle pleine, là où l'on s'attendrait à une réduction vocalique, tels que *a* dans *a vulture fund*. Toutefois, cette remarque ne concerne que peu de

journalistes, mais il faudra en tenir compte dans l'analyse. Il est vrai que la question du discours à partir d'un script lu se pose pour toutes les émissions de télévision ou de radio avec des journalistes[37], à l'exception peut-être des *talk shows* où tout le monde parle à la fois. Ces derniers sont les plus naturels, mais il est difficile de les analyser justement parce qu'il y beaucoup d'interruptions et de bruit. Par ailleurs, ce type d'émission ne se prête pas nécessairement à une étude de composés.

Quels que soient les problèmes que posent les locuteurs, cette émission garde une cohérence d'ensemble dans la mesure où il existe un noyau relativement restreint de locuteurs différents (les journalistes de BBC World). Quant aux reportages, ils témoignent d'une volonté d'informer l'auditeur sur l'actualité dans le monde dans toute sa diversité. La variété des sujets traités dans l'émission *Reporters* est un point fort et nous avons pu retenir un nombre important d'unités non lexicalisées à inclure dans le corpus.

2.3.2.1.2 *Documentaires divers*

Si *Reporters* nous a permis d'étoffer le corpus de façon significative, nous devons tenir compte des réserves que nous avons émises à son sujet. Ainsi, est-il souhaitable de compléter la source documentaire par d'autres types d'émissions.

---

[37] Voir aussi les remarques d'Huart sur le discours radiophonique (Huart, 1984 : 95) : « Le discours radiophonique n'est presque jamais « naturel », soit parce qu'il n'y a pas d'interlocuteur et que les textes sont lus, soit dans le cas d'entretiens, parce que les deux locuteurs ne font pas que se parler, étant conscients du public qui va écouter et juger. La situation de discours étant très particulière, avec la contrainte d'être audible et « compréhensible » pour un grand nombre de personnes, il s'ensuit que l'aspect « musical » de la chaîne subit parfois des déformations ».

Nous avons eu accès à un ensemble de reportages provenant de diverses sources, et surtout de la radio, tous étant d'une durée d'environ trois minutes. Comme pour *Reporters*, nous remarquons le caractère très varié des reportages, tant du point de vue du contenu que des locuteurs. Contrairement à BBC World, il y a peu ou pas de locuteurs non-natifs et le style des journalistes qui mènent les interviews est souvent rapide, toujours fluide et naturel. La cohérence interne de cet ensemble tient au fait que les reportages visent un public anglophone; il n'y a pas de concession ni d'effort particulier pour se faire comprendre. Ce sont des émissions authentiques qui correspondent aux objectifs des responsables de cette sélection qui a été créée pour le concours de l'agrégation (session 2000).

Par ailleurs, la diversité des sujets nous a également permis d'étoffer le corpus, mais les émissions ne se prêtent pas toutes à une étude des composés. Quelques interviews sur l'actualité littéraire ainsi que des reportages sur des élections sont peu porteurs. Le point fort de cet ensemble est incontestablement l'authenticité de la langue.

2.3.2.2 Source littéraire

Le choix de l'œuvre littéraire s'est fait sur deux critères essentiels : d'abord la richesse sur le plan de la création lexicale, et ensuite l'existence d'une version enregistrée par un professionnel. C'est ainsi que nous avons décidé de travailler à partir du livre, *Harry Potter and the Philosopher's Stone*, de J. K. Rowling, et de l'enregistrement lu par Stephen Fry (édité par Cover to Cover).

Ce livre est le premier de sept volumes (tous ne sont pas encore sortis) qui racontent l'histoire d'un jeune garçon qui découvre qu'il est sorcier. Le premier

volume (comme les suivants) se passe principalement dans son école, Hogwarts, qui est une école pour sorciers. L'histoire est ancrée dans un monde magique qui est à l'écart de notre monde, mais non en rupture avec lui. Ceci explique la très grande richesse au niveau de l'innovation lexicale, car l'auteur a dû inventer un grand nombre de termes pour désigner les concepts courants de son monde magique (noms de potions, de créatures, de vêtements, etc.). Bien entendu, beaucoup de ces termes sont des composés.

A l'inverse des sources documentaires dont l'une des forces est la diversité des locuteurs, il n'y a qu'un seul narrateur qui raconte l'histoire dans son intégralité. C'est à la fois un avantage et un inconvénient. Bien entendu, il n'est pas possible d'étudier la variation entre des locuteurs différents, même si le narrateur joue le rôle de tous les personnages et emprunte des voix diverses. Toutefois, cette homogénéité permet de comparer plus facilement les mêmes unités dans des contextes différents parce que le seul facteur qui change est le contexte ; nous n'avons pas besoin de tenir compte d'autres facteurs qui pourraient interférer.

Lorsque nous avons étudié notre choix de sources documentaires, nous avons fait référence au problème que pose le discours lu, qui risquerait de perdre le naturel souhaité. Cet enregistrement est au départ un discours lu ; cependant, Stephen Fry, en narrateur expérimenté, veille à ce que la lecture soit complètement fluide. Comme pour tout acteur qui apprend un texte, les paroles semblent tout à fait naturelles lors de la représentation. Néanmoins, il est certain que le débit est celui d'une narration, donc plus fluide et moins entrecoupé qu'un

vrai discours où il y aurait des hésitations, des reprises, peut-être du bruit (quelqu'un qui tousse par exemple). Une articulation parfois un peu explicite peut ralentir le débit et influer sur la manière dont l'accentuation de certains composés est perçue. Il faudra en tenir compte dans l'analyse.

Malgré ces quelques réserves, le dernier point qui justifie le choix d'une œuvre littéraire est qu'il permet de tester certaines prévisions concernant l'accent. Le texte est d'abord écrit et nous avons commencé par prédire l'accentuation de chacun des composés inventés. Ensuite nous confrontons les prévisions du locuteur natif et l'enregistrement ; les cas de désaccord feront l'objet d'une analyse particulière. Les autres cas devraient nous aider à remonter le mécanisme de la prédiction de l'accent.

Nous allons à présent nous tourner vers la sélection des extraits à inclure dans notre corpus d'analyse. D'abord nous faisons état des contraintes techniques que nous devons prendre en compte, puis nous expliquons nos critères de sélection en commentant les résultats de ce travail.

2.3.3 La sélection des unités

2.3.3.1 Informations techniques

En ce qui concerne la source littéraire, elle a la particularité d'exister d'abord à l'écrit. Le travail s'est donc fait en deux temps. D'abord, nous avons utilisé un concordancier sur la version numérisée du texte de l'œuvre, ce qui nous a permis de rassembler toutes les unités que nous voulions extraire, de les saisir en contexte, et ainsi de prédire l'accent. C'est le fichier écrit d'origine. Ensuite, nous avons travaillé à partir de l'enregistrement, comme pour les autres sources orales.

La sélection des extraits sonores se fait à l'aide d'un logiciel d'édition de son qui permet de sélectionner des séquences qui sont présentées sous la forme oscillographique de la même façon que l'on sélectionne du texte. Ensuite, la séquence doit être copiée et mise dans un autre fichier. Ce fichier est enregistré dans un répertoire (ou sous-répertoire) afin de rassembler les extraits sonores d'une manière logique (voir les commentaires sur cette organisation en Annexe D).

Nous avons utilisé deux logiciels différents selon que nous travaillions sous Mac OS (et là nous utilisions Peak©) ou sous Windows (où nous avons utilisé Audacity[38]). Il a fallu d'abord convertir le support son sous un format lisible par ces programmes, à savoir un format numérisé. En ce qui concerne la source littéraire *Harry Potter*, puisque l'enregistrement était sur CD, la numérisation s'est faite dans de bonnes conditions à partir d'une source constante en utilisant un taux d'échantillonnage de 44100 Hz en stéréo (à 16 bits), et les fichiers numérisés présentaient toujours une représentation graphique où l'oscillogramme était parfaitement clair. Le travail des ingénieurs du son garantit qu'il n'y a aucun écrêtement ; d'ailleurs, l'on sent qu'ils ont dû intervenir parfois pour s'en assurer.

La source *documentaire divers* était également sur CD, et il existait déjà une version numérisée. Toutefois, au départ, nous avons eu quelques difficultés de compatibilité entre les systèmes Mac OS et Windows puisque la version

---

[38] Ce logiciel est disponible en freeware à partir du web sur le site : http://audacity.sourceforge.net.

numérisée était sous Mac OS et ne pouvait pas être lue sous Windows. Une fois le problème résolu, la qualité de la représentation oscillographique était satisfaisante.

Enfin, la source documentaire *Reporters* a posé beaucoup de problèmes au niveau de la conversion parce que l'enregistrement avait été fait depuis une source numérique (la chaîne satellite TPS) sur cassette audio. La numérisation s'est faite en utilisant un taux d'échantillonnage de 22050 Hz (à 16 bits), mais la qualité du son n'était pas constante et nous avons dû moduler le volume selon les reportages, et même à l'intérieur d'un reportage, afin d'obtenir une qualité graphique optimale. Certains enregistrements ont été laissés de côté parce que la qualité n'était pas satisfaisante.

Ces extraits ont été sélectionnés pour être soumis ensuite à une analyse approfondie à l'aide d'un logiciel approprié. Nous utilisons Praat[39] pour effectuer ce travail et les contraintes de ce logiciel nous ont amenée à prendre certaines décisions. Comme ce logiciel traite les deux formats .wav et .aiff (respectivement les formats de fichiers son sous Windows et Mac OS), les extraits sont enregistrés indifféremment dans l'un ou l'autre de ces formats. Par contre, ils ont été sauvegardés en mono et non en stéréo car Praat ne fonctionne que dans ce format. Par ailleurs, cela présente l'avantage d'occuper moins de place. Cependant, le taux d'échantillonnage a été laissé à 44100 Hz pour l'extrait original afin de permettre

---

[39] Praat : http://www.fon.hum.uva.nl.praat. Le concepteur est : Paul Boersma, Institute of Phonetic Sciences, University of Amsterdam, Herengracht 338, 1016CG Amsterdam, The Netherlands. http://www.fon.hum.uva.nl./paul/

une transmission optimale des données, comme le recommandent les concepteurs, même si ensuite ce taux est réduit à 22050 Hz par le logiciel[40].

2.3.3.2 Méthode de sélection

Nous avons procédé en deux temps afin de constituer un corpus oral cohérent, susceptible de fournir des résultats pertinents lorsque les extraits sonores sont soumis à une analyse acoustique. La première étape est celle d'une présélection des extraits sonores correspondant à des composés lexicalisés et à des néologismes.

2.3.3.2.1 Présélection des extraits sonores

Notre objectif était de sélectionner des composés lexicalisés ainsi que des unités non lexicalisées. Au premier abord, nous expliquons notre méthode en ce qui concerne les composés lexicalisés, en commençant par la sélection à partir des sources documentaires, ce travail étant différent selon le type de la source orale.

2.3.3.2.1.1 Composés lexicalisés

2.3.3.2.1.1.1 Critères de sélection : sources documentaires

Les sources documentaires comportent un certain nombre de composés lexicalisés. Nous les avons extraits systématiquement des reportages retenus dans

---

[40] Lorsque l'on numérise du son, on passe d'un signal continu à un signal discret ; la qualité de la numérisation dépend du taux d'échantillonnage ainsi que du nombre de bits choisi pc - stocker ces informations, « quantization » (44100 Hz sera d'une qualité supérieure à 22050 Hz, mais prendra aussi davantage de place ; de même, échantillonner à 32 bits donnera un résultat plus précis, mais occupera deux fois plus de place qu'à 16 bits. Voir Johnson (1997 : 26-32) qui explique l'importance de ces choix. Les opérateurs téléphoniques utilisent une bande allant de 300 Hz à 4000 Hz, ce qui permet de transmettre un signal intelligible. Cette bande filtre le signal acoustique de sorte que seules les fréquences comprises entre ces valeurs sont transmises. Toutefois, pour une

les deux types de sources documentaires. Lorsque le même composé revient plusieurs fois dans un même reportage, nous avons sélectionné chaque occurrence en notant les différences éventuelles de l'oscillogramme, des impressions auditives perçues et du contexte intonatif. Toutefois, si le composé revient plus de trois fois, sans différences notables, nous ne l'avons plus sélectionné, mais comme les reportages ne durent souvent que quelques minutes, il est rare que ce soit le cas.

Ainsi, nous avons rassemblé un certain nombre de composés lexicalisés, ce qui nous permettra d'étudier leur comportement en discours, et non isolément, dans des contextes variés. Nous serons donc en mesure de tester les prévisions concernant l'accent lexical que nous fournissent les sources écrites (voir 4.4).

2.3.3.2.1.1.2 Critères de sélection : source littéraire

Notre méthode est tout à fait différente en ce qui concerne la source littéraire de *Harry Potter*. En effet, sur les 505 minutes d'enregistrement, il y a un très grand nombre de composés lexicalisés, mais nous rappelons que notre objectif principal en retenant cette source est d'étudier le processus de création ; nous nous sommes donc limitée à la sélection de quelques composés lexicalisés tests qui nous semblent avoir un intérêt particulier.

Très souvent, ces composés sont intéressants parce que leur formation ressemble beaucoup à celle des composés inventés ; ainsi, nous avons sélectionné

---

bonne qualité d'enregistrement de la parole, d'après Johnson, il est préférable d'utiliser une bande allant jusqu'à 10000 Hz ; au-delà de ce seuil, le signal ne sera pas significatif.

racing bike pour le comparer à *racing broom*, et *library book* pour le comparer à *spell book*. De même, nous avons retenu les composés lexicalisés *chessboard* et *chessmen* qui seront à analyser par rapport à *wizard chess*, car le même élément entre dans la composition, mais pas dans la même position.

Parfois, nous avons extrait des composés ayant des prononciations variables tels que *icecream* ou *armchair*. Est-ce que la prononciation est sujette à variation par un même locuteur, selon le contexte? Ces extraits nous permettront de répondre à cette question. D'autres extraits sont retenus pour nous permettre de comparer des paires de composés avec des schémas accentuels différents: *Christmas Day* et *Christmas present*, par exemple. Nous avons également retenu quelques composés qui sont souvent répétés dans l'enregistrement comme *fireplace*, *birthday* et *broomstick*.

2.3.3.2.1.2 Composés non lexicalisés

En ce qui concerne les sources écrites, nous nous sommes déjà posé la question de savoir ce qu'est un composé, et notre discussion a été centrée sur la distinction que l'on peut faire entre un composé et un dérivé; nous avons également débattu du cas des quasi-lexèmes. Enfin, nous avons limité notre corpus en prenant la décision d'en exclure la plupart des noms propres, les acronymes, les formes tronquées, les amalgames et les composés étrangers ainsi que les composés historiques devenus opaques (voir 2.1.4). Bien entendu, toutes les entrées sont lexicalisées car nos sources écrites sont des dictionnaires lexico-phonétiques.

2.3.3.2.1.2.1 Critères de sélection : les unités

Pour les sources orales, notre sélection tient compte de ces restrictions, mais la question essentielle se pose différemment. Quels sont les critères que nous devons adopter afin de sélectionner les unités non lexicalisées à retenir pour l'analyse, c'est-à-dire les unités qui ne figurent pas dans les sources écrites ?

Nous avons déjà insisté sur la fonction créative de la composition. Cette créativité peut aboutir à deux types d'unités non lexicalisées : des néologismes, dont on trouve beaucoup d'exemples dans la source littéraire ; et des unités longues comportant plusieurs noms ou adjectifs pour désigner avec un maximum de précision et un minimum de mots une idée connue ou bien que l'on pourrait concevoir. Ces unités ne sont pas des néologismes au même titre que ceux de la source littéraire, mais plutôt des formations où les mots se cumulent afin de désigner tel ou tel (nouveau) concept. On en trouve un grand nombre dans les sources documentaires.

Le terme « composé » convient pour beaucoup de néologismes de la source littéraire qui sont souvent formées à partir de la même construction que des unités lexicalisées (*portrait hole / keyhole*, par exemple), mais il ne peut pas s'appliquer aux unités longues des sources documentaires. Nous devons nous poser la question de savoir comment les appeler.

Certaines unités sont formées à partir d'un composé + élément supplémentaire, comme *grandfather clock*, par exemple ; dans ces cas, Paillard (2000 : 51) leur donne le nom de « surcomposés ». Malgré l'intérêt de cette distinction, elle est trop restrictive au niveau d'une analyse du discours, car elle

exclut les nombreuses unités qui ne sont pas formées à partir d'un composé lexicalisé, mais qui désignent néanmoins un concept unitaire[41]. Cependant, l'appellation de Paillard tient compte de la récursivité de ce processus. Quant à Bauer & Renouf (2001 : 111), ils distinguent les composés de leur corpus (définis provisoirement par la présence d'un trait d'union) des unités lexicales complexes (sans trait d'union) ; ils les appellent « multiword units » ou « MWUs ». Ce terme rappelle la formation complexe de ces unités, mais l'aspect récursif de cette formation nous semble aussi important. Ainsi, nous proposons une nomenclature qui tient compte de ces deux caractéristiques, « construction cumulative », afin de désigner les unités du discours. Nous étudierons ces unités du point de vue acoustique afin de déterminer leur comportement par rapport à celui des unités lexicales qui ont leur place dans un dictionnaire.

Notre méthode consiste donc à retenir toute unité qui fonctionne comme un ensemble et qui désigne un concept. Là où nous hésitons, notamment entre syntagme adjectival et composé, nous préférons sélectionner l'unité plutôt que l'écarter. C'est ainsi que nous avons pré-sélectionné dans nos sources documentaires des extraits tels que *cross-border organized crime* ou bien *talking vodka bottle top* que nous avons déjà cité. Il est clair que ces unités ne figureront sans doute jamais dans un dictionnaire : pourtant, nous pouvons considérer qu'elles forment des unités du discours ; elles sont parfaitement compréhensibles,

---

[41] Voir Picoche (1992 : 16) qui insiste sur la ressemblance de certaines unités aux syntagmes du point de vue de leur construction, mais elle conclut que le fait que le signifiant désigne un concept unitaire suffit pour que l'on considère l'unité comme un composé.

l'ensemble désigne un concept unitaire, et elles entrent donc dans notre problématique.

Nous trouvons d'autres types de composition dans nos sources orales : par exemple, les composés formés à partir d'un nom propre et d'un nom commun qui sont fréquents, que ce soit dans la source documentaire ou dans la source littéraire. Nous avons donc extrait des unités telles que *Mark Rich investment* de *Reporters*, ou bien *Quidditch team* dans *Harry Potter*. Cette dernière source nous fournit beaucoup de néologismes ayant un rôle dans la logique narrative de l'œuvre : certains de ces composés sont présentés comme des noms propres, comme *Invisibility Cloak* (un vêtement que l'on met pour se rendre invisible), d'autres sont des noms communs, comme *flying motorbike* (l'engin qu'utilise Hagrid pour emmener Harry chez son oncle et sa tante Dursley, et qui n'existe pas selon M. Dursley). Nous avons retenu toutes les unités de ce type.

Comme l'un de nos objectifs est de pouvoir tester le comportement des unités en discours, nous avons également été amenée à inclure dans notre sélection quelques séquences qui sont clairement des syntagmes et non des composés à des fins de comparaison. Ainsi, dans *Harry Potter*, nous retenons *dark cupboard* où *dark* joue le rôle d'adjectif qualificatif épithète afin de le comparer à des unités comme *Dark Arts* ou encore *dark wizard*. Nous expliquons de manière plus détaillée les implications de cette démarche dans la section suivante.

2.3.3.2.1.2.2 Critères de sélection : le contexte du discours

Si nous voulons être en mesure d'étudier l'accentuation des composés en discours, la sélection des extraits doit aussi tenir compte d'un certain nombre de

phénomènes du discours. Nous allons préciser les éléments de la prosodie qui en relèvent.

2.3.3.2.1.2.2.1 L'accent de phrase

En discours, l'on peut être amené à insister sur une partie de la phrase afin de lui donner un sens particulier. Certains appellent ce phénomène l'accent de phrase ; Bolinger l'appelle simplement « accent » (Bolinger, 1972). Nous en avons parlé au chapitre 1 (1.3, « Les différents types d'accent »). Il a une influence certaine sur le contour intonatif, mais il peut aussi, dans certains cas, avoir une incidence sur le schéma accentuel de l'unité sur laquelle on insiste. Nous avons retenu quelques exemples d'accent de phrase que nous avons trouvés dans *Harry Potter*, comme le composé *anyone* qui, dans un contexte précis, porte un accent double alors que nous nous attendons à un accent simple. Bien entendu, nous retenons aussi des exemples du même mot dans des contextes neutres afin de les comparer. Ce type d'accent d'insistance est aussi fréquemment rencontré dans les constructions cumulatives : *ˈspring-sown ˌoilseed ˈrape* est opposé à *ˌautumn-sown ˌoilseed ˈrape*, par exemple (voir l'analyse en 4.5.1.2.4.3, « L'insistance »).

2.3.3.2.1.2.2.2 L'intonation

Nous savons que la perception de l'accent est influencée par plusieurs facteurs, dont la hauteur mélodique. Nous avons donc retenu des extraits ayant des contours intonatifs différents afin de pouvoir évaluer l'influence de l'intonation sur l'accent. La sélection de ces entrées est systématique. Dans la source *Harry Potter*, nous avons parfois un grand nombre d'entrées identiques ; dans ce cas, nous nous sommes limitée à trois extraits ayant le même contour intonatif, tout en

notant à quoi correspondent les autres dans le fichier écrit d'origine. Il y a quelques exceptions où nous avons retenu toutes les occurrences (parfois une dizaine) parce que l'unité nous semble particulièrement importante dans la logique narrative de l'œuvre. Dans la mesure où les reportages des sources documentaires ne dépassent pas une dizaine de minutes, nous avons fort peu d'occurrences répétées au-delà de trois fois, et quand c'est le cas, nous n'avons pas jugé utile de les retenir.

2.3.3.2.1.2.2.3 Le déplacement de l'accent

Le déplacement de l'accent est un phénomène tellement répandu, dans les adjectifs composés en particulier, que les dictionnaires lexico-phonétiques utilisent un symbole spécial (le triangle ◄) afin d'indiquer que le schéma accentuel de l'entrée risque d'être modifié dans certains contextes. Nous avons donc voulu sélectionner des unités dans des contextes où l'on pourrait s'attendre à un déplacement de l'accent.

2.3.3.2.2 Présélection des unités: résultats

Le tableau suivant récapitule le travail de présélection des unités. Il précise la durée des enregistrements, et pour les sources documentaires, il donne également la durée des reportages retenus pour sélectionner les unités, ainsi que le nombre de ces reportages. Nous rappelons que pour *Reporters*, la longueur des reportages peut varier de trois à dix minutes alors que pour les *documentaires divers*, tous durent trois minutes. Toutefois, comme *Reporters* est une émission de télévision, il y a aussi une représentation visuelle et certains bruitages qui ont du sens parce qu'on voit ce à quoi ils correspondent. Ces petites interruptions n'ont

pas leur place dans une émission de radio, mais cela explique pourquoi la longueur effective de parole est relativement plus courte dans *Reporters* que pour la source *documentaires divers*.[42] Le débit de parole est peut-être légèrement plus lent aussi, mais ce facteur est variable selon les locuteurs. Nous en tiendrons compte lors de l'analyse acoustique.

Nous distinguons les composés lexicalisés (extraits sonores lexicalisés) et les unités non lexicalisées ou syntagmes (extraits sonores au total). Toutefois, certains extraits sonores sont relevés plusieurs fois, lorsque la même unité est répétée : le chiffre correspondant au nombre d'extraits sonores lexicalisés indique le nombre d'extraits différents (sans répétition dans un reportage ou dans le livre).

Figure 2.3 : Nombre d'extraits sonores présélectionnés

Source	Durée	Nombre de reportages	Nombre d'extraits sonores au total	Nombre d'extraits sonores lexicalisés
Reporters	150 minutes	25 reportages	343	123
Documentaires divers	138 minutes	46 reportages	345	85
Harry Potter	505 minutes		508	33

Si nous avons extrait un certain nombre de composés lexicalisés, surtout à partir des sources documentaires, nous notons cependant que plus des deux-tiers des extraits sonores sont des unités non lexicalisées. Une telle proportion peut paraître surprenante, mais elle s'explique par le fait qu'il s'agit très souvent d'unités complexes formées pour exprimer une idée de façon concise (que nous

[42] Voir la transcription graphique du texte des reportages en Annexe C, et leur durée.

avons appelées «constructions cumulatives»), et non de néologismes à proprement parler. La formation et le contenu de ces extraits sont extrêmement diversifiés puisque les reportages dont ils sont issus traitent des thèmes très variés, couvrant des sujets politiques, sociaux, économiques, culturels, scientifiques[43]. Le nombre d'extraits sonores pour *Harry Potter* est également très important[44]. Ces données pour ont sans doute être exploitées ultérieurement ; toutefois, dans le cadre de ce travail, le nombre et la diversité de ces extraits sonores nous obligent à restreindre ce corpus oral. Nous avons donc abordé la deuxième étape de la sélection où nous avons déterminé des critères plus affinés afin d'aboutir à un corpus d'analyse cohérent.

2.3.3.2.3 Critères de sélection pour le corpus o[illegible] d'analyse

Les sources orales avaient été choisies pour les raisons que nous avons déjà expliquées : le discours plutôt spontané des reportages provenant des sources documentaires qui présentent une grande richesse au niveau du champ lexical couvert, et au niveau de la variété des locuteurs ; par ailleurs, la grande richesse en néologismes de la source littéraire. A l'issue de l'étape de présélection, nous avions plusieurs centaines d'extraits des deux types de sources.

Cependant, la diversité de cet ensemble ne permet pas de faire une analyse approfondie ; nous avons donc décidé de limiter le nombre d'extraits sonores à

---

[43] Voir la description détaillée des sujets traités dans les sources documentaires, par reportage présélectionné, en Annexe C.

[44] Voir les commentaires sur l'organisation des extraits de la source littéraire en Annexe D.

analyser. Pour que les extraits choisis présentent une certaine homogénéité, nous avons restreint le nombre de reportages des sources documentaires qui sont effectivement analysés en sélectionnant ceux qui correspondent le mieux à nos critères d'analyse qui seront expliqués dans les sections suivantes. Nous avons pris six reportages en tout (deux de *Reporters* et quatre de la source *documentaires divers*), d'une durée totale d'une vingtaine de minutes.[45] A des fins de comparaison, nous avons aussi sélectionné quelques extraits d'autres reportages. Nous avons également été amenée à réduire le nombre d'extraits de la source littéraire en tenant compte d'une logique de comparaison entre les extraits des deux types de sources, et des critères d'analyse.[46]

Nous avons constitué le corpus oral d'analyse à partir de cinq critères qui nous semblent les mieux adaptés à nos objectifs. Ces critères tiennent compte des spécificités de chaque type de source et donc permettent de valoriser l'apport de chacune: la source littéraire fournit beaucoup de néologismes alors que les sources documentaires sont riches en constructions cumulatives. Nous avons aussi tenu compte de considérations techniques. Les cinq critères sont les suivants :

1/ Sélection des extraits où il y a un élément en commun

2/ Sélection des extraits selon leur formation

---

[45] Voir la transcription graphique du texte intégral de ces reportages en Annexe C.

[46] Le détail concernant les extraits du corpus oral d'analyse se trouve en Annexe E. Chaque extrait est associé, entre autres, à une concordance, c'est-à-dire, le contexte du discours dans lequel il apparaît. Les extraits sonores peuvent être écoutés à partir du CDROM.

3/ Sélection des extraits afin d'étudier l'influence du contexte de discours

4/ Sélection des extraits : variation et perception de l'accent

5/ Sélection des extraits en vue de la prévision de l'accent

Là où cela a été possible, nous avons sélectionné des extraits qui correspondent à plusieurs critères.

2.3.3.2.3.1 Sélection des extraits où il y a un élément en commun

Nous avons décidé de sélectionner des extraits qui présentent des éléments communs dans la mesure du possible. Ce critère repose sur des considérations techniques dont il faut tenir compte afin de s'assurer de la pertinence de nos analyses. Nous sommes motivée par un principe général connu : moins il y a de paramètres qui changent, plus les différences constatées sont pertinentes. Beaucoup de tests acoustiques en laboratoire utilisent cette démarche afin d'isoler tel ou tel paramètre acoustique de perception. Le chercheur propose la même séquence à l'auditeur, mais fait varier l'intensité (ou le F_0) afin de tester l'influence de ce paramètre dans la perception (voir, par exemple, les travaux de Sluitjer & van Heuven, 1993). De plus, en prenant des éléments communs, nous avons un point de comparaison qui nous permettra d'étudier, par exemple, la longueur relative des segments, la qualité des voyelles, qui *a priori* se ressemblent, dans des contextes de discours semblables et différents. Nous pourrons ainsi mesurer l'impact de ces contextes sur notre perception de l'accent.

Cette démarche permet au corpus oral d'analyse de préserver une cohérence d'ensemble. Cependant, les unités qui ont un élément en commun avec d'autres

sont nombreuses et notre sélection a tenu compte des autres critères pour que les mêmes extraits puissent être analysés dans des contextes différents (de formation, de discours ou encore de perception). Les extraits retenus pour le corpus oral d'analyse correspondent donc non seulement à ce critère mais aussi à au moins un autre.

2.3.3.2.3.2 Sélection des extraits selon leur formation

Un grand nombre d'unités non lexicalisées sont formées à partir d'un nom propre suivi d'un nom commun. Dans la source littéraire, les unités relatives à la vie des «maisons» sont formées de cette manière: *Gryffindor common room*, *Slytherin table*, par exemple. De même, toutes les unités correspondant à la vie des Moldus (qui est la traduction française de «Muggle» qui désigne les personnes n'ayant pas de pouvoirs magiques), telles que *Muggle sweet* ou *Muggle money*, suivent cette construction. Bien entendu, par opposition à notre monde et dans la logique de l'histoire, il existe des bonbons de sorciers et de l'argent de sorcier (en effet, on trouve *wizard money* et une panoplie de bonbons magiques tels que *Chocolate Frogs* et *Bertie Bott's Every-Flavour Sweets*). Ce type de formation se trouve dans les deux sources. Ainsi, un des reportages qui parlait d'un individu nommé Mark Rich contient beaucoup d'unités sur le schéma *Mark Rich investment*. Cependant, nous avons décidé de ne pas cibler ces formations car les questions d'analyse du discours ou de la perception ne sont pas différentes, que la formation comprenne, ou non, un nom propre.

Nous expliquerons notre démarche en ce qui concerne les unités formées à partir de 's, puis les composés réduplicatifs, la construction cumulative, et enfin les unités que l'on doit considérer comme des syntagmes plutôt que des composés.

2.3.3.2.3.2.1 's ou pas

Nous avons sélectionné des extraits où le même élément figure mais est relié à l'autre élément tantôt avec, tantôt sans, 's. C'est dans la source littéraire que nous trouvons le plus d'exemples pertinents. Nous avons retenu les unités commençant par *dragon* ou *dragon's*. Certaines de ces unités sont vraisemblablement des syntagmes (*dragon's head*), d'autres forment un concept bien précis que nous pouvons considérer comme des composés (*dragon's blood*, *dragon egg*, *dragon dealer*, *dragon breeding*). Nous nous intéresserons à leur schéma accentuel.

2.3.3.2.3.2.2 Composés réduplicatifs

Nous avons trouvé un exemple d'une formation de ce type que nous avons donc relevé à titre d'exemple: *hocus pocus*, que l'on trouve dans la source littéraire.

2.3.3.2.3.2.3 Constructions cumulatives

Les sources documentaires ont fourni un nombre conséquent d'unités non lexicalisées; elles sont souvent formées à partir d'une construction cumulative (voir 2.3.3.2.1.2.1, «Critères de sélection: les unités»); le choix des reportages à étudier a tenu compte notamment de ce critère. Certaines sont formées à partir d'éléments qui sont déjà des composés, d'autres à partir d'éléments simples. Les éléments communs correspondent aux thèmes proposés dans le documentaire en

question, et sont réemployés très souvent au cours du reportage pour former de nouvelles unités. Nous avons déjà cité *(talking) (vodka) bottle top*, et *distressed debt* ; nous pouvons en ajouter d'autres telles que : *greenhouse gas emissions*, qui est repris pour former *greenhouse gas emissions trading* ; le composé de base, *greenhouse gas*, n'apparaît pas dans l'enregistrement. Dans d'autres reportages, les éléments *oilseed* et *tobacco* servent dans ces constructions (*genetically modified oilseed rape*, et *tobacco companies*, par exemple).

Dans la source littéraire, ce type de construction est assez marginal. *Bertie Bott's Every-Flavour Beans* est un des quelques exemples ; nous trouvons également *Every-Flavour Beans* une fois que l'existence des bonbons de Bertie Bott est posée. Nous avons donc décidé d'écarter les constructions cumulatives provenant de cette source, car elles sont peu nombreuses.

2.3.3.2.3.2.4 Composés et syntagmes

La formation de certaines unités composées ressemble à celle des syntagmes adjectivaux : comment les distinguer ? Cette question se pose forcément lorsque l'on étudie les composés et c'est pourquoi nous avons voulu analyser des extraits sonores ayant ce type de formation. Nous vérifierons donc l'hypothèse (voir 2.1.1.1.5, « Le critère accentuel ») selon laquelle le schéma accentuel peut servir de critère distinctif : le composé aura un accent sur le premier élément, le syntagme aura un accent sur les deux. Cependant, nous savons que certains composés se comportent comme des syntagmes du point de vue accentuel.

Pour tester cette hypothèse, nous avons sélectionné des unités où le même élément apparaît dans la même position, comme l'adjectif *dark* qui sera comparé

dans *dark cupboard* (de toute évidence un syntagme) et *dark magic*, *Dark Arts*, et d'autres. Nous ne nous sommes pas limitée à ces formations classiques adjectif + nom, et nous nous sommes également intéressée à des formations avec 's comme *dragon's head* que nous comparons à *dragon's blood* et *dragon liver*.

2.3.3.2.3.3 Sélection des extraits afin d'étudier l'influence du contexte de discours

Le corpus oral d'analyse a pour objectif l'étude du comportement des composés dans le discours. Nous en avons donc tenu compte lorsque nous avons sélectionné les extraits sonores à étudier; nous avons préféré les extraits présentant un intérêt pour ce genre d'analyse. Toutefois, nous précisons que ces extraits sont étudiés dans un cadre comparatif. Si nous voulons étudier l'effet d'un certain contour intonatif, par exemple, cela suppose que nous soyons en mesure d'étudier aussi le contexte neutre.

Il est certain que les trois sous-rubriques ci-après sont reliées. Si l'un de ces phénomènes du discours influence la prononciation, nous constatons souvent qu'un autre est également présent. A titre d'exemple, s'il y a un accent de phrase, il est souvent associé à un déplacement d'accent.

2.3.3.2.3.3.1 L'intonation

Les extraits sonores où le contour intonatif est pertinent sont assez rares dans nos sources. Nous avons surtout des contours plats ou bien des contours descendants, et peu de montées. Nous avons donc sélectionné toutes les occurrences où nous avions un contexte neutre (contour plat ou descendant) et un contour montant, afin d'être en mesure de les comparer. Les ensembles avec *dark* et *dragon* seront étudiés de ce point de vue, car il y a un extrait sonore avec un

contour sensiblement différent. Les sources documentaires ne sont pas davantage porteuses. Nous avons deux extraits avec un contour montant pour *farm trials*, et un contexte neutre.

2.3.3.2.3.3.2 L'accent de phrase

Dans la source littéraire, il existe quelques exemples pertinents, comme le compose lexicalisé *anyone*, qui est prononcé comme s'il s'agissait d'un syntagme écrit en deux mots distincts, *any one* (nous avons aussi un contexte neutre), et une des occurrences avec *dark*. Les sources documentaires sont plus riches, lorsque le journaliste (ou la personne qu'il est en train d'interviewer) insiste sur une caractéristique plutôt qu'une autre. Nous avons donc *long-term weather* (sous-entendu par opposition à *short-term weather*, mais ce dernier n'est pas dans le reportage), et *home fuel* est opposé à *home insulation*. L'analyse acoustique nous permettra de déterminer comment le schéma accentuel est affecté par ce type de phénomène et d'évaluer les facteurs acoustiques qui contribuent à faire percevoir l'accent de phrase: l'intensité certainement, mais aussi la longueur et éventuellement le contour intonatif que nous pourrons étudier d'après la fréquence fondamentale et la pente.

2.3.3.2.3.3.3 Le déplacement de l'accent

Nous nous sommes également intéressée aux contextes du discours qui seraient propices à un déplacement de l'accent. Chaque fois que cela a été possible, nous avons sélectionné le composé dans un contexte neutre, puis le même composé dans un contexte de déplacement. Ainsi, nous avons extrait des adjectifs composés tels que *under-age* qui est un composé lexicalisé, puis la

séquence *under-age smoker* qui est une unité non lexicalisée qui nous permettra d'étudier le fonctionnement du composé lexicalisé dans ce contexte. Les unités ayant une construction cumulative sont aussi sujettes au déplacement de l'accent : *autumn-sown* dans la construction *autumn-sown oilseed rape.* Il y a quelques exemples dans la source littéraire : *self-stirring, self-stirring cauldrons.*

2.3.3.2.3.4 Sélection des extraits : la variation et la perception de l'accent

Dans cette rubrique, nous regroupons deux phénomènes apparemment très différents mais qui ont pourtant des caractéristiques communes. Nous devons prendre les mêmes précautions lors de l'analyse des paramètres acoustiques dans les cas de variation comme dans les cas où la perception de l'accent ne correspond pas aux graphiques du logiciel d'analyse acoustique.

Nous avons fait référence à plusieurs reprises à la variation qui est une richesse de la langue mais qui peut aussi perturber l'analyse des résultats. Elle suppose un traitement qui tienne compte des différences entre des voix différentes, mais aussi dans des contextes différents. La seconde rubrique qui traite de la perception attire l'attention sur des différences entre la perception et les paramètres acoustiques tels que le logiciel les présente : c'est-à-dire qu'il y a des occurrences où notre perception est en désaccord avec les faits acoustiques. Il est nécessaire d'exercer une grande vigilance lorsque nous faisons ce genre d'analyse.

2.3.3.2.3.4.1 La variation

La richesse en locuteurs des sources documentaires nous a incitée à étudier la variation telle qu'elle est présentée dans les dictionnaires lexico-phonétiques : c'est-à-dire qu'un composé est prononcé d'une certaine façon par un locuteur A.

et d'une autre façon par un locuteur B. Cette ambition s'est avérée difficilement réalisable en raison de la grande diversité des sujets. Il est vrai que la maladie de la fièvre aphteuse (*foot-and-mouth disease*) réapparaît dans plusieurs reportages, mais il nous semble qu'un travail de ce genre mérite que l'on y consacre tout un corpus, et non quelques extraits[47]. De plus, nous nous sommes intéressée *a priori* à l'anglais britannique standard, ce qui fait que nous n'avons que très peu de locuteurs ayant un accent régional, ou venant d'un autre pays anglophone.

Toutefois, nous avons constaté dans notre corpus une certaine variation interne : c'est-à-dire que le même locuteur prononce le même composé avec des schémas accentuels différents. Bien entendu, cette variation peut être la conséquence de l'influence du contexte de discours dont nous venons de discuter (l'accent de phrase, par exemple), mais ce n'est pas toujours le cas. Nous avons donc retenu des extraits où la prononciation varie selon les occurrences, et nous analyserons quels sont les facteurs qui changent et les raisons de ces changements. Nous avons des extraits des deux types de source : *armchair* de la source littéraire, et les occurrences de *vulture fund* des sources documentaires.

2.3.3.2.3.4.2 Perception et facteurs acoustiques

Lors de l'étape de la préselection des extraits sonores, nous avons constaté que l'oscillogramme laisse parfois deviner un type de schéma qui n'est pas celui que l'on perçoit. Même s'ils sont très minoritaires, si nous voulons mener à bien une analyse acoustique, il est important d'élucider les raisons pour lesquelles les

---

[47] Voir les travaux de Zumstein sur la variation : Thèse en cours, Université de Poitiers.

faits acoustiques ne correspondent pas à notre perception. Il y a des exemples de cela dans les deux types de sources: les ensembles avec *dark* et *dragon*, des composés avec *hole* et le composé réduplicatif *hocus pocus*.

2.3.3.2.3.5 Sélection des extraits en vue de la prévision de l'accent

Nous espérons grâce à ce travail être en mesure de prévoir l'accent dans les néologismes. Nous pouvons tester nos hypothèses à partir de la source littéraire où nous utilisons notre intuition d'anglophone pour prévoir l'accent sur les unités non lexicalisées, puis nous comparons avec la prononciation du narrateur. Nous essayons de remonter le mécanisme de l'attribution de l'accent. Lorsque nous rencontrons une formation de type connue avec des associations nouvelles, nous sommes capables de comprendre l'unité et de la prononcer (*portrait hole* s'accentuera /100/ par analogie avec *keyhole*, /10/).

Notre présélection des unités a tenu compte de cet objectif. Dans l'œuvre considérée, nous rencontrons fréquemment des exemples de cette capacité à créer ou à transformer le sens des mots pour désigner des idées que nous ne rencontrons que dans l'imaginaire: les noms d'ingrédients de potions magiques par exemple, *wormwood*, *beetle eyes* (qui existent peut-être dans notre monde, mais sûrement pas en tant qu'ingrédient), et un des goûts des bonbons de Bertie Bott cités plus haut, qui est décrit comme *vomit-flavoured*. C'est ainsi qu'il existe des personnes qui font du trafic de dragons, appelées *dragon dealers* (que nous comprenons aisément en comparant la formation à celle de *arms dealer*). Le même processus d'interprétation nous permet de comprendre *stoat sandwiches* et de partager le dégoût de Harry par rapport à cette invention culinaire douteuse.

Toutefois, ces composés présentent un ensemble trop disparate pour pouvoir aboutir à une analyse dont les résultats soient pertinents. Nous avons donc décidé de sélectionner pour le corpus oral d'analyse les extraits qui présentent des points de comparaison intéressants entre les composés lexicalisés et les néologismes. Ainsi, dans notre source littéraire, si nous avons extrait *keyhole*, c'est pour le comparer à *portrait hole* (un des portraits est en réalité l'entrée du dortoir), et *staffroom* et *trophy room* serviront comme point de comparaison avec *mirror room* (la pièce où se trouve le fameux miroir magique d'Erised) et *flying-key room* (pièce où se trouvent des milliers de clés volantes). Nous avons également retenu les extraits des sources documentaires contenant ces éléments, là où cela a été possible, comme *washroom*.

2.3.3.2.4 Sélection pour le corpus oral d'analyse : Résultats

Le tableau suivant récapitule le nombre d'extraits sonores effectivement retenus pour le corpus oral d'analyse. Les entrées lexicalisées sont celles indiquées par le dictionnaire, les entrées partiellement lexicalisées correspondent à des unités à construction cumulative qui comportent une unité lexicalisée.

Figure 2.4 : Nombre d'extraits sélectionnés pour le corpus oral d'analyse

Type de source	Nombre d'extraits sonores au total	Nombre d'extraits sonores lexicalisés	Nombre d'extraits sonores partiellement lexicalisés	Nombre d'extraits sonores non lexicalisés
Source documentaire	57	19	10	28
Source littéraire	56	10	0	46

Dans le chapitre 4 (4.4) nous ferons une étude comparative des prévisions des dictionnaires et les réalisations effectives des unités en discours. Toutes ces unités seront soumises à l'analyse acoustique.

2.3.4 Choix des données saisies

Pour les sources orales, la saisie des données est manuelle, comme elle l'était pour la version papier du LPD (voir 2.2.4, « Saisie des données complémentaires : LPD version papier »). Nous devons décider quelles sont les informations que nous pouvons et que nous voulons inclure dans la base de données. De prime abord, nous allons rappeler les indications fournies électroniquement par notre principale source écrite : la graphie, la catégorie grammaticale, la phonie, les variantes, le nombre de syllabes, le schéma accentuel et la graphie inverse. Les autres sources écrites fournissent certaines de ces informations, mais pas toutes, alors que la saisie manuelle de LPD a fourni en plus le découpage en éléments individuels des composés.

Par rapport à ces indications, celle qui est obligatoire, car elle permettra des rapprochements entre les sources, c'est la graphie. Nous l'avons saisie de manière différente selon le type de source orale. Pour la source littéraire, la graphie que nous trouvons dans l'œuvre écrite est associée à l'extrait sonore correspondant.

Comme cette orthographe est vérifiable, nous avons saisi les unités, élément par élément, comme nous l'avons fait pour les composés non compacts de la version papier de LPD90. Pour les sources documentaires, pour chaque extrait sonore, nous associons une graphie que nous ne pouvons pas garantir, puisqu'il s'agit d'une source exclusivement orale. Nous ne pourrions pas faire une étude de la manière dont les éléments sont combinés graphiquement, et c'est la raison pour laquelle nous avons saisi cette information sans tenter d'en montrer le découpage élément par élément.

D'autres indications seront exclues pour des raisons pratiques, comme dans la saisie des composés non compacts de la version papier de LPD90 : la phonie et la graphie inverse (même s'il est possible de la créer à partir de la graphie). Quant à la variation, puisqu'il s'agit d'extraits sonores, nous nous intéressons à la prononciation effective de l'extrait, ce qui pourra nous donner des indications intéressantes sur la variation, mais cette information ne peut pas être saisie en tant que variante puisqu'il s'agit de la prononciation (unique) de l'extrait.

Dans une étude de l'accent lexical, les indications sur le nombre de syllabes et le schéma accentuel sont précieuses ; nous les avons donc incluses pour les extraits sonores, même si ces extraits sont de longueur très variée, souvent sous une forme de construction cumulative (*bottle top*, *vodka bottle top*, *talking vodka bottle top*). Le schéma est indiqué en incluant la borne # entre les éléments, comme pour la saisie des composés non compacts de LPD, version papier. Le type de schéma perçu figure sous forme codé « s » (accentuation simple) ou « d » (accentuation double), ainsi que le type de schéma prévu (par le dictionnaire dans

le cas des unités lexicalisées, et par l'anglophone dans les cas des unités non lexicalisées de la source littéraire). Nous avons également inclus des informations concernant le type de composé, sa catégorie syntaxique, et sa formation.

Afin d'identifier l'extrait[48], nous y associons le nom du fichier son, ainsi que le répertoire où il se trouve : par exemple, « hole\keyhole1 », où le répertoire est « hole » et le fichier son s'appelle « keyhole1 ». Pour chaque source, le nom du fichier contenant l'extrait sonore est le plus proche possible de la graphie, avec un numéro indiquant s'il s'agit de la première (ou de la deuxième...) occurrence de l'extrait dans le reportage. Nous associons également la concordance (l'extrait sonore dans son contexte de discours). Nous reviendrons à la question de l'organisation des données dans le chapitre suivant lorsque nous étudierons la structure du corpus d'analyse (3.2.1).

Conclusion

Nous avons étudié de quelle manière le corpus d'analyse a été constitué à partir de sources écrites et orales, mais il reste à déterminer quelle est la meilleure façon d'exploiter ces informations. Dans le chapitre suivant, nous allons étudier quelles sont les méthodes les mieux adaptées à l'analyse de notre corpus, d'abord en ce qui concerne les extraits sonores, puis pour l'ensemble des données.

---

[48] Les extraits sonores se trouvent sur le CDROM dans le répertoire \corporal, dans le sous-répertoire indiqué en Annexe E.

ANNEXES DU CHAPITRE 2

Annexe A: Extraction des sous-fichiers : programmation 235

Annexe B : Récapitulatif détaillé des fichiers LPD90 : vérification du statut de composé 256

Annexe C : Les reportages de la source documentaire 258

Annexe D : La source littéraire 277

Annexe E : Le corpus oral d'analyse 282

Annexe A: Extraction des sous-fichiers : programmation

Nous nous sommes servi de l'éditeur de programmes MPW (Macintosh Programmer's Workshop) afin d'effectuer les extractions[1] à partir des fichiers suivants : A_Z9CD ;2 (LPD90) et NPD2c[2] (NPD), tous deux dus à Guierre. Les programmes que nous avons écrits sont présentés dans les sections qui suivent, regroupés selon les critères que nous avons utilisés pour extraire les entrées susceptibles de contenir des composés. Nous avons laissé subsister les commentaires qui récapitulent la méthode et explicitent les résultats. Ensuite, les fichiers résultants ont été étudiés, et seuls les composés faisant l'objet de notre analyse ont été retenus. Dans le chapitre 2, section 2.1.3, nous avons commenté cette étape et évalué l'efficacité des critères d'extraction. Ces programmes ont été écrits soit par nous, soit en collaboration avec J.-L. Duchet. Nous indiquons les auteurs et les dates de la programmation au début de chaque script.

Extraction selon des critères catégoriels

Une source, NPD, attribue le symbole {Cp} aux entrées des composés. Nous les avons donc extraites à partir de la programmation suivante qui consiste en une recherche de la séquence {Cp}.

---

[1] Voir le chapitre 2, section 2.1 « Sources écrites : Extraction des Données ».

[2] Voir le site http://www.mshs.univ-poitiers.fr/Forell/Guierre/index.htm.

Recherche dans NPD des entrées marquées {Cp}

```
#JLD & SM lundi 12 juillet 1999
Open 'NPD2c'
Count -l 'NPD2c'
Search -s -q /Cp/ 'NPD2c' > 'NPD2cComp'
Count -l 'NPD2cComp'
Open 'NPD2cComp'
41296
4211
##Objectif: Récupérer les entrées du fichier NPD2c qui sont composées
et notées par Cp.
#4211 entrées marquées {Cp}
```

Extraction selon des critères graphiques

Nous avions émis l'hypothèse que les entrées ayant une graphie avec un trait d'union ainsi que celles écrites avec une apostrophe étaient susceptibles d'être des composés. L'extraction des entrées avec un trait d'union a nécessité une recherche portant exclusivement sur la graphie, puisque le signe graphique du tiret est aussi utilisé dans la partie de l'entrée concernant les variantes. Cette recherche a abouti non seulement à l'extraction des entrées ayant au moins deux éléments reliés par un trait d'union mais aussi à l'extraction des entrées des préfixes, qui se terminent par un trait d'union ainsi que des entrées des suffixes qui commencent par ce signe.

Recherche dans LPD des entrées à trait d'union

```
#JLD & SM mercredi 8 juillet 1998
Open 'A_Z92cop'
Search -s -q /•[¬ ]*-/ 'A_Z92cop' > 'LPDhyphenated2'
Open 'LPDhyphenated2'
Count -l 'LPDhyphenated2'
3201
#Recherche sur séquence (//) allant du début de la ligne (•)
#Point noir de début de ligne: alt @.jusqu'à la fin de la graphie
#de l'entrée (¬ ): recherche orthographique, donc, d'une entrée
#avec trait d'union : fichier investigué, fichier de résultat
#3201 entrées ayant le signe trait d'union dans la graphie
```

```
Open 'LPDhyphenated2'
Search -s -q -f 'ComplémentSuffixes' /•-[¬ ]*/ 'LPDhyphenated2' >
'Suffixes'
Open 'Suffixes'
Open 'ComplémentSuffixes'
Count -l 'Suffixes'
Count -l 'ComplémentSuffixes'
410
2791
#410 entrées suffixes (qui commencent par un tiret)

Open 'ComplémentSuffixes'
Search -s -q -f  'TrueHyphens' /•[¬]*- / 'ComplémentSuffixes' >
'Préfixes'
Open 'Préfixes'
Count -l 'Préfixes'
Count -l 'TrueHyphens'
412
2379
#412 entrées prefixes (qui se terminent par un trait d'union)
#2379 entrées ayant des éléments reliés par un trait d'union
```

Recherche des entrées avec une apostrophe

```
#JLD & SM lundi 26 juillet 1999
#Recherche sur séquence (//) allant du début de la ligne (•)
#Point noir de début de ligne: alt @.
#jusqu'à la fin de la graphie de l'entrée (¬f): recherche
#orthographique, donc d'une entrée où il y a des apostrophes
#fichier investigué, fichier de résultat

Open 'A_Z9CD;2'
Search -s -q /•[¬f]* '/ 'A_Z9CD;2' > 'LPDapostrophe'
Open 'LPDapostrophe'
Count -l 'LPDapostrophe'
293
#293 entrées avec apostrophe dans LPD

Open 'NPD2c'
Search -s -q /•[¬f]* '/ 'NPD2c' > 'NPDapostrophe'
Open 'NPDapostrophe'
Count -l 'NPDapostrophe'
102
#102 entrées avec apostrophe dans NPD
```

Extraction selon des critères phonologiques / phonétiques

Dans cette programmation, nous avons utilisé un système de numérotation des voyelles. Le tableau suivant récapitule les correspondances entre cette numérotation et les symboles IPA.

Figure II.1 : Système de correspondance entre les voyelles de l'anglais et N° voyelle

Numéro	Symbole IPA	Numéro	Symbole IPA	Numéro	Symbole IPA	Numéro	Symbole IPA
1	/iː/	7	/ɔː/	13	/eɪ/	18	/ɪə/
2	/ɪ/	8	/ʊ/	14	/əʊ/	19	/eə/
3	/e/	9	/ɑː/	15	/aɪ/	20	/ʊə/
4	/æ/	10	/ʌ/	16	/aʊ/		
5	/ɑː/	11	/ɜː/	17	/ɔɪ/		
6	/ɒ/	12	/ə/				

Recherche des dissyllabes en /10/ ayant une voyelle pleine dans la deuxième syllabe

Méthode d'extraction 1

Avec cette méthode subsistent les entrées à trait d'union, d'où la présence de quelques doublons dans notre base de données. Les voyelles ne sont pas traitées séparément, sauf lorsque la programmation le nécessite : c'est-à-dire, le traitement des diphtongues, pour les distinguer des schwas seuls qui représentent des voyelles réduites, ainsi que de /ɪ/ dont une proportion conséquente risque d'être des voyelles réduites aussi. Cette méthode a été retenue pour les entrées en /ɪ/.

```
#JLD&SM mercredi 19 juillet 2000
#rechercher les composés dissyllabiques compacts (soudés)
#en fonction de la voyelle.
Open 'A_Z9CD;2'
Search -s -q /£2/ 'A_Z9CD;2' > 'A_Z9dissyll'
Open 'A_Z9dissyll'
Count -l 'A_Z9dissyll'
24472

Open 'A_Z9dissyll'
Search -s -q /.01/ 'A_Z9dissyll' > 'A_Z9dissyll_01'
Open 'A_Z9dissyll_01'
Count -l 'A_Z9dissyll_01'
2515
Search -s -q /.10/ 'A_Z9dissyll' > 'A_Z9dissyll_10'
Open 'A_Z9dissyll_10'
Count -l 'A_Z9dissyll_10'
20156
```

```
#Recherche des dissyllabiques en /10/ qui ont un schwa dans la
#deuxième syllabe. Récupération du fichier complémentaire
#de ceux qui n'ont PAS de schwa.
Open 'A_Z9dissyll_10'
Search -s -q -f 'A_Z9dissyll_10_nonschwa' /•≈ʃ≈…≈@≈Ÿ/ 'A_Z9dissyll_10'
> 'A_Z9dissyll_10_schwa'
Open 'A_Z9dissyll_10_schwa'
Count -l 'A_Z9dissyll_10_schwa'
6876
Open 'A_Z9dissyll_10_nonschwa'
Count -l 'A_Z9dissyll_10_nonschwa'
13280

#Recherche des sonantes syllabiques
#Récupération complémentaire des dissyllabes_10_non-schwa
#délestés des sonantes syllabiques
Open 'A_Z9dissyll_10_nonschwa'
Search -s -q -f 'A_Z9dissyll_10_nonschwa;2' /•≈ʃ≈…≈°≈Ÿ/
'A_Z9dissyll_10_nonschwa' > 'A_Z9dissyll_10_sonsyll'
Open 'A_Z9dissyll_10_nonschwa;2'
Count -l 'A_Z9dissyll_10_nonschwa;2'
11400
Count -l 'A_Z9dissyll_10_sonsyll'
1880

#Mais il faut ajouter les mots transcrits avec un schwa dans une
diphtongue repris dans le fichier 'A_Z9dissyll_10_schwa'

Open 'A_Z9dissyll_10_schwa'
Search -s -q /•≈ʃ≈…≈[IUe]@ ≈Ÿ/ 'A_Z9dissyll_10_schwa' >
'A_Z9dissyll_10_diphtcent'
Open 'A_Z9dissyll_10_diphtcent'
Count -l 'A_Z9dissyll_10_diphtcent'
123
#123 voyelles 18, 19, 20, diphtongues centralisantes, /ɪə/, /eə/, /ʊə/.

Open 'A_Z9dissyll_10_schwa'
Search -s -q /•≈ʃ≈…≈@U≈Ÿ/ 'A_Z9dissyll_10_schwa' > 'A_Z9dissyll_10_@U'
Open 'A_Z9dissyll_10_@U'
Count -l 'A_Z9dissyll_10_@U'
  988
#988 voyelle 14 /əu/

#Donc on a comme dissyllabes en /10/ ayant une voyelle non réduite en
#schwa ou en sonante syllabique, un effectif qui se monte à :
#11400 + 123 + 988=12511
#Il reste là-dedans des secondes syllabes réduites à I, des préfixés
du type abstract, des noms propres, etc.
```

```
#A partir des non-schwa (aussi sans les sonantes syll)
#on recherche les 2e syllabes en I et le fichier complémentaire
#donc celui des mots dont la 2e syllabe n'a ni schwa ni I

Open 'A_Z9dissyll_10_nonschwa;2'
Search -s -q -f 'A_Z9dissyll_10_non_I' /•≈ƒ≈…≈I≈Ÿ/
'A_Z9dissyll_10_nonschwa;2' > 'A_Z9dissyll_10_nonschwa_I'
Open 'A_Z9dissyll_10_nonschwa_I'
Count -l 'A_Z9dissyll_10_nonschwa_I'
4197
#4197 voyelle 2 et diphtongues 13 et 15

#Fichier complémentaire des non I non@
Open 'A_Z9dissyll_10_non_I'
Count -l 'A_Z9dissyll_10_non_I'
7203
#7203 voyelles autres que les diphtongues, schwa et /I/

#Recherche des diphtongues en I dans le fichier des non-schwa en I
#et établissement du fichier complémentaire de mots purement en I œref
Open 'A_Z9dissyll_10_nonschwa_I'
Search -s -q -f 'A_Z9dissyll_10_nonschwa_pureI' /•≈ƒ≈…≈[aeO]I≈Ÿ/
'A_Z9dissyll_10_nonschwa_I' > 'A_Z9dissyll_10_nonschwa_diphtI'
Open 'A_Z9dissyll_10_nonschwa_diphtI'
Count -l 'A_Z9dissyll_10_nonschwa_diphtI'
1309
Open 'A_Z9dissyll_10_nonschwa_pureI'
Count -l 'A_Z9dissyll_10_nonschwa_pureI'
2888
#1309 voyelle 13 & 15, diphtongues fermantes /eɪ/, /ɔɪ/.
#2888 voyelle 2 /ɪ/.
```

méthode d'extraction 2

Avec cette méthode, nous avons écarté dès le départ les entrées à trait d'union de notre extraction. Ensuite, nous avons extrait les voyelles une à une, à l'exception de /ɪ/ dont le traitement avait déjà été effectué.

```
#SM 6 juillet 2001
##extraction des dissyllabes en 10 sans trait d'union
Open 'A_Z9dissyll_10'
Search -s -q -f 'dislInonhyph' /•[¬ ]*-/ 'A_Z9dissyll_10' >
'disyllhyph'
Open 'dislInonhyph'
Count -l 'dislInonhyph'
Open 'disyllhyph'
Count -l 'disyllhyph'19624
  532
```

```
#c'est bon  Maintenant on va extraire voyelle par voyelle les voyelles
#pleines de la 2eme syllabe
##commence par voyelle 1
Open 'dis1lnonhyph'
Search -s -q /•≈f≈…≈i: ≈Ÿ/ 'dis1lnonhyph' > 'voy1nonacc'
Open 'voy1nonacc'
Count -1 'voy1nonacc'  447
#447 voyelle 1 /i:/

#maintenant voyelle 3 (2 est fait, méthode 1)
Open 'dis1lnonhyph'
Search -s -q /•≈f≈…≈e≈Ÿ/ 'dis1lnonhyph' > 'voy3nonacc'
Open 'voy3nonacc'
Count -1 'voy3nonacc' 1231
#attention, diphtongues donc autre traitement nécessaire
## on continue avec les voyelles pures n'étant pas utilisées dans les
symboles des diphtongues

###voyelle 4
Open 'dis1lnonhyph'
Search -s -q /•≈f≈…≈&≈Ÿ/ 'dis1lnonhyph' > 'voy4nonacc'
Open 'voy4nonacc'
Count -1 'voy4nonacc'  567
#567 voyelle 4 /æ/

#voyelle 5
Open 'dis1lnonhyph'
Search -s -q /•≈f≈…≈A: ≈Ÿ/ 'dis1lnonhyph' > 'voy5nonacc'
Open 'voy5nonacc'
Count -1 'voy5nonacc'  365
#365 voyelle 5 /ɑ:/

#voyelle6
Open 'dis1lnonhyph'
Search -s -q /•≈f≈…≈Q≈Ÿ/ 'dis1lnonhyph' > 'voy6nonacc'
Open 'voy6nonacc'
Count -1 'voy6nonacc'  622
#622 voyelle6 /ɒ/

#voyelle 7
Open 'dis1lnonhyph'
Search -s -q /•≈f≈…≈O: ≈Ÿ/ 'dis1lnonhyph' > 'voy7nonacc'
Open 'voy7nonacc'
Count -1 'voy7nonacc'  464
#464 voyelle 7 /ɔ:/

#voyelle 9
Open 'dis1lnonhyph'
Search -s -q /•≈f≈…≈u: ≈Ÿ/ 'dis1lnonhyph' > 'voy9nonacc'
Open 'voy9nonacc'
Count -1 'voy9nonacc'  272
#272 voyelle 9 /u:/
```

```
#voyelle10
Open 'dis11nonhyph'
Search -s -q /•≈f≈…≈V≈Ÿ/ 'dis11nonhyph' > 'voy10nonacc'
Open 'voy10nonacc'
Count -l 'voy10nonacc'  180
#180 voyelle10 /ʌ/

#voyelle 11
Open 'dis11nonhyph'
Search -s -q /•≈f≈…≈œ:≈Ÿ/ 'dis11nonhyph' > 'voy11nonacc'
Open 'voy11nonacc'
Count -l 'voy11nonacc  281
#281 voyelle 11 /ɜ:/

#voyelle 15
Open 'dis11nonhyph'
Search -s -q /•≈f≈…≈aI≈Ÿ/ 'dis11nonhyph' > 'voy15nonacc'
Open 'voy15nonacc'
Count -l 'voy15nonacc'  560
#560 voyelle 15 /aɪ/

#voyelle 17
Open 'dis11nonhyph'
Search -s -q /•≈f≈…≈OI≈Ÿ/ 'dis11nonhyph' > 'voy17nonacc'
Open 'voy17nonacc'
Count -l 'voy17nonacc'  139
#139 voyelle 17 /ɔɪ/

#voyelle 18 I@
Open 'dis11nonhyph'
Search -s -q /•≈f≈…≈I@≈Ÿ/ 'dis11nonhyph' > 'voy18nonacc'
Open 'voy18nonacc'
Count -l 'voy18nonacc'   33
#33 voyelle 18 /ɪə/

#passons à l'extraction des voyelles 13 19 et 3 en utilisant le
#fichier complémentaire
Open 'dis11nonhyph'
Search -s -q -f 'non_eI' /•≈f≈…≈eI≈Ÿ/ 'dis11nonhyph' > 'voy13nonacc'
Open 'voy13nonacc'
Count -l 'voy13nonacc'
Open 'non_eI'
Count -l 'non_eI'  700
18924
#700 voyelle 13 /eɪ/
```

```
#extraction de 19 depuis non_eI
Open 'non_eI'
Search -s -q -f 'non_e@' /•≈ƒ≈…≈e@≈Ÿ/ 'non_eI' > 'voy19nonacc'
Open 'voy19nonacc'
Count -l 'voy19nonacc'
Open 'non_e@'
Count -l 'non_e@'   66
18858
#66 voyelle 19 /eə/

#extraction de 3 depuis non_e@ (qui est aussi  non_eI)
Open 'non_e@'
Search -s -q  /•≈ƒ≈…≈e≈Ÿ/ 'non_e@' > 'voy3nonacc'
Open 'voy3nonacc'
Count -l 'voy3nonacc'  465
#465 voyelle 3 /e/

#extraction de @U 14 aU 16 U@ 20 et enfin U
Open 'dis1lnonhyph'
Search -s -q -f 'non_14' /•≈ƒ≈…≈@U≈Ÿ/ 'dis1lnonhyph' > 'voy14nonacc'
Open 'voy14nonacc'
Count -l 'voy14nonacc'
Open 'non_14'
Count -l 'non_14'  962
18662
#963 voyelle 14 /əu/

#extraction de 16 depuis non_14
Open 'non_14'
Search -s -q -f 'non_1416' /•≈ƒ≈…≈aU≈Ÿ/ 'non_14' > 'voy16nonacc'
Open 'voy16nonacc'
Count -l 'voy16nonacc'
Open 'non_1416'
Count -l 'non_1416'  238
18424
#238 voyelle 16 /au/

#extraction de 20 depuis non_1416
Open 'non_1416'
Search -s -q -f 'non_141620' /•≈ƒ≈…≈U@≈Ÿ/ 'non_1416' > 'voy20nonacc'
Open 'voy20nonacc'
Count -l 'voy20nonacc'
Open 'non_141620'
Count -l 'non_141620'   17
18407
#17 voyelle 20 /uə/

#extraction de 8
Open 'non_141620'
Search -s -q /•≈ƒ≈…≈U≈Ÿ/ 'non_141620' > 'voy8nonacc'
Open 'voy8nonacc'
Count -l 'voy8nonacc'  194
#194 voyelle 8
```

Extraction selon des critères accentuels

Comme nous avons constaté que l'accent /3/ était utilisé dans la notation des composés à accentuation simple dont le second élément est plurisyllabique, nous avons voulu extraire toutes les entrées ayant un accent /3/. Nous avons ensuite écarté celles où l'accent /3/ était en variante, ainsi que les entrées où l'accent /3/ vient avant l'accent /1/ et qui sont donc des dérivés, et non des composés. Restaient des entrées en *-ism* qui se voient attribuer une accentuation tertiaire dans LPD90 et secondaire dans LPD00 (mais les dictionnaires EPD ne le marquent pas). Nous les avons enlevées du fichier. Ayant écarté les variantes, cette extraction a abouti aux prononciations principales avec un accent /3/. Les prononciations secondaires seront récupérées par la programmation suivante, mais les variantes concernent surtout les dérivés.

Recherche des entrées avec un accent 3 susceptibles d'être des composes

```
#SM mardi 8 août 2000
## Extraction des accents /3/
Open 'A_Z9CD;2'
Search -s -q /°/ 'A_Z9CD;2' > 'Acc3'
Open 'Acc3'
Count -l 'Acc3'
2737
#total de 2730 entrées où figure un accent /3/

Open 'Acc3'
Search -s -q -f 'acc3var' /£≈.≈3≈/ 'acc3' > 'acc3p'
Open 'acc3var'
Count -l 'acc3var'
Open 'acc3p'
Count -l 'acc3p'
  683 'acc3var'
 2054 'acc3p'
##683 ont un accent /3/ en variante, 2054 en accent principal.
```

```
Open 'acc3p'
Search -s -q -f 'acc3deriv' /£≈.≈1≈3≈/ 'acc3p' > 'acc3comp'
Open 'acc3comp'
Count -l 'acc3comp'
Open 'acc3deriv'
Count -l 'acc3deriv'
 1657 'acc3comp'
  397 'acc3deriv'
##397 entrées sont des dérivés où l'accent /3/ précède l'accent/1/

Open 'acc3comp'
Search -s -q -f 'acc3vrcomp' /°≈Iz…'@'m/ 'acc3comp' > 'mots_ism'
Open 'acc3vrcomp'
Count -l 'acc3vrcomp'
Open 'mots_ism'
Count -l 'mots_ism'
 1281 'vrais composés'
  376 'mots supposés en -ism'
##376 mots contiennent -ism
#donc les 1281 entrées qui restent sont par hypothèse des composés.

Open 'mots_ism'
Search -s -q /≈msi/ 'mots_ism' > 'mots_ismJLD'
Open 'mots_ismJLD'
Count -l 'mots_ismJLD'
  373
#Cette vérification relève un petit problème mais tout petit: une
différence de 3.
#Mots en -asm sont aussi extraits par cette programmation
```

Recherche des entrées non extraites par les critères précédents susceptibles d'être des composés à accentuation double

La méthode consiste à isoler les entrées ayant un nombre donné de syllabes {£n} et ayant une suite de n'importe quel caractère, suivi de l'accent /2/, suivi de n'importe quel caractère, suivi de l'accent /1/, suivi de n'importe quel caractère (c'est-à-dire la séquence {*2*1*} en langage SQL qui correspond à {≈2≈1≈} dans MPW), puis nous avons écarté celles avec un trait d'union qui ont déjà été extraites.

```
#SM août 2001
#recherche des séquences acc2 pu    acc1
## les fichiers créés à marquer     appelés LPD(n)noH
###où (n) correspond au nombre de syllabes
#2 syllabes
Open 'A_Z9CD;2'
Search -s -q  /£2. 2 1 / 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£' 1609
#2 syllabes en /21/ sans trait d'union
```

```
Open 'LPD£'
Search -s -q -f 'LPD2noH' /•[¬ ]*-/ 'LPD£' > 'LPDH'
Open 'LPD2noH'
Count -l 'LPD2noH'
Open 'LPDH'
Count -l 'LPDH' 1122
  487
1122 entrées en £2 avec /*2*1*/

#3 syllabes
Open 'A_Z9CD;2'
Search -s -q  /£3. 2 1 / 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'  2447
#3 syllabes en /21/ sans trait d'union
Open 'LPD£'
Search -s -q -f 'LPD3noH' /•[¬ ]*-/ 'LPD£' > 'LPDH'
Open 'LPD3noH'
Count -l 'LPD3noH'
Open 'LPDH'
Count -l 'LPDH'   1789
  658
1789 entrées en £3 avec /*2*1*/

#4 syllabes
Open 'A_Z9CD;2'
Search -s -q  /£4. 2 1 / 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£' 3583
#4 syllabes en /21/ sans trait d'union
Open 'LPD£'
Search -s -q -f 'LPD4noH' /•[¬ ]*-/ 'LPD£' > 'LPDH'
Open 'LPD4noH'
Count -l 'LPD4noH'
Open 'LPDH'
Count -l 'LPDH'     3237
  346
3237 entrées en £4 avec /*2*1*/

#5 syllabes
Open 'A_Z9CD;2'
Search -s -q  /£5. 2 1 / 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'   2452
#5 syllabes en /21/ sans trait d'union
Open 'LPD£'
Search -s -q -f 'LPD5noH' /•[¬ ]*-/ 'LPD£' > 'LPDH'
Open 'LPD5noH'
Count -l 'LPD5noH'
Open 'LPDH'
Count -l 'LPDH'     2341
  111
2341 entrées en £5 avec /*2*1*/
```

```
#6 syllabes
Open 'A_Z9CD;2'
Search -s -q  /£6. 2 1 / 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'      726
#6 syllabes en /21/ sans trait d'union
Open 'LPD£'
Search -s -q -f 'LPD6noH' /•[¬ ]*-/ 'LPD£' > 'LPDH'
Open 'LPD6noH'
Count -l 'LPD6noH'
Open 'LPDH'
Count -l 'LPDH'     699
   27
699 entrées en £6 avec /*2*1*/

#7 syllabes
Open 'A_Z9CD;2'
Search -s -q  /£7. 2 1 / 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'   171
#7 syllabes en /21/ sans trait d'union
Open 'LPD£'
Search -s -q -f 'LPD7noH' /•[¬ ]*-/ 'LPD£' > 'LPDH'
Open 'LPD7noH'
Count -l 'LPD7noH'
Open 'LPDH'
Count -l 'LPDH'  166
    5
166 entrées en £7 avec /*2*1*/

#8 syllabes
Open 'A_Z9CD;2'
Search -s -q  /£8. 2 1 / 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'          33
#8 syllabes en /21/ sans trait d'union
Open 'LPD£'
Search -s -q -f 'LPD8noH' /•[¬ ]*-/ 'LPD£' > 'LPDH'
Open 'LPD8noH'
Count -l 'LPD8noH'
Open 'LPDH'
Count -l 'LPDH' 31
    2
31 entrées en £8 avec /*2*1*/
```

```
#9 syllabes
Open 'A_Z9CD;2'
Search -s -q  /£9. 2 1 / 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'              2
#9 syllabes en /21/ sans trait d'union
Open 'LPD£'
Search -s -q -f 'LPD9noH' /*[¬ ]*-/ 'LPD£' > 'LPDH'
Open 'LPD9noH'
Count -l 'LPD9noH'
Open 'LPDH'
Count -l 'LPDH'
0 entrée en £9 avec /*2*1*/
```

Recherche des entrées non extraites par les critères précédents susceptibles d'être des composés à accentuation simple

La méthode consiste à isoler les entrées ayant un nombre donné de syllabes {£n} ayant une suite de n'importe quel caractère, suivi de l'accent /1/, suivi de n'importe quel caractère, suivi de l'accent /0/ (c'est-à-dire la séquence {*1*0} en langage SQL correspondant au {≈1≈0} dans MPW), puis nous avons écarté les entrées dissyllabiques, celles avec un accent /3/, celles avec un accent /2/, et enfin celles avec un trait d'union puisque toutes ces entrées ont déjà été extraites. Une tentative pour associer ces trois critères a échoué (la programmation fait intervenir l'opérateur logique ET et non OU).

```
#SM août 2001
###statistiques concernant des entrées possibles à acc simple avec E2
monosyllabique
##recherche 3 syllabes avec /010/ ou /100/
Open 'A_Z9CD;2'
Search -s -q  /£3. 1 0/ 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'
14836
##écarter les acc 3
Open 'LPD£'
Search -s -q -f 'LPDno3' /. 3  / 'LPD£' > 'LPDa3'
Open 'LPDno3'
Count -l 'LPDno3'
Open 'LPDa3'
Count -l 'LPDa3'
13851
  985
##écarter les acc 2
Open 'LPDno3'
Search -s -q -f 'LPDno23' /. 2  / 'LPDno3' > 'LPDno3a2'
Open 'LPDno23'
Count -l 'LPDno23'
Open 'LPDno3a2'
Count -l 'LPDno3a2'
12985
  866
##écarter les traits d'union
Open 'LPDno23'
Search -s -q -f 'L3mono-H' /•[¬ ]*-/ 'LPDno23' > 'LH'
Open 'L3mono-H'
Count -l 'L3mono-H'
Open 'LH'
Count -l 'LH'
12842
  143
#12842 entrées en 3 syllabes en /100/ ou /010/

###tentative pour regrouper qui NE MARCHE PAS
##extraire les acc 2 et 3
Open 'LPD£'
Search -s -q -f 'LPDmono' /. 2 3  / 'LPD£' > 'LPDa23'
Open 'LPDmono'
Count -l 'LPDmono'
Open 'LPDa23'
Count -l 'LPDa23'14835
    1
###mauvaise programmation car cela comprend et acc3 et acc2 non OU
```

```
####recherche 4 syllabes avec /0100/ ou /1000/
Open 'A_Z9CD;2'
Search -s -q  /£4. 1 0/ 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'
 7929
##écarter les acc 3
Open 'LPD£'
Search -s -q -f 'LPDno3' /. 3  / 'LPD£' > 'LPDa3'
Open 'LPDno3'
Count -l 'LPDno3'
Open 'LPDa3'
Count -l 'LPDa3'
  7367
  562
##écarter les acc 2
Open 'LPDno3'
Search -s -q -f 'LPDno23' /. 2  / 'LPDno3' > 'LPDno3a2'
Open 'LPDno23'
Count -l 'LPDno23'
Open 'LPDno3a2'
Count -l 'LPDno3a2'
 4049
 3318
##écarter les traits d'union
Open 'LPDno23'
Search -s -q -f 'L4mono-H' /•[¬ ]*-/ 'LPDno23' > 'LH'
Open 'L4mono-H'
Count -l 'L4mono-H'
Open 'LH'
Count -l 'LH'
 4043
    6
#4043 entrées en 4 syllabes en /1000/ ou /0100/

####recherche 5 syllabes avec /01000/ ou /10000/
Open '\_Z9CD;2'
Search -s -q  /£5. 1 0/ 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'
  2897
##écarter les acc 3
Open 'LPD£'
Search -s -q -f 'LPDno3' /. 3  / 'LPD£' > 'LPDa3'
Open 'LPDno3'
Count -l 'LPDno3'
Open 'LPDa3'
Count -l 'LPDa3'
 2664
  233
```

```
##écarter les acc 2
Open 'LPDno3'
Search -s -q -f 'LPDno23' /. 2  / 'LPDno3' > 'LPDno3a2'
Open 'LPDno23'
Count -l 'LPDno23'
Open 'LPDno3a2'
Count -l 'LPDno3a2'
  347
 2317
##écarter les traits d'union
Open 'LPDno23'
Search -s -q -f 'L5mono-H' /•[¬ ]*-/ 'LPDno23' > 'LH'
Open 'L5mono-H'
Count -l 'L5mono-H'
Open 'LH'
Count -l 'LH'
  346
    1
#346 entrées en 5 syllabes en /10000/ ou /01000/

####recherche 6 syllabes avec /010000/ ou /100000/
Open 'A_Z9CD;2'
Search -s -q  /£6. 1 0/ 'A_Z9CD;2' > 'LPD£'
Open 'LPD£'
Count -l 'LPD£'
  756
##écarter les acc 3
Open 'LPD£'
Search -s -q -f 'LPDno3' /. 3  / 'LPD£' > 'LPDa3'
Open 'LPDno3'
Count -l 'LPDno3'
Open 'LPDa3'
Count -l 'LPDa3'
  601
  155
##écarter les acc 2
Open 'LPDno3'
Search -s -q -f 'LPDno23' /. 2  / 'LPDno3' > 'LPDno3a2'
Open 'LPDno23'
Count -l 'LPDno23'
Open 'LPDno3a2'
Count -l 'LPDno3a2'
   24
  577
```

```
##écarter les traits d'union
Open 'LPDno23'
Search -s -q -f 'L6mono-H' /*[¬ ]*-/ 'LPDno23' > 'LH'
Open 'L6mono-H'
Count -l 'L6mono-H'
Open 'LH'
Count -l 'LH'
   24
    0
#24 entrées en 6 syllabes en /100000/ ou /010000/

#Les entrées en 7, 8 ou 9 syllabes sont des doublons
```

Prémarquage des fichiers contenant potentiellement des composés à accentuation simple et des fichiers contenant potentiellement des composés à accentuation double

Les fichiers résultant de ces extractions sont très grands (17255 et 9387 entrées respectivement); donc nous avons voulu faire une présélection des entrées qui sont des composés avant de les transférer sur EXCEL. La méthode est la suivante : chaque entrée considérée comme un composé a été marquée manuellement par {!>} en début de ligne ce qui permet ensuite d'extraire le fichier des composés pour les entrées ayant le même nombre de syllabes {£n}. De cette façon, nous avons travaillé sur des ensembles de taille raisonnable. Ensuite, nous avons copié le contenu des fichiers ainsi créés susceptibles de contenir des composés à accentuation simple dans un seul et même fichier appelé LPDdivS, et ceux susceptibles de contenir des composés à accentuation double dans un fichier appelé LPDdivD.

Le fichier LPDdivS comporte 1078 entrées extraites d'un ensemble de 17255 ; le fichier LPDdivD comporte 1701 entrées extraites d'un ensemble de 9387. Ces deux fichiers ne contiennent pas de noms propres, mais beaucoup d'acronymes et de préfixés ont été retenus. L'importance relative du fichier à accentuation double, LPDdivD, reflète le grand nombre d'hésitations concernant les préfixés et quasi-lexèmes associés aux

morphèmes libres. Ces fichiers ont ensuite été transférés sous EXCEL pour le marquage affiné qui a permis de constituer le corpus final puisque les questions sur la composition du corpus d'analyse n'ont été résolues qu'à partir du marquage sous EXCEL.

fichiers contenant potentiellement des composés à accentuation simple

```
#SM août 2001
#0 entrée sur 24 en 6 syllabes sont à considérer
##marquage des composés des fichiers créés sous MPW avec !>
###extraction du fichier des entrées marquées (5 syllabes)
Open 'L5mono-H'
Search -s -q /•[¬f]*!>/ 'L5mono-H' > 'L5s-CP'
Open 'L5s-CP'
Count -l 'L5s-CP'     1
#1 entrée sur 346 en 5 syllabes est à considérer

###extraction du fichier des entrées marquées (4 syllabes)
Open 'L4mono-H'
Search -s -q /•[¬f]*!>/ 'L4mono-H' > 'L4s-CP'
Open 'L4s-CP'
Count -l 'L4s-CP'
   50
#50 entrées sur 4043 en 4 syllabes sont à considérer

#remarque programmation - nécessité d'indiquer le nb de £
Open 'A_Z9CD;2'
Search -s -q /£3.010/ 'A_Z9CD;2' > 'test'
Open 'test'
Count -l 'test' 4409
## 8067 si on ne précise pas £3.010 qui limite la recherche aux
#trisyllabes

Open 'A_Z9CD;2'
Search -s -q /£3.100/ 'A_Z9CD;2' > 'test'
Open 'test'
Count -l 'test'
 9548

####division en /100/ et /010/ du fichier contenant les £3
Open 'L3mono-H'
Search -s -q -f 'L3-100' /.010/ 'L3mono-H' > 'L3-010'
Open 'L3-100'
Count -l 'L3-100'
Open 'L3-010'
Count -l 'L3-010'
 8478
 4364
```

```
###travail de marquage des fichiers divisés puis extraction du fichier
CP des /100/
Open 'L3-100'
Search -s -q /•[¬f]*!>/ 'L3-100' > 'L3-100CP'
Open 'L3-100CP'
Count -l 'L3-100CP'
  997
#997 entrées sur 8478 en 3 syllabes /100/ sont à considérer

###travail de marquage des fichiers divisés puis extraction du fichier
CP des 010
Open 'L3-010'
Search -s -q /•[¬f]*!>/ 'L3-010' > 'L3-010CP'
Open 'L3-010CP'
Count -l 'L3-010CP'
   30
#30 entrées sur 4364 en 3 syllabes /010/ sont à considérer
```

fichiers contenant potentiellement des composés à accentuation double

```
#SM août 2001
#0 entrée sur 2 en 9 syllabes sont à considérer

###travail de marquage des fichiers en *2*1* puis extraction du
##fichier CPdouble des £8
Open 'LPD8noH'
Search -s -q /•[¬f]*!>/ 'LPD8noH' > 'LPD8dCP'
Open 'LPD8dCP'
Count -l 'LPD8dCP'
    3
#3 entrées sur 31 en 8 syllabes sont à considérer

###travail de marquage des fichiers en *2*1* puis extraction du
##fichier CPdouble des £7
Open 'LPD7noH'
Search -s -q /•[¬f]*!>/ 'LPD7noH' > 'LPD7dCP'
Open 'LPD7dCP'
Count -l 'LPD7dCP'
   28
#20 entrées sur 166 en 7 syllabes sont à considérer

###travail de marquage des fichiers en *2*1* puis extraction du
##fichier CPdouble des £6
Open 'LPD6noH'
Search -s -q /•[¬f]*!>/ 'LPD6noH' > 'LPD6dCP'
Open 'LPD6dCP'
Count -l 'LPD6dCP'
  133
#133 entrées sur 699 en 6 syllabes sont à considérer
```

```
###travail de marquage des fichiers en *2*1* puis extraction du
##fichier CPdouble des £5
Open 'LPD5noH'
Search -s -q /•[¬f]*!>/ 'LPD5noH' > 'LPD5dCP'
Open 'LPD5dCP'
Count -l 'LPD5dCP'
  266
#266 entrées sur 2341 en 5 syllabes sont à considérer

###travail de marquage des fichiers en *2*1* puis extraction du
##fichier CPdouble des £4
Open 'LPD4noH'
Search -s -q /•[¬f]*!>/ 'LPD4noH' > 'LPD4dCP'
Open 'LPD4dCP'
Count -l 'LPD4dCP'
  279
#279 entrées sur 3237 en 4 syllabes sont à considérer

###travail de marquage des fichiers en *2*1* puis extraction du
##fichier CPdouble des £3
Open 'LPD3noH'
Search -s -q /•[¬f]*!>/ 'LPD3noH' > 'LPD3dCP'
Open 'LPD3dCP'
Count -l 'LPD3dCP'
  568
#568 entrées sur 1789 en 3 syllabes sont à considérer

###travail de marquage des fichiers en *2*1* puis extraction du
##fichier CPdouble des £2
Open 'LPD2noH'
Search -s -q /•[¬f]*!>/ 'LPD2noH' > 'LPD2dCP'
Open 'LPD2dCP'
Count -l 'LPD2dCP'   424
#424 entrées sur 1122 en 2 syllabes sont à considérer
```

Conclusion

Les programmes décrits ici ont permis d'extraire les entrées susceptibles d'être des composés en les rassemblant par rapport à certains critères. Il reste encore à étudier les fichiers de dissyllabes en /01/ pour s'assurer qu'il n'y a pas de composés. Sur la base de critères accentuels, cela ne devrait pas être régulièrement possible, mais il peut y avoir des exceptions. De même, nous étudierons le fichier des dissyllabes en /10/ avec schwa en /0/ ce qui représente les exceptions à la règle de non-réduction d'un élément de composé. Nous savons qu'il y en aura, notamment en *-man* et *-land*.

Annexe B : Récapitulatif détaillé des fichiers LPD90 : vérification du statut de composé

Le tableau reprend les catégories que nous avons utilisées à partir de notre balayage des entrées fournies par le dictionnaire LPD afin de déterminer lesquelles seraient incluses dans le corpus d'analyse, et lesquelles en seraient exclues. Nous indiquons d'abord la source des entrées extraites selon certains critères : sous-fichiers des mots à trait d'union, les mots avec apostrophe, les dissyllabes en /10/ sans réduction vocalique sur la deuxième syllabe, et enfin les mots ayant un accent /3/ (colonne 1), suivi du nombre total des entrées du sous-fichier (colonne 2). Les autres colonnes donnent la ventilation des entrées, catégorie par catégorie.

Les catégories des entrées prises en considération pour le corpus d'analyse

N° colonne	Entrées incluses / exclues	Contenu de la colonne
colonne 3	Entrées incluses	Les composés communs
colonne 4		Les composés noms propres désignant des objets, des événements, des noms de rues
colonne 5	Entrées exclues	Noms propres qui sont sans doute des composés
colonne 6		Noms propres simples
colonne 7		Acronymes, amalgames, formes tronquées, et les composés opaques ou historiques ou étrangers
colonne 8		Lexèmes simples (communs, non composés)

Figure II.2: Tableau récapitulatif : nombre d'entrées incluses / exclues du corpus d'analyse, par catégorie

		Entrées incluses		*Entrées exclues du corpus d'analyse*			
Type de fichier	Total entrées	Composés communs	Composés noms propres	*Noms propres qui sont sans doute composés*	*Noms propres simples*	*Acronymes Formes tronquées Amalgames Composés historiques /étrangers*	*Lexèmes simples*
Trait d'union	2365	2048	37	*54*	*80*	*69*	*77*
Apostrophes[3]	259	5	3	*50*	*38*	*119*	*44*
Dissyllabes en /10/ Qualité de la voyelle non-accentuée[3]							
iː	447	135	3	*105*	*65*	*4*	*136*
e	465	155	8	*58*	*103*	*13*	*128*
æ	567	279	10	*23*	*121*	*25*	*118*
ɑː	364	145	4	*20*	*111*	*7*	*77*
ɒ	621	215	5	*58*	*167*	*14*	*162*
ɔː	464	206	1	*146*	*53*	*4*	*54*
ʊ	194	67	0	*66*	*19*	*2*	*40*
uː	272	110	2	*4*	*47*	*4*	*105*
ʌ	180	121	1	*14*	*7*	*2*	*35*
ɜː	281	135	1	*65*	*41*	*2*	*37*
eɪ	700	277	10	*114*	*65*	*5*	*229*
əʊ	963	202	2	*63*	*299*	*12*	*385*
aɪ	560	232	6	*31*	*72*	*3*	*216*
aʊ	238	168	0	*43*	*9*	*1*	*17*
ɔɪ	139	33	3	*9*	*15*	*1*	*78*
ɪə	33	7	0	*14*	*5*	*0*	*7*
eə	66	37	2	*9*	*6*	*2*	*10*
ʊə	17	0	0	*4*	*6*	*1*	*6*
TOTAL	6571	2515	58	*846*	*1211*	*102*	*1840*
ɪ	2844	256	7	*177*	*906*	*15*	*1483*
accent /3/[3]	1034	708	16	*94*	*40*	*30*	*146*

[3] Hors entrées ayant un trait d'union.

Annexe C : Les reportages de la source documentaire :

Thèmes et nombre d'extraits relevés

Dans la colonne « thème », les rubriques correspondent aux sens suivants :

Culturel : sujet à caractère culturel : entretiens avec des personnages connus tels que des artistes, acteurs, musiciens, sportifs…

Economique : sujet à caractère économique : la finance et l'économie, projets et problèmes

Politique : sujet à caractère politique

Social : sujet à caractère social : problèmes juridiques, sociaux, historiques

Technologique : sujet à caractère technologique dans des domaines tels que l'informatique, la science, la médecine

La transcription graphique du texte intégral des reportages qui figurent en caractères gras est à la suite du tableau récapitulatif. Pour ces reportages, le nombre total d'extraits sonores figure en premier, suivi d'un nombre entre parenthèses et en caractères gras correspondant à ceux définitivement retenus pour le corpus oral d'analyse. Pour les reportages marqués d'un astérisque, seul un extrait a été retenu, donc le texte de l'ensemble du reportage n'est pas fourni. Son contexte figure dans la concordance de l'Annexe E qui récapitule les extraits sonores du corpus oral d'analyse.

Figure II.3 : *« Documentaires divers »* : Reportages retenus pour la présélection des extraits sonores

Répertoire	Sous-répertoire	Sujet traité	Thème	Nombre d'extraits composés
M-03	ELVIS3	Entretien : femme imitatrice de Elvis Presley	Culturel	16
M-04	INTERVIEW4	Entretien : ancienne chanteuse devenue actrice	Culturel	10
M-06	FISHWOMAN6	Entretien : femme pêcheur	Social	6
M-07	EXCHANGE7	Taux de change et les vacanciers britanniques	Economique	8
M-08	RURAL8	Entretien : Liberal Democrats, problèmes ruraux	Politique	10
M-09	**TOBACCO9**	**Les fabricants de tabac demandent à augmenter l'âge légal pour acheter les cigarettes**	**Social**	12 (7)
M-10	VAT-FAT10	Un médecin propose d'augmenter la TVA sur les produits gras	Social	7
C-01	GMC1	Les Liberal Democrats et les OGM (organismes génétiquement modifiés)	Politique	11
C-02	DOME2	Le coût de l'entrée au Millennium Dome en question	Politique	8
C-03	QUAKE3	Tremblement de terre à Taiwan	Social	11
C-04	EDUCPAY4	Discussion du projet d'augmentation de salaire liée à la performance pour les enseignants	Social	11
C-05	**WEATHER5**	**Un satellite météorologique pour Mars**	**Technologique**	11 (3)
C-06	**LEGAL6**	**Friends of the Earth : action en justice**	**Social**	12 (5)
C-07	PEACEFOR7	Les forces de paix au Timor	Politique	8
C-08	**WARMING8**	**Prescott en Pologne : conférence sur le réchauffement de la planète**	**Politique**	19 (5)
C-09	BIRDS9	Crimes contre des oiseaux sauvages	Social	6
C-10	PATENT10	Pas de brevets sur le projet des gènes humains	Technologique	6
V-02	TAXTIPS2	Comment éviter la taxe sur l'héritage	Economique	3
V-05	WALES5	Le système électoral gallois	Politique	4
V-07	COMPUTE7	Vos questions sur les ordinateurs : émission	Technologique	6
V-08	SPORTS8	Anecdote sportive : noms de stars et équipes	Culturel	2
V-09	GAYRIGH9	Age de consentement et les droits des homosexuels	Social	2
V-10	USTALK10	US talk show – dispute	Social	2
V-11	ONLINE11	Un journal en ligne	Technologique	6

Répertoire	Sous-répertoire	Sujet traité	Thème	Nombre d'extraits composés
P-02	LIBEL2	ITN : procès de diffamation	Social	4
P-04	FILMS4	Femme réalisatrice – réaliser des films au Kenya	Culturel	6
P-06	MUSIC6	Des instruments de musique traditionnels africains	Technologique	3
P-07	MEATMEAL7	Projet pour distribuer la viande au Pays de Galles	Economique	6
P-08	VISTA8	Le projet de photographie Vista dans l'Espace	Technologique	7
P-09	TRADPUB9*	Pour les pubs traditionnels	Social	9
P-10	ZIMBAB10	Des squatters au Zimbabwe	Social	5
P-11	PINOCH11	Le retour de Pinochet au Chili	Politique	4
P-12	PRINTING12	Emission sur l'histoire : les effets de l'imprimerie	Social	4
G-01	DOTCOM1	Le E-business : boo.com	Technologique	15
G-02	WITCHES2	Une commémoration des chasses aux sorcières	Social	5
G-03	BOOKS3	Les livres et les ordinateurs	Social	7
G-05	FANBAN5	Interdiction des supporters turcs aux matches de football	Social	5
G-06	HEALTH6	Aide médicale pour les personnes âgées	Social	12
G-07	EDUCUNI7	Taux bas d'entrants à l'université venant des milieux défavorisés	Social	6
G-08	INTERNET8	Connections Internet	Technologique	10
D-01	SPIELBER1	Entretien : Spielberg et la famille	Culturel	3
D-02	GANGS2	La lutte contre la violence des gangs à Cicero, Etats-Unis	Social	7
D-04	HOMELESS4	Entretien : un évêque controversé	Social	11
D-05	CLINTON5	Entretien: Clinton, la fin du mandat	Politique	6
D-09	UNNEG9	Problèmes des décisions d'intervention pour les Nations-Unies	Politique	7
D-10	MRSCLINT10	Mrs Clinton et New York	Politique	6

Figure II.4: *Reporters* : reportages retenus pour la présélection des extraits sonores

Répertoire	Sous-répertoire	Sujet traité	Thème	Nombre de composés extraits
REPJAN01	AFGHANIS	Aide pour l'Afghanistan	Economique	10
REPJAN01	CHERNOBY	L'après-Tchernobyl	Social	13
REPJAN01	MAFIA	Lutte anti-mafia	Social	17
REPJAN01	**PERUDEBT**	**Problèmes financiers au Pérou**	**Economique**	**24 (20)**
REPJAN01	TFHERAN*	Problèmes sociaux à Téhéran	Social	16
REPFEB28	MARKRICH	Documentaire sur l'escroc Mark Rich	Economique	39
REPFEB28	NIGERIA	La loi musulmane pour les Chrétiens au Nigéria	Politique	8
REPFEB28	RUSSIA	Militaires russes	Politique	8
REPFEB28	TALIBAN	Sanctions des Nations-Unies pour les Talibans	Politique	12
REPFEB28	ROME	Archéologie à Rome	Social	9
REPFEB28	HINDFEST	Festival hindou	Social	3
REPMAR21	ISRAELBL	La situation en Israël	Politique	26
REPMAR21	FOOTANDM	La maladie de la fièvre aphteuse	Economique	13
REPMAR21	LAPLAND	Les droits d'herbage des rênes	Social	11
REPMAR21	PAKISTAN	Action sociale au Pakistan	Social	14
REPMAR21	CONGOTR*	Les affaires au Congo	Economique	5
REPMAR21	**VODKA**	**Une invention: le capuchon parlant de bouteilles de vodka**	**Technologique**	**13 (13)**
REPMAR28	ARGENTIN	La maladie de la fièvre aphteuse en Argentine	Economique	17
REPMAR28	MONGOLIA	La pauvreté en Mongolie	Economique / Social	20

Répertoire	Sous-répertoire	Sujet traité	Thème	Nombre de composés extraits
REPMAR28	PESTICID	Les pesticides dans les bananeraies	Social	19
REPMAR28	TUNISIA	Irrigation : de l'eau pour le désert	Economique	14
REPMAR28	IRANFEST	Le festival solaire de l'Iran	Social	11
REPMAR28	HYDROVIA	Projet sur le fleuve au Paraguay	Economique	15
REPMAR28	VENICE	Une nouvelle matière pour les gondoles de Venise ?	Technologique	10
REPMAR28	VOODOO	Un mariage vaudou	Social	6

Script des reportages retenus pour le corpus oral restreint à partir des sources documentaires

Reporters

DR1 : janvier 2001 : Reportage sur les problèmes économiques du Pérou, et les difficultés que ce pays a rencontrées au niveau du remboursement de sa dette (codé DR1 dans l'Annexe E de ce chapitre).

DR2 : 21 mars 2001 : Reportage sur une nouvelle invention russe : le capuchon parlant de la bouteille de vodka (code DR2 dans l'Annexe E de ce chapitre).

« Documentaires divers »

DD1 : Reportage sur le lancement d'un satellite sur Mars pour contrôler le climat (codé DD1 dans l'Annexe E de ce chapitre).

DD2 : Reportage sur les OGM (organismes génétiquement modifiés) en Grande-Bretagne et les problèmes légaux des essais (codé DD2 dans l'Annexe E de ce chapitre).

DD3 : Reportage sur les problèmes de l'environnement en ce qui concerne les gaz (codé DD3 dans l'Annexe E de ce chapitre).

DD4 : Reportage sur les enjeux de la publicité pour les cigarettes (codé DD4 dans l'Annexe E de ce chapitre).

Figurent en caractères gras les unités analysées. Le récapitulatif des extraits, associés à leur concordance, ainsi que des indications relatives à leur schéma et à leur formation, se trouvent dans l'Annexe E de ce chapitre.

DR1 : *Reporters*, janvier 2001 : Reportage sur les problèmes économiques du Pérou, et les difficultés que ce pays a rencontrées au niveau de sa dette

GB female reporter

Now Peru has been a troubled country of late. First it saw the departure of its president under the shroud of links to a corrupt official, then it discovered that the president had in fact been a citizen of Japan all along where he's now residing. And the country has not enjoyed much economic luck either. But to cap it all, Peru has also found itself a victim of some aggressive behaviour by a firm of **New York investors.** The company, described as a **vulture fund**, recently upset Peru's attempts to get some of its debts written off.

NOISE

GB male reporter - Evan Davis

Feeding time for Frankie and Chance, the vultures of the Bronx zoo. Three dead rats each. It's not a pretty sight. But these are not the only vultures you'll find in New York.

NOISE - outside

Because you'll also find them nesting in the office blocks of downtown Manhattan or indeed in the suburbs. Though the vultures here are a particular breed of investor : investors who pick over the remains of **dead companies and bad debts.** Like their feathered counterparts, they have their uses. They search out for what are called **distressed securities** and they find value in them even when the rest of the world had thought there was none.

Investing in dying companies or bankrupt property is nothing new: it can even help messy situations, but a recent episode of **distressed investing** in the **bad debt** of a **distressed country** has created a worrying precedent.

US male voice

The Peruvian government was attempting to reach an agreement with its creditors so that it would be financially responsible and paid debt that was owed and (hmm) a **vulture fund** from **New York** came and tried to tried to disrupt the process. Peru successfully reached an agreement, but the **vulture fund** ended up collecting 58 million dollars of **hard-earned cash** in Peru.

GB male reporter - Evan Davis

Peru had found itself unable to pay its debts in the nineteen eighties. Peru's leaders agreed to write off half that debt. Just as they did that, a **New York company** got in the way, bought some **distressed Peruvian debt** at a knockdown price, and then sued Peru for payment in full. The company won, and stopped Peru from paying **anyone anything** until it had got all its cash. Peru had to pay up or the litigation would have derailed its attempts to get its debts back on course.

NOISE – outside

The company involved is Elliot Associates. And it resides on the **thirty-sixth floor** of this building. Alas, after several requests, **publicity-shy Elliots** still did not want to grant us an interview. Security saw us off the premises. Other lenders are less shy. In London, banks expecting payment from **restructured Peruvian debt** saw their returns threatened. The precedent is worrying them.

GB male voice

If everyone acted in the way that Elliots Associates acted, none of these countries would ever be able to reservice their debt. And would never therefore gain access to the capital markets again.

GB male reporter - Evan Davis

Everyone knows vultures can be scary, though the tactics of certain **vulture funds** are threatening to give the **birds of prey** a bad name.

Evan Davis, BBC News, New York.

***Reporters*, 21 mars parlant de la bout** **Reportage sur une nouvelle invention russe : le capuchon vodka (DR2)**

GB female reporter

Russian scientists have come up with a way to ensure that no-one will ever have to drink alone ever again. They've designed a **vodka bottle top** which talks back to you. **Mass production** is due to begin next month but our correspondent in Moscow's been given a taste of things to come.

GB male reporter - Steve Rotenberg

Russians have a saying: drinking is bad for you, but not drinking is even worse, and if you don't believe me, listen to this.

NOISE

It's the world's first ever **talking vodka bottle top**. The voice belongs to the **vodka genie** who recites a whole string of popular Russian toasts and encourages you to have a tipple. Each time you take the top off the bottle the **vodka genie**'s speech grows more and more slurred.

NOISE

The **talking bottle top** was developed by scientists at this **workshop** in Moscow. It's packed with the very latest **microchip technology** but it's very simple to use.

FOREIGN VOICE

If these scientists have their way, then very soon, everything in your home could be talking back to you: **armchairs**, tables, even the **kitchen sink**. It's all part of a **ground breaking scheme** designed to give each and every object a tiny part of the Russian soul. And that includes the country's national drink. Mind you, with over a billion litres of vodka consumed each year in Russia you wouldn't think there was a need for a **talking bottle top** to convince people to drink it.

NOISE

But most of the Russians we spoke to were quite taken with the idea. They'd certainly seen nothing like it ever before. Some though missed the point entirely.

NOISE

When the vodka's flowing freely, it's surprising just how quickly you can get through a bottle. But even that's not enough to silence the **vodka genie**.

NOISE

By that stage though chances are you won't be paying much attention to what it's saying.

Steve Rotenberg, BBC News, Moscow.

***Documentaires divers* : C-05 : launching a weather satellite to Mars (DD1)**

GB reporter

The weather forecast: from Mars. The first interplanetary weather satellite is about to begin circling the red planet. Frederick Taylor, the professor of atmospheric physics at Oxford is working at the Jet Propulsion Laboratory in California. Good morning to you.

GB interviewee – Frederick Taylor

Good morning.

GB reporter

Er, is it there yet, has it arrived at Mars?

GB interviewee – Frederick Taylor

Er, no not quite. We've got a couple more hours to go.

GB reporter

And then what?

GB interviewee – Frederick Taylor

Well the action starts round about nine o'clock when the spacecraft will fire its motors and er that will last about **fifteen or sixteen minutes** and then it should er go into Mars orbit and we'll start getting signals back round about ten o'clock.

GB reporter

And they'll tell you what, you hope?

GB interviewee – Frederick Taylor

Well, er, it'll hopefully tell us it's in the correct orbit, er we won't begin observing the planet straight away.

GB reporter

But when you do, when will that be?

GB interviewee – Frederick Taylor

That actually won't be until next February.

GB reporter

Oh I see, a long time away.

GB interviewee – Frederick Taylor

Yes, er, the **spacecraft** really parked in orbit for quite a long time 'cos it's got to adjust the orbit over a long period to get into the right one. We want to be in very close circular orbit to get close to the planet.

GB reporter

Why do you need to be in such a specific position in order to tell about the weather?

GB interviewee – Frederick Taylor

Well because we want to er map the atmosphere in a lot of detail and er that requires getting the instrument very close to the planet.

GB reporter

And why do you want to do that? I mean apart from simply wanting to know of course.

GB interviewee – Frederick Taylor

Er why are we doing the whole thing? Er, well, because we've been doing this sort of thing on the Earth for a long time now and we think we know how to do it pretty well and it's time to look at another planet. Mars is the planet which has an atmosphere the most like the Earth and we expect to see similar kinds of weather systems and things

going on there. We're also trying to look at the Martian climate which means sort of long-term weather because we know it's changed dramatically. Mars used to be much warmer and wetter than it is now, and we don't know why it's changed. So of course we're very obsessed about **climate change** on the Earth at the moment. So we've got to understand the basic principles and Mars is a good place to look.

GB reporter

Oh I see, so that may tell us about what's happening to us. How similar to our climate is Mars or is it completely dissimilar?

GB interviewee – Frederick Taylor

No, it's fairly similar in many ways. Er in fact it used to be very – very similar to the Earth as far as we can tell from the geological evidence, er, but it's now, the atmosphere is now much thinner which means that there's no more liquid water left on the surface. But, er, there used to be as far as we can tell. We can see evidence for river beds and oceans and so on.

GB reporter

So does that mean that there was life there, or might have been?

GB interviewee – Frederick Taylor

Well that's the sixty-four thousand dollar question. I think – I think we can say with a fair amount of confidence that the conditions for life were quite good in the past, but whether that actually means there really was life, that's the thing we'd really like to know.

***Documentaires divers* : C-06 : Friends of the Earth legal challenge on genetically modified crops (DD2)**

GB reporter

The government's decision not to contest the main part of the legal challenge by Friends of the Earth to trials of **autumn-sown oilseed rape** is an embarrassment to ministers. The crops were to be part of larger scale farm trials set up by the government to assess the impact of **modified crops** on the environment.

The biotechnology company Agrivo already had permission to plant **spring-sown genetically modified oilseed rape** but officials in the Department of the Environment altered the licence, allowing the company to plant the crop in autumn and to quadruple the area on which it was tested to five thousand hectares. Friends of the Earth issued a legal challenge to the decision and earlier this month were given clearance to seek a judicial review.

Today the Environment Minister, Michael Meecher, admitted the government had been wrong not to insist that Agrivo made up an entirely new application for the new trials.

GB male interviewee – Michael Meecher

We accept that the government did act illegally but there is no question of health or safety or of the environment concerned. What is at fault is a narrow, legal tec. nicality and as soon as it became apparent to us er that we had acted beyond our powers er then we have immediately taken steps to put it right. We've informed the court, we have informed the company Agrivo, we have informed Friends of the Earth.

GB reporter

Agrivo said a week ago that it would continue with the trials of the **genetically modified autumn sown oilseed** rape despite the legal action and has planted it on three of the four sites. Ministers say they won't make Agrivo end the trials because at the time they were sown the company was acting in good faith on the basis of the

permission it has obtained. However, Friends of the Earth say they're considering further legal action over the trials. The group's legal advisor, Peter Roderick, says it was clear some time ago that the rules had been broken.

GB male interviewee – Peter Roderick

The idea that the seeds have been planted in good faith, it's the first time I've actually heard that said, is incredible. These – these seeds were planted last week after the government had written to the court asking for a quick hearing - um – knowing that if the courts had ordered the quick hearing and had found for us they would have to have been dug up. These seeds have been planted knowing full well that the legalities of the case were on our side. Now the government's admitted it.

GB reporter

Michael Meecher said that he was sure that Agrivo would have obtained permission for the extended trials had it made a fresh application. He stressed that the programme of farm trials of **genetically-modified crops** would continue next spring unaffected by today's decision. A total of seventy-five are planned by next year. Meanwhile, however, the government could face the embarrassment of seeing its latest farm trials declared illegal and dug up.

***Documentaires divers* : C-08 : Report on global warming (DD3)**

(GB female reporter)

Deputy Prime Minister John Prescott is in Poland today. He's in Warsaw at a conference reviewing international progress on reducing global warming. Well our environment correspondent Robert Piggot is on the line. Robert, we're going to hear from John Prescott himself later in the programme. He'll no doubt claim that Britain's doing rather better than most people in trying to reduce global warming. Is that right?

(GB male reporter, Robert Piggot)

Well in many ways yes. Britain actually agreed to cut emissions by 12.5% on 1990 levels compared to an average for the industrialized world of 5%. But in this country many coal-fired stations which produced a lot of carbon dioxide have been replaced with cleaner burning gas powered stations so we are well on course in a perhaps slightly fortuitous way to meeting that target so there is sort of latitude for John Prescott to be something of a leader in this forum.

(GB female reporter)

This is all about looking back to the Kyoto **climate change conference** two years ago and and sort of making markers all along the way to the next one seeing who's doing what?

(GB male reporter, Robert Piggot)

That's right. A number of nations were agreed at Kioto to reduce the **emission of greenhouse gases**: that included a timetable later on in Buenos Aires in fact last November. Those measures include such things as **emissions trading** which people may remember: that's buying the right to produce each ton of carbon dioxide you produce and that penalizes those who produce a lot, provides an incentive to technology to reverse that situation. And but there are a lot of loose ends to tie up before those measures can actually practically start working. Things like what proportion of the **total greenhouse gas emissions** a country's allowed to buy or sell and how that we can be

sure they're sticking to the rules. So it's getting that immense detail worked out and the **timetable** adopted that this meeting's all about, yes.

(GB female reporter)

And some some people might say at home the government's sending out a mixed message, because home insulation is taxed more highly because of VAT than home fuel.

(GB male reporter, Robert Piggot)

Yes, the, again I think a lot of the things that are happening in the UK are because of internal politics and and the need to tackle other issues. And

(GB female reporter)

John Prescott isn't in charge of tax.

(GB male reporter, Robert Piggot)

That's right and he's frequently there's been frequent grumbling among environmental groups that he hasn't been seen to control things like that or or have much control on how the government performs on tax. In a sense because the UK has its own target well under control the priority is for him certainly more to persuade other countries to do their bit particularly people like the United States, a a country with four or five percent of the population of the world but which produces almost a quarter of its emissions, and getting these mechanisms up and running in order to persuade the developing world, poorer countries like India and China who have potentially huge emissions of their own to actually join the process. At the moment they say they won't until the industrialized world gets its act together and and of course the United States is is refusing to do anything without the developing world, so it's breaking this log jam by getting these practical measures up and running that's so important now.

Documentaires divers : **M-09 : Report on tobacco companies (DD4)**

(GB male reporter, Humphreys)

Britain's biggest tobacco firm, British American Tobacco, has called on Tony Blair to raise the minimum legal age for buying cigarettes to eighteen. Well on the line is John Carlisle who's the Public Affairs Director of the Tobacco Manufacturers' Association. What do you make of this Mr Carlisle?

GB male interviewee, John Carlisle

Well, BAT have had this line for quite some time and been joined recently by Philip Morison calling for the er the same approach. I think what they're trying to say is let's move this argument on. There's been too much talk I think in the last few years about old history, what happened in the past and so on and what BAT are I think very sensibly saying is we want to talk to the government, we want a better dialogue between our customers, government er and and indeed the general public. Let's move the argument forward and this could be a way perhaps of soving, solving the terrible problem which we've got at the moment which is **under-age smoking**. There's other abuses of course **under-age**, of sex, alcohol and so on, but as far as tobacco's concerned, they believe that the legal age should be raised to eighteen of purchasing a cigarette.

(GB male reporter, Humphreys)

If if you're so worried about young people buying cigarettes, why don't you stop advertising them the way you do?

GB male interviewee, John Carlisle

Well, advertising doesn't actually appeal to young people. I think you'll appreciate ...

(GB male reporter, Humphreys)

What?!

GB male interviewee, John Carlisle

You'll appreciate Mr Humphreys that we can't appeal to young people anyway by the voluntary code, em, but I've got to say to you as a **non-smoker** that I find a lot of the advertisements that are put out by **cigarette companies** frankly somewhat perplexing. What we do in advertising is to try and switch people from one brand to another. No advert actually appeals to young people, no child sees an advertisement like a block of ice with a piece of of a **razor blade** and a piece of silk in it and actually goes out and says gosh I see this advert I I must have a fag. That is not...

(GB male reporter, Humphreys)

So who does then? I mean do you have to be sort of seventeen eighteen nineteen twenty before you get that message?

GB male interviewee, John Carlisle

Oh yes, you have to be sophisticated enough to be an existing **adult smoker** to understand that perhaps by the that that advertisement you may switch brands.

(GB male reporter, Humphreys)

You simply can't prove something like that.

GB male interviewee, John Carlisle

Well, I think it's common sense frankly. You go out and look at the advertisements which we're still allowed to portray, which I say again we're not allowed to target, quite rightly, **under-age smokers**, and tell me if really somebody looking at a rather morose person standing by a **phone by by a pillar box** em of a particular colour is going to say, well because I've seen that advert I think I'll take up smoking.

Annexe D : La source littéraire

Organisation en répertoire et nombre des extraits sonores

Le tableau suivant récapitule l'organisation en répertoires des extraits sonores et indique le nombre d'extraits par répertoire. Nous commentons ce tableau à la suite.

Figurent en caractères gras les répertoires contenant les extraits retenus pour ie corpus oral d'analyse. Les chiffres entre parenthèses correspondent au nombre d'extraits qui se trouvent aussi dans un autre répertoire.

Figu. · U.5 : Les répertoires des extraits de la source littéraire Harry Potter

Répertoires	Commentaire	Sous-répertoires	Sous-sous-répertoires	Nombre de composés extraits
Activities	Les composés relatifs à des jeux et activ 's, subdivisés par rapport au jeu des sorciers Quidditch et d'autres activités diverses.	Quidditch		35
		general		2 (+2)
animals	Les composés désignant des espèces d'animaux, subdivisés par rapport aux animaux souvent rencontrés dans l'œuvre. Le répertoire \dragon comprend un sous-répertoire complémentaire regroupant les noms de types de dragons.	cat		3
		doghound		4
		dragon	general	11
			type	7
		owl		4
		troll		6
		werewolf		4
comparison	Les composés relevés à des fins de comparaison, divisés pour regrouper les adjectifs composés et ceux ayant l'élément *dark*, *magic* et *sorting*.	adjectives		11
		dark		16
		magic		10
		sorting		7
discourse	Ce répertoire regroupe les extraits ayant un intérêt au niveau de l'analyse en discours.			10 (+ 22)
food	Les composés désignant différentes sortes de nourriture, divisés pour regrouper ceux ayant l'élément *cake* et *chocolate*, et les autres.	cakes		6 (+1)
		chocolate		8
		other		15
lessons	Les composés formés à partir des différents cours			6 (+3)
NP	Les composés formés à partir de noms propres, divisés en fonction des « maisons », des personnes, des lieux et des objets.	houses	Hufflepuff	5
			Gryffindor	26
			Ravensclaw	3
			Slytherin	13
		people	Dumbledore	2
			Harry	3
			Malfoy	7
			others	13
			Voldemort	6
			Muggle	11
		places	Hogwarts	17
			Gringotts	6
			others	16
		things		11

Répertoires	Commentaire	Sous-répertoires	Sous-sous-répertoires	Nombre de composés extraits
objects	Les composés désignant des objets divers, divisés pour regrouper les mots relatifs aux balais, aux vêtements et divers autres objets magiques. Les objets existant dans notre monde sont dans \tests.	**brooms**		9 (+2)
		clothes		13
		others		19
		tests		26
		others		5 (+1)
places	Les composés désignant des lieux, divisés pour regrouper les composés avec l'élément *corridor*, *hole*, *room* puis des adverbes et les autres lieux.	adverbs		3
		corridor		3
		hole		5
		others		17
		room		8
spellcharm	Tous les composés relatifs à la magie, divisés en ingrédients de potions, noms de potions, la magie (ou les sortilèges), puis les composés ayant l'élément *spell*, *wand* et *wizard*.	curse		17
		ingredient		26 (+1)
		others		12
		potion		12
		spell		5
		wand		3 (+3)
		wizard		20 (+1)

Les répertoires en caractères gras correspondent à ceux qui contiennent les extraits retenus pour le corpus oral d'analyse. Le détail concernant ces extraits se trouve dans l'annexe suivante.[4]

Il est essentiel de faire une répartition des extraits sonores en répertoires à cause de leur nombre. Dans l'analyse acoustique, il faut pouvoir accéder rapidement aux extraits qui nous intéressent. Nous ne pouvons pas adopter la même méthode pour la source littéraire que pour les sources documentaires car l'enregistrement compose un ensemble et les mêmes extraits apparaissent tout au long de l'œuvre : il serait vain de les

[4] Informations concernant le schéma accentuel, la formation, la consonance, ainsi que l'identification du fichier son.

classer selon les chapitres par exemple. Nous avons donc cherché un classement thématique en créant un certain nombre de répertoires. Grâce à cette ventilation des données, nous pouvons rassembler les extraits, ce qui facilite le traitement et en clarifie la présentation. Cette organisation nous permet de faire les rapprochements que l'analyse acoustique exige, et d'arriver à manier des répertoires de taille raisonnable (aucun ne dépasse 30 extraits).

Le choix des répertoires s'est fait selon deux critères : le thème auquel l'extrait peut être associé ; et les éléments individuels de l'extrait en fonction des comparaisons que nous voudrons faire. C'est ce dernier critère qui a motivé l'organisation des sous-répertoires : c'est-à-dire que nous tentons à chaque fois de rassembler des éléments que nous voudrons comparer.

La plupart des répertoires sont thématiques, à l'exception de « \comparison » et « \discourse ». Dans le répertoire « \comparison » nous avons rassemblé des mots fonctionnant comme des adjectifs dans les extraits ; si nous avions adopté une organisation thématique pour ces extraits, nous aurions perdu les rapprochements que nous souhaitons faciliter. Nous avons suivi la même logique que pour le classement en sous-répertoires.

En ce qui concerne les composés lexicalisés, le seul répertoire qui les rassemble systématiquement à part est « objects\test ». Ces composés ont souvent été regroupés avec les composés nouveaux avec lesquels nous voulons les comparer pour faciliter l'accès aux données. Ainsi, *staffroom* est dans le répertoire « \places\room » avec *flying-key room*.

Certaines unités peuvent se trouver dans deux répertoires différents : par exemple, *dragon's blood* se trouve à la fois dans « animal\dragon » et dans « spellcharm\ingredient ». En effet, il y a 543 fichiers, mais seulement 507 extraits au total ; il y a donc quelques 36 extraits dans deux répertoires différents.

Enfin, l'autre répertoire non thématique, « discourse », regroupe des extraits relevés afin d'étudier le comportement en discours. Ce répertoire contient souvent des séquences non composées comme *Dudley's birthday*, *twelve-foot mountain troll*, et *dark cupboard*. Il comporte aussi quelques adjectifs composés suivis de noms comme *jewel-bright birds*, et des séquences se terminant par *one*, comme *vomit-flavoured one*. Nous avons aussi relevé quelques exemples d'accent de phrase. Nous avons donc inclus des extraits sonores utiles pour notre travail sur les composés mais aussi des extraits qui seront intéressants à étudier ultérieurement, dans le cadre d'une analyse du discours plus vaste.

Annexe E : Le corpus oral d'analyse

Cette annexe contient les extraits sonores retenus pour le corpus oral d'analyse. Ces extraits ont été rassemblés et classés dans des répertoires correspondant à leur origine et / ou à leur intérêt sur le plan de l'analyse. Nous présentons d'abord un récapitulatif des correspondances entre les colonnes du tableau et les informations.

Les extraits sonores du corpus oral d'analyse

Figure II.6 : Informations correspondant au tableau récapitulatif (Figure II.7)

N° colonne	Titre	Contenu	Codage
1	Source	Indique de quelle source provient l'extrait	HP : *Harry Potter and the Philosopher's Stone* DR1: *Reporters*, reportage sur l'endettement du Pérou (texte intégral Annexe C). DR2: *Reporters*, reportage sur le capuchon parlant (texte intégral Annexe C). DR3: *Reporters*, reportage sur les Talibans (extrait). DR4: *Reporters*, reportage sur le commerce au Congo (extrait) DD1: *« Documentaires divers »*, reportage sur le satellite envoyé sur Mars (texte intégral Annexe C) DD2: *« Documentaires divers »*, reportage sur les essais d'OGM (texte intégral Annexe C). DD3: *« Documentaires divers »*, reportage sur le contrôle des gaz (texte intégral Annexe C). DD4: *« Documentaires divers »*, reportage sur la publicité pour les cigarettes (texte intégral Annexe C). DD5: *« Documentaires divers »*, reportage sur les *pubs* (extrait)
2	Fichier son[5]	Indique le nom du fichier son qui est précédé de celui du répertoire. L'extrait se trouve dans ce répertoire sur le CD.	
3	Concordance	Indique le contexte de l'extrait. Les caractères gras correspondent à l'étendue de l'extrait sonore.	
4	Graphie	Indique la graphie (supposée) de la partie analysée de l'extrait. Cette graphie pourra être comparée avec celles des dictionnaires de la base de données.	
5	Nb syll	Indique le nombre de syllabes correspondant à la partie analysée de l'extrait.	
6	Schéma perçu	Indique le schéma accentuel effectivement perçu de l'extrait (la partie analysée), en incluant les bornes de frontières morphologiques (#) comme pour la table des composés non compacts de LPD90.	
7	s/d perçu	Indique le code « s » ou « d » correspondant au type de schéma.	« s » pour l'accentuation simple « d » pour l'accentuation double Dans le cas de constructions complexes, il y a deux codes (ou plus). Le premier code correspond à l'unité la plus grande. Lorsqu'il y a des unités plus petites à l'intérieur de la construction, leur type de schéma est indiqué de la même façon, après celui de l'ensemble : *underage smoker* indique d'abord le schéma de l'ensemble, puis celui de *underage* dans cet ensemble.

[5] Les extraits sonores se trouvent sur le CDROM dans le répertoire \corporal, dans le sous-répertoire indiqué.

8	s/d prévu	Indique le type de schéma prévu par le dictionnaire[6] pour les unités lexicalisées[7].	« s » pour l'accentuation simple « d » pour l'accentuation double Pour les unités non lexicalisées formées à partir d'une unité lexicalisée, la prévision est indiquée entre parenthèses.
9	Type cp	Indique le code du type de composé (ou autre type de formation)[8].	CT: composé radical (*root compound*) CT' : composé radical avec apostrophe CO : composé co-ordonné CX : composé exocentrique (*bahuvrihi*) CS : composé synthétique (*verbal compound*) CR : composé réduplicatif OPC : composé formé à partir d'un nom propre CC : composé dans une construction cumulative QL : quasi-lexème P : préfixe S : syntagme
10	Cat gr	Indique la catégorie grammaticale du composé.	N : nom Q : adjectif SN : syntagme nominal
11	Formation	Indique la formation de la partie analysée de l'extrait.	N : nom Pron : pronom Q : adjectif V : verbe Ving : verbe en -ing Adv : adverbe Prep : préposition Conj : conjonction

[6] Les indications sont celles de LPD90 sauf si l'unité est absente de ce dictionnaire mais figure dans un autre. Des précisions sur l'accent lexical et le corpus oral se trouvent en annexe au chapitre 4 (Annexe C).

[7] Pour la source *Harry Potter and the Philosopher's Stone*, figurent des prévisions pour les unités non lexicalisées correspondant à l'intuition de l'anglophone (nous-même) lors de la lecture du livre. Ces prévisions sont indiquées en italiques.

[8] Pour une discussion de ces différentes catégories, voir chapitre 4 : 4.3.2.4, « Etude de la catégorie syntaxique »..

Figure II.7 : Tableau récapitulatif des extraits sonores du corpus oral d'analyse, classés selon le répertoire et le fichier son

Source	Fichier son	Concordance	graphie	Nb syll	Schéma perçu	s/d perçu	s/d prévu	TypeCp	Cat gr	Formation
HP	broom\broomshed1	to take his Nimbus Two Thousand back to the **broomshed**. He couldn't ever remember feeling happier. (165)	broomshed	2	1#0	s	*s*	CT	N	N+N
HP	broom\broom1	their noses pressed against a window with **broomsticks** in it. (56)	broomsticks	2	1#0	s	s	CT	N	N+N
HP	broom\broom2	and so were twenty **broomsticks** lying in neat lines on the ground. (109)	broomsticks	2	1#0	s	s	CT	N	N+N
DR4	broom\DR-bike	Meet the **original mountain bike**: strong, durable...	mountain bike	3	10#0	s		CT	N	N+N
HP	broom\racingbike1	bike. Exactly why Dudley wanted a **racing bike** was a mystery to Harry, as (20)	racing bike	3	10#0	s	*s*	CT	N	N/Ving+N
HP	broom\racingbr1	I'm going to drag them off to look at **racing brooms**. I don't see why first-years (60)	racing brooms	3	10#0	s	*s*	CT	N	N/Ving+N
HP	broom\schbroo1	complain about the **school brooms**, saying that some of them started to vibrate (109)	school broom	2	?#1?	d	*d*	CT	N	N+N
HP	broom\schowl1	to sleep in the owlery with the other **school owls**. This morning, however, she (101)	school owls	2	P#1	d	*d*	CT	N	N+N
HP	dark\antidark1	and Quirrell would have done some **anti-Dark Arts** spell which Snape needs to (165)	anti-Dark Arts spell	5	R0#P#1#0	s d	*s*	CC/CT	N	Q+N+N+N
HP	dark\antidark2	If Quirrell's told him how to break his **Anti-Dark Force** spell - There's still Fluffy, (180)	Anti-Dark Force spell	5	R0#P#1#0	s d	*d*	CC/CT	N	Q+N+N+N
DR3	dark\darkage1	in its attempt to create the purest Islamic society, the Taliban has turned the clock back to the **Dark Ages**, on purpose	Dark ages	3	1#30	s	s	CT	N	Q+N
HP	dark\darkart1	job. Knows an awful lot about the **Dark Arts**, Snape.' Harry watched Snape for (94)	Dark Arts	2	1#0	s	*d*	CT	N	Q+N
HP	dark\darkart2	it, or turn it into a school for the **Dark Arts**! Losing points doesn't matter (196)	Dark Arts	2	1#0	s	*d*	CT	N	Q+N
HP	dark\darkcupb1	Harry lay in his **dark cupboard** much later (27)	dark cupboard	3	P#10	d	*d*	S	SN	Q+N
HP	dark\darkforce1	to Find Them by Newt Scamander **The Dark Forces**: A Guide to Self-Protection by (53)	Dark Forces	3	1#30	s	*d*	CT	N	Q+N
HP	dark\darkmag1	with a broomstick except powerful **Dark Magic** - no kid could do that to a (140)	Dark Magic	3	1#30	s	*d*	CT	N	Q+N
HP	dark\darkmag2	[146] the books containing powerful **Dark Magic** never taught at Hogwarts and (146)	Dark Magic	3	P#10	d	*d*	CT	N	Q+N

Source	Fichier son	Concordance	graphie	Nb syll	Schéma perçu	s/d perçu	s/d prévu	TypeCp	Cat gr	Formation
HP	dark\darksid1	ter want anythin' ter do with the **Dark Side**. 'Maybe he thought he could (45)	Dark Side	2	P#1	d	*s*	CT	N	Q+N
HP	dark\darksid2	didn't need an excuse to go over to the **Dark Side**.'He turned to Hermione. 'Can we (82)	Dark Side	2	P#1	d	*s*	CT	N	Q+N
HP	dark\darksid3	because I'm never going over to the **Dark Side**! I'm going through that trapdoor (197)	Dark Side	2	1#0	s	*s*	CT	N	Q+N
HP	dark\darkwiz1	famous for his defeat of the **dark wizard** Grindelwald in 1945, for the (77)	dark wizard	3	1#30	s	*s*	CT	N	Q+N
HP	dark\darkwiz2	My dad says it must've been a powerful **Dark wizard** to get round Gringotts, but (80)	Dark wizard	3	P#10	d	*d*	CT	N	Q+N
HP	dark\darkwiz1A	famous for his defeat of the **dark wizard Grindelwald** in 1945, for the (77)	dark wizard Grindelwald	6	P#30#100	d s	*d*	CC/S	SN	Q+N+N
HP	dark\darkwiz1B	July, widely believed to be the work of **dark wizards or** witches unknown. (105)	dark wizards or witches	6	P#P0#0#10	d	*d*	S	SN	Q+N+CONJ+N
HP	dark\defdark1	had really been looking forward to was **Defence Against the Dark Arts**, but (100)	Defence Against the Dark Arts	7	0P#0P#0#P#1	d d	*d*	CC/S	SN	N+PREP+N
HP	dark\defdark2	by older students studying advanced **Defence Against the Dark Arts**. 'What are (146)	Defence Against the Dark Arts	7	0P#0P#0#P#1	d d	*d*	CC/S	SN	N+PREP+N
HP	dark\defdark3	is he?' [162] Next morning in **Defence Against the Dark Arts**, while (162)	Defence Against the Dark Arts	7	0P#0P#0#P#1	d d	*d*	CC/S	SN	N+PREP+N
DR1	debt\baddebt1	but a recent episode of distressed investing in the **bad debt of a distressed country** created a worrying precedent.	bad debt	2	1#0	s	d	CT	N	Q+N
DR1	debt\deadcom1	investors who pick over the remains of **dead companies and bad debts**. Like their feathered counterparts,	bad debts	2	P#1	d	d	CT	N	Q+N
DR1	debt\baddebt1	but a recent episode of distressed investing in the **bad debt of a distressed country** created a worrying precedent.	distressed country	4	0P#10	d		CT	N	Q+N
DR1	debt\distinv1	but a recent episode of **distressed investing** in the bad debt of a distressed country created a worrying precedent.	distressed investing	5	0P#010	d		CT	N	Q+N
DR1	debt\distper1	a New York company got in the way, bought some **distressed Peruvian debt at a knockdown price**, then	distressed Peruvian debt	7	0P#0P0#1	d d		CC/S	SN	Q+Q+N
DR1	debt\distsec1	They search out for what are called **distressed securities** and then find value in them	distressed securities	6	0P#0100	d		CT	N	Q+N
DR1	debt\restruct1	banks expecting payment from **restructured Peruvian debt** saw their returns threatened.	restructured Peruvian debt	8	0P0#0P0#1	d d		CC/S	SN	Q+Q+N
DR1	debt\vultfun1	The company, described as a **vulture fund**, recently upset Peru's attempts to get some of its debts written off.	vulture fund	3	10#0	s		CX	N	N+N

Source	Fichier son	Concordance	graphie	Nb syll	Schéma perçu	s/d perçu	s/d prévu	TypeCp	Cat gr	Formation
DR1	debt\vultfun2	paid debt that was owed and (hmm) a vulture fund from New York came and tried to tried to disrupt the process	vulture fund	3	10#0	s		CX	N	N+N
DR1	debt\vultfun3	Peru successfully reached an agreement, but the vulture fund ended up collecting 58 million dollars of hard-earned cash in Peru.	vulture fund	3	10#0	s		CX	N	N+N
DR1	debt\vultfun4	though the tactics of some of the vulture funds are threatening to give the birds of prey a bad name.	vulture funds	3	P0#1	d		CX	N	N+N
HP	dragon\dragbite1	Madam Pomfrey - would she recognise a dragon bite? By the afternoon, though, he (174)	dragon bite	3	10#0	s	*s*	CT	N	N+N
HP	dragon\dragdeal1	the pub down in the village. Mighta bin a dragon dealer, mightn' he? I never saw his (193)	dragon dealer	4	10#30	s	*s*	CT	N	N+N
HP	dragon\dragegg2	Let's see ... yeah, then he said he had the dragon egg an' we could play cards fer it if (193)	dragon egg	3	10#0	s	*s*	CT	N	N+N
HP	dragon\dragegg3	I told him! Yeh could've died! All fer a dragon egg! I'll never drink again! I should (219)	dragon egg	3	10#0	s	*s*	CT	N	N+N
HP	dragon\dragegg1	How many people wander around with dragon eggs if it's against wizard law? (193)	dragon eggs	3	10#0	s	*s*	CT	N	N+N
HP	dragon\dragheart1	Mr Potter. Try this one. Beechwood and dragon heartstring. Nine inches. Nice and (64)	dragon heartstring	4	10#3#0	s	*s*	CT	N	N+N
HP	dragon\dragliver1	shaking her head as they passed, saying, 'Dragon liver, seventeen Sickles an ounce, (56)	Dragon liver	4	10#30	s	*s*	CT	N	N+N
HP	dragon\dragbreed1	'But it's against our laws,' said Ron. 'Dragon-breeding was outlawed by the ...(169)	Dragon-breeding	4	10#30	s	*s*	CT	N	N+N
HP	dragon\dragbreed2	'Got this outta the library - Dragon-Breeding for Pleasure and Profit (170)	Dragon-breeding	4	P0#10	d	*s*	CT	N	N+N
HP	dragon\dragblood2	for the discovery of the twelve uses of dragon's blood and his work on alchemy (167)	dragon's blood	3	10#0	s	s	CT'	N	N's+N
HP	dragon\dragblood1	for the discovery of the twelve uses of dragon's blood and his work on alchemy (77)	dragon's blood	3	10#0	s	s	CT'	N	N's+N
HP	dragon\draghead1	He reached out a hand to stroke the dragon's head. It snapped at his fingers, (172)	dragon's head	3	P0#1	d	*d*	S	SN	N+'s+N
DD3	emiss\green1	to reduce the emission of greenhouse gases: that included a timetable	greennouse gases	4	P#0#10	d s	d (s)	CC	SN	N+PREP+N
DD3	emiss\emiss1	Those measures include such things as emissions trading which people may remember	emissions trading	5	0P0#10	d		C	N	Npl+N

Source	Fichier son	Concordance	graphie	Nb syll	Schéma perçu	s/d perçu	s/d prévu	TypeCp	Cat gr	Formation
DD3	emiss\emiss2	what proportion of the **total greenhouse gas emissions** a country's allowed to buy or sell	greenhouse gas emissions	6	P#0#1#030	s/d	(d s)	CC/S	SN	Q+N
HP	hole\keyhole1	it had already been caught and stuffed roughly into the **keyhole**. 'That one!' (203)	keyhole	2	1#0	s	s	CT	N	N+N
HP	hole\portrh1	shadows. They had almost reached the **portrait hole** when a voice spoke from the (115)	portrait hole	3	10#0	s	*s*	CT	N	N+N
HP	hole\portrh2	She followed Ron [116] through the **portrait hole**, hissing at them like an (116)	portrait hole	3	10#0	s	*s*	CT	N	N+N
HP	hole\portrh3	common room and climbed through the **portrait hole**. 'Who's there?' squawked the (151)	portrait hole	3	10#0	s	*s*	CT	N	N+N
HP	hole\portrh4	wall had just chimed midnight when the **portrait hole** burst open. Ron appeared out (173)	portrait hole	3	10#0	s	*s*	CT	N	N+N
DR1	newyork\newyork2	a vulture fund **from New York** came and tried to tried to disrupt the process.	New York	2	P#1	d	d	NP	N	Q+N
DR1	newyork\newyork3	**a New York company** got in the way, bought some distressed Peruvian debt at a knockdown price	New York company	5	P0#10	d s	(d +SS)	CC/OPC	N	Q+N
DR1	newyork\newyork1	some aggressive behaviour by a firm of **New York investors**. The company,	New York investors	5	P0#010	d s	(d +SS)	CC/OPC	N	Q+N+N
DD2	oilseed\oilseed1	trials of **autumn-sown oilseed rape** is an embarrassment to ministers.	autumn-sown oilseed rape	6	P0#0#P0#1	d	(s)	CC	N	Q+N+N+N
DD2	oilseed\oilseed3	the trials of **genetically modified autumn sown oilseed rape** despite	genetically modified autumn-sown oilseed rape	14	0P00#P00#P0#0#P0#1	d	(s)	CC	N	Q+Q+N+N+N
DD2	oilseed\farmtri2	He stressed that the programme of farm trials of **genetically modified crops** would continue next spring	genetically modified crops	7	P00#P00#1	d		CC	N	adv+Q+N
DD2	oilseed\modif1	assess the impact of **modified crops** on the environment.	modified crops	4	P00#1	d		CC	N	Q+N
DD2	oilseed\oilseed2	permission to plant **spring-sown genetically modified oilseed rape** but officials	spring-sown genetically modified oilseed rape	13	P#0#0P00#P00#P0#1	d	(s)	CC	N	Q+Q+N+N+N
HP	room\charmsclass1	hurtled along it and came out near their **Charms classroom**, which they knew was (118)	Charms classroom	3	1#30	s s	*d s*	CC/OPC	N	N+N
HP	room\flykeyrm	get Ron - [208] grab brooms from the **flying-key room**, they'll get you out of the (208)	flying-key room	4	10#0#0	s s	*s d*	CC	N	Ving+N+N
HP	room\mirrorm1	that he might not be able to find the **mirror room** again. With Ron covered in the (154)	mirror room	3	10#0	s	*s*	CT	N	N+N

Source	Fichier son	Concordance	graphie	Nb syll	Schéma perçu	s/d perçu	s/d prévu	TypeCp	Cat gr	Formation
HP	room\staffroom1	He made his way down to the **staff room** and knocked. (134)	staff room	2	1#0	s	*s*	CT	N	N+N
HP	room\towaerrm1	up their wands and crept across the **tower room**, down the spiral staircase and (115)	tower room	3	10#0	s	*s*	CT	N	N+N
HP	room\trophy1	all right? We'll meet you in the **trophy room**, that's always unlocked.' When (114)	trophy room	2	10#0	s	*s*	CT	N	N+N
HP	room\trophy2	corner when they heard Filch enter the **trophy room**. 'They're in here somewhere,' (117)	trophy room	2	10#0	s	*s*	CT	N	N+N
HP	room\trophy3	which they knew was miles from the **trophy room**. 'I think we've lost him,' Harry (118)	trophy room	2	10#0	s	*s*	CT	N	N+N
DD5	room\washroom	The **washrooms** are disgusting	washrooms	2	1#0	s	s	CT	N	N/V+N
DD4	shift\ci_comp1	advertisements that are put out by **cigarette companies** frankly somewhat perplexing.	cigarette companies	6	P0#10	d		CT	N	N+N
DD1	shift\climat1	we're very obsessed about **climate change** on the Earth at the moment.	climate change	3	P0#1	d		CT	N	N+N
DD3	shift\clim1	the Kioto **climate change conference** two years ago	climate change conference	6	P0#0#10	d s		CC	N	N+N+N
DD1	shift\fifteen	that will last about **fifteen or sixteen minutes** and then it should er go into Mars orbit	fifteen or sixteen minutes	7	P#0#0#P#0#10	d s	(d/s)	CT	N	N+N
DR1	shift\hardear1	Peru successfully reached an agreement, but the vulture fund ended up collecting 58 million dollars of **hard-earned cash** in Peru.	hard-earned cash	3	P#P#1	d d	d s	CC	N	Q+N
DR1	shift\publici1	Alas, after several requests, **publicity-shy Elliots** still didn't want to grant us an interview.	publicity-shy Elliots	8	0P0#0#10	d s		CC/S	SN	Q+N
HP	shift\self1	Sizes - Copper, Brass, Pewter, Silver - **Self-Stirring**-Collapsible said a sign (56)	Self-Stirring	3	1#30	s		CO	Q	pron+Ving
HP	shift\self2	about batty old wizards who'd invented **self-stirring cauldrons** and they'd be free, (192)	self-stirring cauldrons	5	P#30#10	d s		CC	N	Q+N
DR1	shift\thirty1	And it resides on the **thirty-sixth floor** of this building.	thirty-sixth floor	4	P0#1#0	s d		CC/S	SN	Q+N
DD4	shift\underag2	There's other abuses of course **under-age**, of sex, alcohol and so on,	under-age	3	10#0	s	d	CT	Q	adv+N
DD4	shift\underag3	we're not allowed to target, quite rightly, **under-age smokers**, and tell me	under-age smokers	5	P0#0#10	d s	(d +SS)	CC	N	Q+N
DD4	shift\underag1	the terrible problem which we've got at the moment which is **under-age smoking**. There's other abuses	under-age smoking	5	P0#0#10	d s	(d +SS)	CC	N	Q+N

Source	Fichier son	Concordance	graphie	Nb syll	Schéma perçu	s/d perçu	s/d prévu	TypeCp	Cat gr	Formation
DR1	test\anyone1	The company won, and stopped Peru from paying **anyone anything** until it had got all its cash.	anyone	3	P0#0	s	s	CT	N	pron+N
DR1	test\anyone1	The company won, and stopped Peru from paying **anyone anything** until it had got all its cash.	anything	3	P0#0	s	s	CT	N	pron+N
DR1	test\birdsof	though the tactics of some of the vulture funds are threatening to give the **birds of prey** a bad name.	birds of prey	3	P#0#1	d	d	CS	N	N=prep+N
HP	test\hocuspo1	your little bit of **hocus pocus**. I'm waiting. (166)	hocus pocus	4	P0#10	d	d	CR	N	N+N
DD4	test\pillarb1	a rather morose person standing **by a phone by by a pillar box** em of a particular colour is going to say,	pillar box	3	10#0	s	s	CT	N	N+N
DD4	test\razorbl1	a piece of of a **razor blade** and a piece of silk in it	razor blade	3	10#0	s	s	CT	N	N+N
DD1	test\space1	the **spacecraft** really parked in orbit for quite a long time	spacecraft	2	1#0	s	s	CT	N	N+N
DD3	test\time1	it's getting that immense detail worked out and the **timetable** adopted that this meeting's all about, yes	timetable	3	1#30	s	s	CT	N	N+N
DD4	tobacco\adultsm1	you have to be sophisticated enough to be **an existing adult smoker** to understand	adult smoker	4	P0#10	d		CT	N	N/Q+N
DD4	tobacco\nonsmok1	I've got to say to you as a **non-smoker** that I find a lot of the advertisements	non-smoker	3	P#10	d	d	P	N	pref+N
HP	vodka\HParmcha3	Hagrid was sitting in an **armchair** outside his house (193)	armchair	2	P#1	d	d	CT	N	N+N
DR2	vodka\armchai1	everything in you home could be talking back to you: **armchairs**, tables, even the kitchen sink.	armchairs	2	P#1	d	s/d	CT	N	N+N
HP	vodka\HParmcha1	turning all the **armchairs** into hunched black shadows (115)	armchairs	2	1#0	s	s/d	CT	N	N+N
HP	vodka\HParmcha2	so they were able to get the good **armchairs** by the fire. (146)	armchairs	2	1#0	s	s/d	CT	N	N+N
DR2	vodka\groundb1	It's all part of a **ground breaking scheme** designed to give each and every object	ground-breaking scheme	4	P#30#1	d s	s	CC/S	N	Q+N
DR2	vodka\kitchen1	everything in your home could be talking back to you: armchairs, tables, even the **kitchen sink**.	kitchen sink	3	P0#1	d	d	CT	N	N+N
DR2	vodka\masspro1	a vodka bottle top which talks back to you. **Mass production** is due to begin next month	mass production	4	P#010	d	d	CT	N	N+N
DR2	vodka\microch1	It's packed with the very latest **microchip technology** but it's very simple to use.	microchip technology	7	P0#0#0100	d s	(s)	CT/QL	N	N+N
DR2	vodka\talkbot2	The **talking bottle top** was developed by scientists at this workshop in Moscow	talking bottle top	5	P0#10#0	s s d		CC/S	N	Q+N+N
DR2	vodka\talkbot3	you wouldn't think there was a need for a **talking bottle top** to convince people to drink it.	talking bottle top	5	P0#P0#1	d d d		CC/S	N	Q+N+N

Source	Fichier son	Concordance	graphie	Nb syll	Schéma perçu	s/d perçu	s/d prévu	TypeCp	Cat gr	Formation
DR2	vodka\talkbot1	It's the world's first ever talking vodka bottle top. The voice	talking vodka bottle top	7	P0#P0#P0#1	d d d d		CC/S	N	Q+N+N+N
DR2	vodka\vodbott1	They've designed a vodka bottle top which talks back to you.	vodka bottle top	5	P0#P0#1	d d d		CC/S	N	N+N
DR2	vodka\vodgeni1	The voice belongs to the vodka genie who recites a whole string of popular Russian toasts	vodka genie	4	10#30	s		CT	N	N+N
DR2	vodka\vodgeni2	Each time you take the top off the bottle the vodka genie's speech grows more and more slurred.	vodka genie	4	10#30	s		CT	N	N+N
DR2	vodka\vodgen3	But even that's not enough to silence the vodka genie.	vodka genie	4	P0#10	d		CT	N	N+N
DR2	vodka\worksh1	The talking bottle top was developed by scientists at this workshop in Moscow.	workshop	2	1#0	s	s	CT	N	N/V+N

UNIVERSITÉ DE POITIERS

Faculté des Lettres et Langues

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE POITIERS

LES MOTS COMPOSÉS :

ANALYSE DE SCHÉMAS ACCENTUELS DE L'ANGLAIS BRITANNIQUE STANDARD

(deux volumes, accompagnée d'un CD-ROM : VOLUME 2)

présentée et soutenue publiquement par

Susan MAUROUX

le 14 décembre 2002

sous la direction de

Monsieur Jean-Louis DUCHET

Professeur à l'université de POITIERS

JURY

Monsieur Nicolas BALLIER, Maître de conférences à l'université de Rouen

Monsieur Jean-Louis DUCHET, Professeur à l'université de Poitiers

Madame Ruth HUART, Professeur à l'université de Lille III

Monsieur Michel PAILLARD, Professeur à l'université de Poitiers

Monsieur Peter ROACH, Professeur à l'université de Reading, Angleterre

CHAPITRE 3

Chapitre 3 : Méthodes d'analyse du corpus

Nous avons examiné de quelle manière le corpus a été constitué dans le chapitre précédent. Nous allons à présent aborder la question des méthodes d'analyse que nous adoptons pour travailler sur ce corpus. Il y a deux grands axes de recherche ; tous deux reposent sur l'utilisation d'outils informatiques. Nous commençons par l'analyse acoustique qui servira pour tester des extraits des sources orales, et qui nous permettra d'étudier le comportement accentuel des composés en discours. Certains résultats de cette recherche seront incorporés dans une base de données sous Access[1], qui sera utilisée pour mettre en relation les deux types de source, et pour étudier plus particulièrement les sources écrites. L'utilisation de la base de données est étudiée dans la deuxième partie de ce chapitre.

3.1 Analyse acoustique

Nous avons déjà expliqué comment nous avons restreint le corpus oral d'analyse et pour quelles raisons. L'analyse acoustique suppose un travail minutieux de préparation des données et deux questions préalables se posent : quelles sont les mesures pertinentes pour notre analyse ? Et quelles sont les unités pour lesquelles nous devons disposer de mesures ? La résolution de ces questions nous permettra de choisir les outils les plus adaptés à l'analyse des extraits sonores du corpus. Nous allons d'abord nous intéresser aux corrélats acoustiques de l'accent.

---

[1] Base de données du pack Office Pro de Microsoft.

3.1.1 Les corrélats acoustiques de l'accent

Nous commençons par citer Ladefoged (Ladefoged, 1962 : 13) qui explique de quelle manière les sons peuvent se distinguer:

> These three factors - loudness, pitch and quality - provide the most convenient method of differentiating between all sounds. They can be regarded as three ways in which sounds can differ. Whenever you hear two sounds it is possible to describe the differences between them by comparing them in these three ways.

Comme nous l'avons vu dans le premier chapitre où nous avons étudié la proéminence (voir 1.1.1, « L'accent et la proéminence : un ensemble de facteurs »), ces trois facteurs sont également des paramètres essentiels de la perception de l'accent. Il faut y ajouter la longueur. Du point de vue acoustique, quels sont les corrélats de ces facteurs psychologiques ? Fry (Fry, 1958 : 126) nous explique :

> Differences of stress are perceived by the listener as variations in a complex pattern bounded by four psychological dimensions: length, loudness, pitch and quality. The physical correlates of these perceptual factors are the duration, intensity, fundamental frequency and formant structure of the speech sound waves

Encore aujourd'hui, ce sont ces quatre paramètres acoustiques qui sont considérés comme étant significatifs dans la perception de l'accent : la durée (corrélat de la longueur), l'intensité (corrélat de la force, « loudness » en anglais) la fréquence fondamentale (corrélat de la hauteur mélodique, « pitch » en anglais) et les formants (corrélat de la qualité vocalique). L'importance relative de ces quatre paramètres a été étudiée par différents chercheurs depuis Fry, qui a fait des expériences sur des paires dissyllabiques « nom » / « verbe » de l'anglais qui se distinguent par leur schéma accentuel. Ses conclusions (Fry, 1955 & 1958) tendent à démontrer surtout la pertinence de la fréquence fondamentale, et aussi celle de la durée et de l'intensité mais à un degré moindre. Nous allons étudier ces quatre paramètres acoustiques, afin de déterminer quelle place leur attribuer dans

notre analyse du corpus oral, et comment interpréter les mesures qui y sont associées.

3.1.1.1 La fréquence fondamentale

La fréquence fondamentale est la mesure acoustique qui correspond à la hauteur mélodique, appelée « pitch » en anglais. Cette fréquence représente le nombre de pics par seconde ; elle se mesure en hertz ce qui représente une valeur absolue, ou bien en demi-tons, ce qui permet de disposer d'une échelle de valeurs relatives. Ce paramètre acoustique est celui qui est le plus souvent cité en relation avec l'accent ; c'est pourquoi nous commençons notre étude des paramètres acoustiques par la fréquence fondamentale (aussi appelée F_0[2]).

3.1.1.1.1 La relation entre la fréquence fondamentale (F_0) et l'accent

Dès Fry (1955 & 1958), l'importance de ce paramètre est reconnue. Ses conclusions sont corroborées par les travaux de Lieberman qui reprend les mêmes paires de mots que Fry en contexte. Ses conclusions (Lieberman, 1960 : 453) sont en accord avec celles de Fry :

> The data indicate that higher fundamental frequencies and envelope amplitudes are the most relevant of the unidimensional acoustic correlates of stressed syllables; the fundamental frequency seems most relevant ...

Beckman fait une étude comparative de l'anglais et du japonais : elle prend les mêmes paires de mots que Fry pour l'anglais et des mots comparables du point de vue de leur structure syllabique pour le japonais. Elle a fait deux expériences, l'une de production, l'autre de perception. Elle arrive aux mêmes conclusions : la

---

[2] Il y a plusieurs abréviations pour la fréquence fondamentale : *F0* (Terken, 1991), F0 (Dujardin, 2001) ; F_0 (Gussenhoven & Rietveld, 1988 ; Beckman & Edwards, 1994 ; Diana, 1994).

fréquence fondamentale, l'intensité et la durée sont tous des paramètres importants; mais en position nucléaire, c'est la fréquence fondamentale qui est déterminante. D'autres expériences (Brown & McGlone, 1974 et Terken, 1991) corroborent ces résultats.

Les propos de Sluitjer et van Heuven (1993 : 246) résument les travaux précédents :

> The acoustical correlates of stress are pitch, duration, loudness and vowel quality. Of these, pitch and duration have been found the most important perceptual cues; intensity and vowel quality are generally claimed to be of lesser importance.

Ils expliquent qu'un mouvement mélodique est décisif dans la perception de l'accent (Sluitjer & van Heuven, 1993 : 249) : « [...] pitch movement will always be the overriding cue in pitch perception ». Leurs expériences ont donc tenu compte de ce fait en étudiant la perception de l'accent dans les deux types de contextes : là où l'accent correspond effectivement à un mouvement mélodique, et là où il n'y correspond pas. Ils démontrent ainsi la signification du paramètre de la fréquence fondamentale, mais aussi la signification d'autres paramètres. Ils expliquent leur hypothèse puis leur démarche (Sluitjer & van Heuven, 1993 : 246 & 249) :

> Indeed, an appropriate pitch movement provides an overriding cue, but is present only when the stress coincides with an accent (marking focus). When words are spoken outside focus, the position of the stress has to be inferred from the remaining cues.
>
> Stimuli will be presented outside focus (without a pitch movement on the target) because of the fact that a pitch movement will always be the overriding cue in pitch perception.

Dujardin (2001) souligne l'importance de l'intensité et de la durée dans la perception de l'emphase mais rappelle que sur des mots courts, c'est la fréquence fondamentale qui sert pour la marquer. L'importance de la fréquence

fondamentale a donc été clairement démontrée, mais quels sont les facteurs dont il faut tenir compte pour interpréter la portée de ce paramètre acoustique ? Nous abordons cette question sous la rubrique suivante.

3.1.1.1.2 Comment interpréter la fréquence fondamentale

Nous devons tenir compte de plusieurs paramètres que nous avons classés en trois catégories : ceux qui sont constants et dont on tiendra compte à chaque analyse ; ceux qui peuvent troubler l'interprétation des données, que nous avons appelés « paramètres contextuels » ; et enfin la déclinaison[3] de la fréquence fondamentale sur l'ensemble de l'unité tonale et qu'il faut absolument intégrer dans l'analyse des résultats.

3.1.1.1.2.1 Paramètres physiologiques

La fréquence fondamentale varie selon le type de voyelle : Gray (Gray, 2001 : 29) rappelle : « Les voyelles fermées ont tendance à être réalisées sur un niveau de F_0 plus haut que les voyelles ouvertes ». Dans ce cas, Beckman (Beckman, 1986 : 129) parle de « intrinsic F_0 », mais elle nous met en garde, car les effets de « intrinsic pitch », le corrélat psychologique de la fréquence fondamentale, est inverse.

De même, comme le rappelle Dujardin (2001 : 66), la voix d'une femme est généralement plus élevée que celle d'un homme, donc la proéminence correspondant à une plus haute fréquence fondamentale est moins perçue, car elle est interprétée comme étant normale et non exceptionnelle. Ces ajustements se

---

[3] Nous avons emprunté ce terme à Dujardin (2001). Il correspond à « declination » en anglais.

reposent sur des facteurs psychologiques qui sont en relation avec les attentes de l'interlocuteur, et qui sont difficilement mesurables. Nous devons en tenir compte dans nos analyses.

3.1.1.1.2.2 Paramètres contextuels

Beckman (Beckman, 1986 : 126) nous met en garde contre les effets des segments environnant la voyelle qui peuvent rehausser ou rabaisser la valeur de la fréquence fondamentale de sorte que la corrélation avec la hauteur mélodique est faussée : « breathy voiced consonants sharply depress the F_0 of adjacent vowels ». De manière générale, les occlusives voisées tendent à faire baisser la fréquence fondamentale de la voyelle qui suit, alors que les occlusives non voisées ont l'effet inverse. La différence de la fréquence fondamentale ainsi affectée peut être de quatre demi-tons, et la valeur peut ainsi être modifiée sur une assez grande partie de la voyelle qui suit le segment.

Dujardin (2001 : 120-122) explique la nécessité de faire des corrections manuelles lorsque l'on travaille sur PRAAT car la courbe peut être faussée par plusieurs facteurs : des sons non voisés, le r, les coups de glotte, une voix grinçante ou trop basse[4].

La perception de la modulation du F_0 peut aussi être affectée par le contour intonatif. Dans l'interprétation d'un énoncé, le locuteur et l'interlocuteur font des ajustements automatiques en fonction de ces facteurs. Gray rappelle la difficulté

---

[4] Pour vérifier qu'il s'agit bien d'une erreur, on peut également se référer aux formants du spectrogramme pour s'assurer que la fréquence des harmoniques ne correspond pas à la valeur calculée de F_0. Dans ce cas, le calcul de F_0 est effectivement erroné. Cette démarche a été suggérée par G. Boulakia lors de la journée « Logiciels de la parole » à Poitiers, 24 mai 2002.

pour les francophones à distinguer la syllabe accentuée lorsqu'elle se trouve sur un ton montant (Gray, 2001 : 124) : « Il ressort clairement de nos analyses que l'intonation montante rend difficile la tâche, apparemment simple, qui consiste à identifier la syllabe accentuée dans un corpus de dissyllabes ».

Enfin, Beckman (1986: 109) affirme que la durée influence également la perception de la hauteur mélodique : « This probable effect of duration on frequency discrimination should be borne in mind whenever the significance of a small difference in fundamental frequency is being evaluated ». Nous prenons conscience de la complexité de l'interprétation des données acoustiques car les paramètres sont souvent corrélés[5].

3.1.1.1.2.3 Déclinaison de F_0

Le dernier facteur dont il faut tenir compte dans une analyse de la fréquence fondamentale est ce que l'on appelle en anglais « F_0 declination », c'est-à-dire la tendance de la fréquence fondamentale à baisser à la fin de l'unité tonale. Ce phénomène a été remarqué par plusieurs chercheurs (Pierrehumbert, 1979 ; Liberman and Pierrehumbert, 1984 ; Gussenhoven and Rietveld, 1988 ; Diana, 1994). Terken (1991 : 1768) résume certaines de leurs découvertes :

> It was found that pitch was not a direct function of *F*0, but that the relation between *F*0 and pitch was mediated by position in the utterance: Of two equal *F*0 maxima, the one later in the utterance was perceived as having higher pitch. The same was demonstrated for prominence judgments by Gussenhoven and Rietveld (1988). These findings were accounted for with reference to *F*0 declination; this is the phenomenon that "the range of *F*0 values employed is narrower and lower at the end of a phrase than at the beginning" (Liberman and Pierrehumbert, 1984, p. 161). Thus it appears that listeners take into account expectations about the narrowing of the *F*0 range and compensate for it when making judgments about pitch and prominence.

---

[5] Voir aussi les propos de Ladefoged, 2001: 61-3.

> listeners take into account expectations about the narrowing of the *F*0 range and compensate for it when making judgments about pitch and prominence.

Nous constatons à nouveau l'ajustement automatique que fait l'interlocuteur dans son interprétation de l'accent. Les expériences de Terken lui permettent de tirer les mêmes conclusions (Terken, 1991 : 1771) : « under both instructions, the adjustments appear to be influenced by expectations about *F*0 declination in the course of the phrase ». Il faut noter que l'interlocuteur aura davantage tendance à faire des ajustements, lorsque la variation de la fréquence est importante (Terken, 1991 : 1773) :

> The present results confirm the findings for experiment 1: Listeners compensate for expected declination of *F*0 maxima in making judgments about pitch and prominence, and this tendency is stronger for higher P1 values

Dans cette section, nous avons constaté l'importance de la fréquence fondamentale comme paramètre acoustique et nous avons expliqué quels sont les facteurs dont il faut tenir compte pour l'interpréter : le contexte segmental, le contour intonatif et le type de voix. Toutefois, même si la fréquence fondamentale paraît être le paramètre acoustique le plus révélateur, du moins en position nucléaire, il n'est pas le seul qui ait de l'importance dans la perception de l'accent. Nous passons à l'étude d'un autre paramètre acoustique qui a souvent été cité avec la fréquence fondamentale : l'intensité.

3.1.1.2 L'intensité

3.1.1.2.1 La relation entre l'intensité et l'accent

Comme nous l'avons vu au début de ce chapitre, un des facteurs psychologiques de la perception de l'accent est la force, appelé « loudness » en anglais. En termes d'acoustique, la force se mesure par rapport à l'intensité. Sa valeur est toujours donnée en décibels (dB).

Ce paramètre a longtemps été considéré comme le facteur le plus révélateur de la perception de l'accent, jusqu'aux travaux de Fry (1955 & 1958), qui démontrent que la fréquence fondamentale est encore plus importante. D'autres travaux tentent de chercher le rôle précis de l'intensité dans la perception de l'accent, notamment dans des cas où la fréquence fondamentale est moins révélatrice[6], lorsque la syllabe accentuée n'est pas en position nucléaire (Beckman, 1986; Sluitjer & van Heuven, 1993). Les résultats de ces expériences démontrent à la fois l'importance de ce paramètre et les difficultés d'interprétation des mesures qui lui sont associées.

3.1.1.2.2 Comment interpréter l'intensité

La correspondance même entre l'intensité et la force est remise en question par Beckman (Beckman, 1986: 107): « It is probably less erroneous to equate hertz or semitones with pitch than it is to equate decibels with loudness ». Le calcul de la valeur de l'intensité est très influencé par les conditions d'enregistrement. Le moindre bruit de fond ou changement de position par rapport au micro entraînera nécessairement des valeurs basées sur des calculs différents. En effet, une analyse acoustique se doit de prendre des mesures constantes où le calcul se fait de la même façon, afin que les comparaisons puissent être pertinentes[7].

---

[6] Voir les propos de Ladd (1993: 11): « Once we see beyond the overwhelming effects of pich accent on stress judgements, we realize that duration and intensity are considerably more important than Fry's work make it appear ».

[7] Philippe Carré (IRCOM-SIC, université de Poitiers), spécialiste du traitement du signal, a fait cette remarque lors de la journée « Logiciels de la parole » à Poitiers, 24 mai 2002. Il commentait l'efficacité relative des différents types de calcul que propose PRAAT et expliquait que chaque type de calcul est un compromis entre la précision des valeurs du temps et celles de la fréquence

Blevins insiste sur la valeur relative de l'intensité par rapport à d'autres facteurs. Elle explique la théorie de la sonorité de la syllabe, et elle distingue justement dans son propos la force («loudness») de la mesure acoustique de l'intensité parce que la perception est influencee par la hauteur de la fréquence fondamentale et que les valeurs ne correspondent pas. Cette distinction est expliquée dans la note de bas de page (Goldsmith, 1995 : 234) :

> Sonority is defined here in terms of a perceptual measure, loudness, not in terms of the acoustic property of intensity. Though these will often coincide, there are cases where they do not. For instance, a vowel with a fundamental frequency of 500 hertz will be judged by the human ear as louder than an otherwise identical vowel of the same intensity at 5,000 hertz. Here and elsewhere, the human ear imposes a sonority ranking which is not present in the acoustic signal.

Beckman (1986) fait état des mêmes constats en indiquant qu'il y a une bande critique et qu'en dehors de cette fourchette les valeurs varient beaucoup. C'est-à-dire que l'oreille perçoit les variations d'intensité de manière à peu près stable entre 2000 Hz et 5000 Hz, mais en dehors de cette bande, elles sont moins perçues.[8].

Les valeurs pour l'intensité qui sont données par PRAAT se situent entre 50 et 100 dB. Yoo[9] émet des réserves par rapport à la validité de ces mesures en tant que valeur absolue. Toutefois, si nous considérons que ces valeurs sont des valeurs relatives, nous pouvons très bien les comparer entre elles. Beckman arrive à la conclusion que l'intensité a une grande importance dans la perception de la

---

ou de l'intensité. Il conseille donc au chercheur d'en sélectionner un et de s'y tenir afin que les comparaisons soient faites à partir des mêmes bases.

[8] Voir Johnson (1997 : 54). Ce phénomène rejoint la question du taux d'échantillonnage optimal que nous avons soulevée dans le chapitre 2 (2.3.3.1 « Informations techniques ».

[9] Journée « Logiciels de la parole », Poitiers, 24 mai 2002. Yoo, étudiante en thèse à Paris VII sous la direction de G. Boulakia, dit que ces mesures posent problème. Est-ce le mode de calcul lui-même qui serait inapproprié ?

syllabe accentuée en tant que valeur relative (Beckman, 1986 : 200) : « Here I make the further proposal that the defining feature is relative loudness, as reflected not in the peak amplitude level but rather in varying quantities of total amplitude ».

De même, les travaux de Sluitjer & van Heuven (1993) s'appuient sur l'hypothèse de la relativité des valeurs de l'intensité. Ils pensent que lorsque l'on manipule le signal, il faut faire varier l'intensité, non en termes d'absolu en élevant ou en abaissant les valeurs de façon uniforme, mais en y appliquant un coefficient en fonction de la hauteur de la fréquence pour les faire varier. Ils appellent cette manipulation « realistic intensity manipulations (i.e. concentrated in the higher frequency bands) » (Sluitjer et van Heuven, 1993: 249). La perception de la force serait donc influencée par la hauteur de la fréquence fondamentale.

L'association de l'intensité avec un autre facteur a aussi été constatée par Beckman. D'après ses expériences, l'intensité (donc la perception de la force) serait associée à la durée[10]. Elle explique la signification de la relation entre les deux (Beckman, 1986 : 139) :

> This relationship may be crucial to understanding the physical correlates of stress accent. It may be the source of the correlation between duration and stress, and it may explain why stress-detecting algorithms that use the intensity integral have vastly better success than do algorithms that rely on peak intensity levels.

[10] Voir aussi les expériences de Turk & Sawusch (1996) qui étudient la relation entre ces deux paramètres, et ils concluent qu'ils sont corrélés, mais que la durée reste le paramètre le plus important.

Nous allons à présent aborder ce troisième paramètre acoustique de la perception de l'accent : la durée.

3.1.1.3 La durée

3.1.1.3.1 La durée et l'accent

La durée est fréquemment citée comme étant un facteur associé à la perception de l'accent (Fry, 1955 & 1958 ; Lieberman, 1960 ; Beckman, 1985). Pour Fry, ses expériences de 1955 et de 1958 montrent que c'est un paramètre plus efficace que celui de l'intensité. Une syllabe accentuée dure plus longtemps qu'une syllabe inaccentuée. Comme l'explique Ladefoged (Ladefoged, 1993: 95) : «Stressed syllables are longer than the corresponding unstressed syllables»[11]. Nous avons également vu qu'elle est citée comme un facteur qui influence le F_0 (Beckman, 1986).

3.1.1.3.1 Comment interpréter la durée

Dans l'interprétation de la durée, nous devons tenir compte de deux facteurs. D'abord, la durée moyenne d'un segment est variable. Beckman nous rappelle que les voyelles basses sont plus longues que les voyelles hautes (Beckman, 1986 : 141) : « Other things being equal, a low vowel will be longer than a high vowel » et elle appelle ce phénomène « intrinsic duration » que nous pouvons rapprocher du concept de « intrinsic frequency » auquel nous avons fait référence sous la première rubrique (3.1.1.1.2.1).

---

[11] Voir aussi les travaux de Beckman & Edwards (1994 : 30) : « If a syllable has a full vowel specification as the head of its stress foot, it will necessarily be much longer than a syllable with a reduced vowel, and will have a much larger, faster opening movement ».

Le second phénomène concerne le rythme général de l'anglais que définit Ladefoged (Ladefoged, 1993: 95) : « [...] general tendency in English to equalize the lengths of syllables that differ in the number of segments they contain». Cette tendance remet en cause une quelconque valeur absolue de la durée d'un segment et nous oblige à considérer la durée en tant que valeur relative. Ce phénomène d'égalisation doit également se comprendre par rapport au rythme de la langue qui fait alterner les syllabes accentuées et les syllabes inaccentuées dans un contexte de discours.

3.1.1.4 La qualité vocalique

3.1.1.4.1 La qualité vocalique et l'accent

La qualité vocalique est une des manières dont on distingue les sons, comme nous l'avons vu dans l'introduction à cette partie. En ce qui concerne l'accent, une voyelle réduite signale forcément une syllabe inaccentuée ; par contre, une voyelle pleine est une condition nécessaire mais non suffisante pour qu'une syllabe soit accentuée. Ainsi, toutes les syllabes accentuées contiennent une voyelle pleine, mais un nombre significatif de syllabes inaccentuées en ont une aussi.

3.1.1.4.2 Comment interpréter la qualité vocalique

La qualité vocalique est représentée par la structure des formants de la voyelle. Le schwa d'une syllabe inaccentuée est donc reconnaissable d'après ses formants ; nous pourrons ainsi identifier les syllabes inaccentuées ayant la voyelle réduite schwa. Toutefois, dans le cas de /ɪ/, les formants seront identiques qu'il s'agisse de la voyelle réduite [ɪ] d'une syllabe inaccentuée (de *forces* ou *defence*,

par exemple) ou de la voyelle pleine et accentuée [ɪ] (de *wizard*). La portée de ce paramètre seul est donc limitée.

Nous sommes également réservée quant à l'utilisation de ce paramètre dans notre analyse, car une des caractéristiques des composés à accentuation simple est la non-réduction de la voyelle du deuxième élément monosyllabique. Ainsi, il est rare que la syllabe inaccentuée corresponde à une voyelle réduite.

3.1.1.5 Un ensemble de paramètres

Il apparaît clairement que la perception de l'accent doit être attribuée non à un seul mais à un ensemble de paramètres qui ont chacun plus ou moins de portée selon le contexte. D'après Beckman, c'est justement le fait d'interpréter plusieurs facteurs qui distingue les langues à accentuation comme l'anglais des langues comme le japonais (Beckman, 1986 : 145) : « [...] the stress-accent hypothesis which posits that stress-accent languages use phonetic attributes other than pitch to a greater extent than do non-stress-accent languages ». Cet avis est partagé par d'autres chercheurs (Fry, 1955 & 1958 ; Lieberman, 1960 ; Ladefoged 1971 ; Sluitjer & van Heuven, 1993 ; Diana, 1994).

Comment ces paramètres peuvent-ils s'interpréter ? Bucellato reprend les paramètres de la fréquence et de l'intensité et insiste sur leur importance, mais ses conclusions montrent que l'analyse nécessite une prise en compte très fine des données (Bucellato, 1985 : 282) : « Il apparaît que ce sont les changements de niveau de la fréquence et de l'intensité qui engendrent l'impression d'accent plutôt que le niveau élevé de l'une ou de l'autre ». Il faudra donc tenir compte de

cet ensemble de données et de l'interaction des différents paramètres acoustiques dans notre analyse des extraits sonores du corpus.

Pour conclure cette partie sur les corrélats acoustiques de l'accent, nous citons les propos de Gussenhoven & Jacobs qui résument la discussion de ces différents paramètres (Gussenhoven & Jacobs, 1998 : 206), et ouvrent le débat sur un autre domaine :

> While at first, loudness was taken to be the relevant phonetic parameter, experiments by Fry (1955; 1958) showed that a number of different parameters were involved in creating the impression of stress in English, specifically, duration, pitch variation and vowel quality. However, beginning with Liberman & Prince (1977), it was realized that stress was not a phonological feature that was given some content by the phonetic implementation rules, as assumed in the linear theory of *SPE*, but a structural position. The structural position is the foot which is a phonological constituent above the syllable and below the word.

Dans la section suivante, nous allons aborder ces notions et déterminer si, en effet, le pied est l'unité la plus appropriée à notre analyse, ou bien si nous préférons rester au niveau de la syllabe afin d'effectuer les comparaisons entre les extraits sonores.

3.1.2 Les unités de comparaison

Il faut d'abord définir la syllabe, car elle se trouve à un niveau inférieur au pied ; nous commençons donc notre étude de ces deux unités par elle. Dans cette section, nous tenterons d'apprécier la signification des unités en question, puis nous étudierons leur pertinence pour notre analyse à partir d'exemples précis.

3.1.2.1 La syllabe

Ladefoged (Ladefoged, 1971 : 83) souligne l'importance de la syllabe dans le rythme naturel de l'anglais en faisant référence aux travaux d'Allen (Allen,

1971) qui a mené des expériences afin de déterminer l'emplacement exact des battements rythmiques de l'anglais et leur perception :

> Like syllabicity, stress is an organizing principle. In many languages the timing of utterances is organized at least partly in terms of stress groups. As has been shown by Allen (1971) syllables and stresses are the only units involved in the rhythm of English phrases.

Toutefois, il résume bien la difficulté qui existe à la définir (Ladefoged, 1993 : 47) : « Although nearly everybody can identify syllables, almost nobody can define them ». Leur existence, comme celle du mot, a une réalité psychologique, démontrée par les nombreux jeux de langue qui reposent sur cette unité : *eg language*, par exemple, qui consiste à rajouter « eg » devant chaque voyelle, donc « I go » se dit « egI gego » (voir aussi Hammond, 1999).

Jensen insiste sur l'intérêt fondamental de cette unité (Jensen, 1993 : 47) : « The syllable is the basic unit of pronunciation [...] speakers are generally aware of how many syllables there are in words of their language [...] ». Comme nous l'avons vu dans le premier chapitre, c'est la syllabe qui est associée à l'accent. Fry précise (Fry, 1958 : 127) :

> Hence stress is a term that refers to a relation between syllables and successive variations in this relation constitute the rhythmic pattern of an utterance just as successive variations in tone-relations make up the intonation pattern.

Il résume en partie la définition de l'accent que nous avions donnée dans le premier chapitre : il s'agit d'une syllabe proéminente, et cette proéminence est ressentie comme étant une valeur relative ; les syllabes sont perçues comme étant proéminentes les unes par rapport aux autres. Cette définition de la syllabe n'est

pas la seule[12], mais elle montre bien le lien entre cette unité et l'accent, qui est notre propos. Sous la rubrique suivante, nous aborderons certaines approches théoriques de la syllabe afin de comprendre les enjeux en ce qui concerne la coupe syllabique.

3.1.2.1.1 Application à l'analyse des composés

Les langues ont des structures syllabiques différentes, et la manière dont ces syllabes sont structurées est expliquée par différentes théories que résume Ladefoged (Ladefoged, 1993 : 248) :

> In summary, we can say that there are two types of theories attempting to define syllables. First, there are those in which the definitions are in terms of properties of sounds, such as sonority (acoustic energy) or prominence (some combination of sonority, length, stress and pitch). Second, there are theories based on definitions that are in terms of activities of the speaker, principally the notion that a syllable is a unit in the organisation of the sounds of an utterance.

Blevin (Goldsmith, 1995 : 206-7) souscrit plutôt à la première de ces théories. Elle explique en quoi consiste la théorie de la sonorité :

> What are syllables? [...] syllables can be viewed as the structural units providing melodic organization to such strings. This melodic organization is based for the most part on the inherent sonority of phonological segments, where the sonority of a sound is roughly defined as its loudness relative to other sounds produced with the same input energy (i.e., with the same length, stress, pitch, velocity of airflow, muscular tension, etc.).[13] Hence, melodic organization of a phonological string into syllables will result in a characteristic sonority profile: segments will be organized into rising and falling sonority sequences, with each sonority peak defining a unique syllable. The syllable then is the phonological unit which organizes segmental melodies in terms of sonority; syllabic segments are equivalent to sonority peaks within these organizational units.

Il reste très difficile d'établir une échelle de sonorité complètement satisfaisante, mais cette théorie permet d'expliquer pourquoi l'on hésite sur le

---

[12] Pour une discussion générale de la syllabe, voir, entre autres, Crépin (1982), Goldsmith (1990 : 208), Giegerich (1992 : 132-170), Ladefoged (1993 : 246-249), Blevin in Goldsmith (1995). Pour une discussion de la signification des caractéristiques phonétiques par rapport à la coupe syllabique, voir Turk (1994) et Nolan (1994).

[13] Renvoi dans le texte à la note que nous avons déjà citée pour expliquer les problèmes liés à la définition de l'intensité (voir 3.1.1.2).

nombre de syllabes dans certains mots (*prism*, par exemple), et surtout de déterminer la coupe syllabique dans un mot comme *helmet*, qui se coupera *helm.et* selon ce principe. Certains chercheurs, dont Ladefoged (1993 : 249) ont proposé le principe d'ambisyllabicité, donc dans *happy*, /p/ est à la fois rattaché à la première et à la deuxième syllabe.

Pour notre analyse, nous serons amenée à comparer des syllabes dans des extraits sonores différents : la coupe syllabique est donc importante et il faut adopter une démarche cohérente afin que les comparaisons soient pertinentes. Nous citons les propos de Wells dans la première édition de son dictionnaire (Wells, 1990 : xx) ; il résume les différentes approches possibles, puis propose la sienne, qui s'apparente à la deuxième approche théorique citée par Ladefoged :

> i A syllable boundary is found wherever there is a word boundary, and also coincides with the morphological boundary between elements in a compound.
>
> ii Affricates cannot be split.
>
> iii Where (i), (ii), and the phonotactic constraint will allow, consonants are syllabified with whichever of the two adjacent vowels is more strongly stressed; or, if they are equally stressed, with the leftward one.

Wells explique que cette théorie prédit correctement les allophones, ce qui justifie son choix de coupe syllabique dans le dictionnaire de prononciation. Il est intéressant de noter que, dans la deuxième édition, il ne parle plus des autres théories, mais explique la coupe syllabique en termes d'allophones uniquement. Il cite, à titre d'exemple, des phénomènes tels que l'aspiration : le /t/ de *invi.tation* débute la syllabe accentuée (il est aspiré) alors que le /t/ de *invit.ed* fait partie de la coda (il n'est pas aspiré).

La démarche de Wells nous intéresse plus particulièrement car le dictionnaire LPD de 1990 est notre principale source écrite. En ce qui concerne

les composés, la coupe syllabique se fera toujours à la frontière morphologique entre les éléments de composition. Dans le cas de composés dont la graphie indique la frontière des éléments, comme *trophy room*, la coupe se fera évidemment au niveau de l'espace graphique : *trophy.room*. Pour *flying-key room*, la coupe se fera au niveau de l'espace graphique et du trait d'union : *flying.key.room*. Ensuite, pour les éléments plurisyllabiques, la coupe se fera selon les critères cités plus haut : *fly.ing* pour des raisons morphologiques, *troph.y* pour la raison qui est donné en (iii) (la consonne est associée à la syllabe accentuée qui précède dans ce cas).

L'important dans une analyse des composés est de disposer d'une segmentation en unités qui permet de comparer les extraits sonores entre eux. Il n'est donc pas primordial de savoir si *trophy* doit être segmenté *tro.phy* ou *troph.y*. A l'inverse, il est nécessaire de s'assurer que l'élément *trophy* sera coupé de la même façon à chaque fois qu'il apparaît, ainsi que les éléments présentant des similitudes et que nous sommes susceptible de vouloir comparer.

3.1.2.2 Le pied

L'autre unité rythmique que nous pourrions utiliser pour notre analyse est celle du pied. Abercrombie en donne une définition très claire (1964 : 217) :

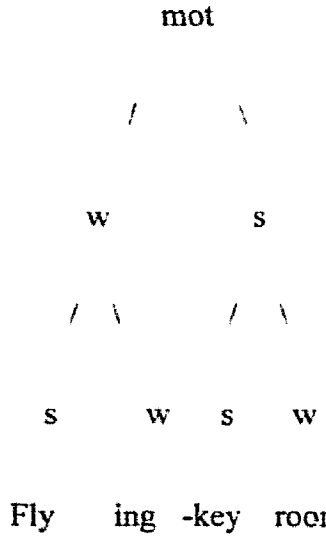
> English utterances may be considered as being divided by the isochronous beat of the stress pulse into feet of (approximately) even length. Each foot starts with a stress and contains everything that follows that stress up to, but not including, the next stress.

Cette unité rythmique est la base de l'analyse de Gussenhoven & Jacobs (1998) ainsi que de celle de Liberman & Prince (1977), et de tous les chercheurs qui s'inscrivent au sein de la phonologie métrique (Goldsmith, 1995).

3.1.2.2.1 Application à l'analyse des composés

Contrairement à la syllabe, les approches théoriques concernant le pied concordent. A partir de la définition d'Abercrombie (1964), nous allons donc pouvoir étudier comment appliquer cette unité dans notre analyse. Pour illustrer de quelle manière elle fonctionne, nous prenons un des exemples de notre corpus oral que nous avons présenté pour expliquer la coupe syllabique : *flying-key room*. Le tableau représente un arbre métrique, appelé « metrical tree » en anglais ; il montre le rythme du composé et fait ressortir le lien entre les éléments.

Figure 3.1 : Représentation de *flying-key room* : arbre métrique



Cette segmentation est forcément binaire et le composé comporte deux pieds : *flying* puis *key room*. L'autre type de représentation qu'utilise la phonologie métrique est la grille, appelée « metrical grid » en anglais. A la différence de l'arbre, la représentation n'est pas forcément binaire et a l'avantage de montrer des cas de conflits accentuels (là où il y a des croix consécutives sur le niveau 1), comme nous l'avons vu dans le chapitre 1 (1.5.2, « La phonologie

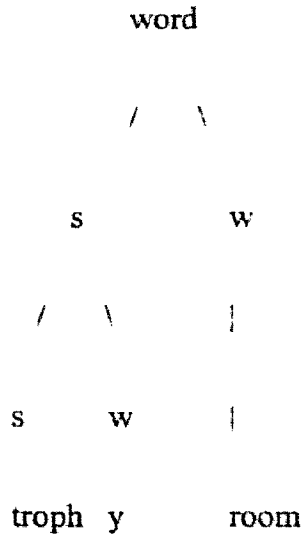
métrique »). Toutefois, le lien entre les éléments est perdu. Nous reprenons le même exemple sous cette forme.

Figure 3.2 : Représentation de *flying-key room* : grille métrique

		X		Niveau 2
X		X		Niveau 1
X	X	X	X	Niveau 0
fly	ing	key	room	

Nous prenons un second exemple où la segmentation binaire du composé est impossible : *trophy room*

Figure 3.3 : Representation de *trophy room*: arbre métrique



L'arbre montre bien la relation entre les éléments. Cette fois-ci, le composé comporte un seul pied, car seule la première des trois syllabes est accentuée. Le même exemple sous forme de grille montre à nouveau la hiérarchie accentuelle.

Figure 3.4 : Representation de *trophy room*: grille métrique

X			Niveau 2
X			Niveau 1
X	X	X	Niveau 0
troph	y	room	

Ces exemples peuvent être analysés en termes de pied, et leur schéma accentuel correspondrait apparemment à un schéma à accentuation simple s'il n'y a qu'un pied, et à un schéma à accentuation double, s'il y en a deux. Toutefois, nous devons émettre deux réserves. Premièrement, dans le cas de composés complexes comme *flying-key room*, le statut d'accentuation double n'est pas du tout certain. Ce composé se forme à partir de *flying key* (un syntagme) puis *room* : il forme donc plutôt un composé à accentuation simple, car le dernier élément ne porte pas l'accent /1/. De même, dans le cas de composés à accentuation simple ayant un deuxième élément plurisyllabique comme *ˌyellow ˌhammer*, par exemple, il y aurait également deux pieds à cause du battement rythmique sur la première syllabe de *hammer*. Le schéma accentuel parallèle entre ce composé et *trophy room* n'est pas du tout mis en valeur par une analyse par pied.

Un second problème se pose dans les cas où la première syllabe du composé est inaccentuée car cette syllabe ferait partie du pied précédent. Dans une analyse du rythme en termes de pied, c'est l'énoncé tout entier qui doit être étudié ; cela permet de rapprocher telle ou telle syllabe du pied auquel il est associé. Dans notre analyse des composés, il ne nous paraît pas du tout satisfaisant de devoir diviser le composé pour en associer une partie à une syllabe (ou à une pause) qui précède.

3.1.2.3 La syllabe ou le pied : quelle unité pour l'analyse des composés ?

Nous venons d'émettre quelques réserves en ce qui concerne la pertinence du pied pour l'analyse des composés. L'unité de la syllabe serait donc plus pertinente. De plus, il est nécessaire de connaître la structure syllabique et le schéma accentuel de l'énoncé afin de lui attribuer une structure en termes de pied, comme nous l'avons vu d'après la définition d'Abercrombie (1964). Le passage par la syllabe est donc obligé.

L'unité du pied est étroitement liée aux représentations de la phonologie métrique. Est-ce que la représentation sous forme d'arbre ou de grille métrique pourra nous être utile ? L'arbre métrique montre le lien entre les éléments, ce qui pourrait nous servir ; la grille métrique occulte cet aspect ; cependant, elle montre la hiérarchie accentuelle. Dans le cas des composés à accentuation double, elle nous permet de visualiser les conflits accentuels et surtout de comprendre les cas de déplacement de l'accent que nous risquons de trouver dans notre analyse des composés en discours. Nous n'écartons donc pas complètement ces principes dans notre analyse, mais nous retenons la syllabe comme unité de comparaison car elle semble plus adaptée à notre analyse.

Notre étude porte sur plusieurs axes. Nous comparons les syllabes d'un composé entre elles afin d'analyser leur durée respective, les valeurs de la fréquence fondamentale et l'intensité. A partir de deux extraits différents pour le même composé, nous comparons les deux réalisations afin de tirer des enseignements sur la variation entre locuteurs, ou entre contextes.

Pour l'analyse de notre corpus oral, nous prendrons donc des mesures pour l'ensemble de chaque syllabe correspondant aux corrélats acoustiques qui semblent les plus pertinents : la fréquence fondamentale, l'intensité et la durée. L'étude de l'échantillonnage qui suit nous permet d'évaluer la portée des mesures prises et de l'unité de comparaison choisie. Dans cette section, notre discussion des unités s'est limitée au pied et à la syllabe. Nous serons amenée à nous demander dans la section suivante quelles sont les autres unités qui pourraient nous servir.

3.1.3 Echantillonnage des extraits ayant l'élément «dark»

Nous avons procédé à l étude d'un échantillonnage représentatif du corpus oral d'analyse afin de tester l'efficacité de notre démarche avant de l'appliquer à l'ensemble de ce corpus. Nous allons expliquer notre méthode puis commenter les résultats qui nous permettent d'évaluer la pertinence des unités de référence et des mesures sélectionnées.

3.1.3.1 La méthode

3.1.3.1.1 Le choix de l'échantillonnage

Nous avons analysé un échantillonnage d'extraits ayant le mot « dark ». Cet échantillonnage a été choisi parce qu'il représente une certaine homogénéité car, sur les dix-huit extraits, dix-sept sont tirés de la source littéraire : un seul est tiré de *Reporters* ce qui fournit déjà un point de comparaison entre sources. La comparaison entre les extraits eux-mêmes est facilitée par le nombre d'unités lexicales qui se trouvent dans plusieurs extraits différents : par exemple, *Dark Arts* apparaît dans *anti-Dark Arts spell* (une fois), *Defence against the Dark Arts* (trois fois) et en tant qu'unité lexicale seule (deux fois). Dans cet échantillonnage, les

extraits correspondent à des composés et à des syntagmes. Les deux types de schémas accentuels sont représentés, il y a des contours intonatifs différents, des cas d'insistance, et enfin, l'unité ne se trouve pas toujours dans la même position par rapport à l'ensemble de l'extrait. Cela nous a permis de tirer un certain nombre de conclusions préliminaires qui nous serviront dans l'analyse du reste du corpus oral.

3.1.3.1.2 Le choix des mesures et des unités de référence

Nous avons pris des mesures concernant les corrélats acoustiques de la perception de l'accent, comme nous les avons définis dans la première partie de cette section : le F_0, l'intensité (I) et la durée. En ce qui concerne le F_0, nous en avons pris cinq mesures ; les indications ci-après précisent le nom de la variable suivi de l'explication de son contenu :

F_0Max : la valeur maximale de F_0, en Hz

F_0Min : la valeur minimale de F_0, en Hz

F_0-MAS : la pente moyenne de F_0, en demi-tons/seconde

F_0-Mean : la valeur moyenne de F_0, en demi-tons

F_0-SD : l'écart type de F_0, en demi-tons

Chaque mesure est prise pour une unité de référence donnée. Le choix des hertz pour les maxima et minima correspond à un objectif de précision sur ces valeurs absolues ; le choix des demi-tons pour la pente, la moyenne et l'écart type est inspiré par les travaux de Dujardin (2001 : 141) qui utilise cette mesure afin de comparer plus facilement des locuteurs différents : il s'agit d'une valeur relative et

non absolue. Comme explique Terken (1991 : 1772) : « The logarithmic scale was chosen because musical intervals are equal on a logarithmic scale (slope measured in semitones/s) ».

Pour l'intensité, nous avons procédé de la même manière ; toutefois, les mesures sont toujours en décibels. Nous avons pris les quatre m suivantes :

Imax : la valeur maximale de l'intensité, en dB

Imin : la valeur minimale de l'intensité, en dB

I-Mean : la valeur moyenne de l'intensité, en dB

I-SD : l'écart type de l'intensité, en dB

En ce qui concerne la durée, nous avons relevé les temps de début et de fin de l'unité ainsi que sa durée, et, à l'instar de Dujardin (2001), les temps des valeurs maximales et minimales du F_0 et de l'intensité ont été pris : les variables s'appellent T-F_0Max, T-F_0Min, T-Imax et T-Imin.

Au début du test, nous n'avions prévu que les mesures concernant les maxima et minima du F_0 et de l'intensité. Comme elles se sont révélées tout de suite insuffisantes, nous avons intégré l'ensemble des mesures citées ci-dessus pour l'analyse. La syllabe a été prise comme seule unité de référence, car c'est sur elle que porte l'accent. Afin de déterminer la coupe syllabique, nous nous sommes référée à LPD et nous avons appliqué la coupe syllabique qui y est proposée à notre corpus oral, comme nous l'avons expliqué dans la section précédente. Nous verrons dans nos conclusions que nous avons été amenée à apporter quelques modifications à ce choix, ainsi qu'au choix de certaines mesures.

Ces mesures ont été intégrées dans un TextGrid sous PRAAT. Cette fonction permet de visualiser l'oscillogramme, d'écouter l'extrait sonore, puis de découper le signal selon l'unité choisie. La grille comporte autant de niveaux que l'on souhaite. Nous avions d'abord créé des grilles comportant une ligne pour chacune des mesures, en plus d'une ligne pour la segmentation syllabique étiquetée d'après la transcription phonétique correspondante[14], une ligne correspondant aux mots (transcription graphique)[15] et une ligne correspondant au schéma accentuel. Nous avons adopté la nomenclature que nous avions proposée à la fin du chapitre 1, à savoir :

/1/ : syllabe susceptible de porter la chute intonative /1/

/2/ : accent secondaire, rythmique, sur un élément plurisyllabique (sur un élément portant l'accent /1/)

/3/ : syllabe accentuée du deuxième élément d'un composé à accentuation simple

/P/ : syllabe accentuée du premier élément d'un composé à accentuation double

/R/ accent secondaire, rythmique sur un élément plurisyllabique (sur un élément portant l'accent /P/ ou l'accent /3/)

Nous avons ainsi pu mettre à l'épreuve ce système de notation dans un contexte d'anglais oral ce qui nous permettra d'évaluer son intérêt.

Une fois que toutes les mesures ont été prises, elles ont été transférées dans le tableur EXCEL©[16] ce qui permet une visualisation claire de l'ensemble des

---

[14] Cette transcription correspond à un codage spécifique à PRAAT qui permet l'étiquetage des graphiques en IPA. Voir Annexe B pour les correspondances.

[15] Ces données ont été modifiées par la suite : voir la rubrique suivante, 3.1.3.2.1 : « L'analyse : Les unités de référence »

[16] Tableur du pack Office Pro de Microsoft.

données, et, le cas échéant, permet d'effectuer des calculs qui n'ont pas été faits par PRAAT[17].

3.1.3.2 L'analyse

3.1.3.2.1 Les unités de référence

Lors de l'analyse des données, est très vite apparue la nécessité de définir une ou plusieurs autres unités de référence. La syllabe seule n'est pas une indication suffisante, même si elle permet de tirer un certain nombre de conclusions. Il est vrai que les valeurs maximales et minimales du F_0 et de l'intensité pour un ensemble quelconque peuvent être déduites (voire calculées) à partir des données concernant chaque syllabe ; toutefois, les mesures plus affinées relatives aux valeurs moyennes et à l'écart type ne sont pertinentes que pour l'ensemble que nous voulons analyser. Cette remarque s'applique *a fortiori* pour la pente moyenne de F_0

Nous avons donc rajouté trois autres niveaux d'analyse correspondant aux unités de référence suivantes: le mot (au sens d'élément potentiel de composition), l'unité lexicale (correspondant à l'unité composée ou au syntagme), et enfin l'extrait dans son ensemble, sans les éventuelles pauses ou bruit qui peuvent se trouver au début ou à la fin de l'extrait sonore. Pour ces trois niveaux, nous avons pris les mesures suivantes: la pente du F_0 (F_0-MAS), la valeur moyenne du F_0 (F_0-Mean) et l'écart type (F_0-SD); la valeur moyenne de l'intensité (I-Mean) et l'écart type (I-SD); enfin, la durée.

---

[17] Voir Annexe A pour les tableaux récapitulatifs des mesures de cet échantillonnage.

Ces niveaux supplémentaires nous ont permis de tirer des conclusions par rapport à la durée relative du mot *dark* dans les différentes unités lexicales où il figure. Cela permet également de mieux interpréter la pente du F_0 qui se poursuit souvent sur plusieurs syllabes, d'où l'intérêt de prendre des mesures pour les mots, les unités lexicales et l'ensemble de l'extrait. En effet, nous constatons parfois une pente moyenne importante sur une syllabe inaccentuée ce qui correspond à la poursuite de la chute entamée sur la syllabe accentuée qui précède.

Nous n'avons pas segmenté l'extrait par pieds. C'est la syllabe, avec la coupe proposée par LPD, qui a été retenue, ainsi que les autres unités citées plus haut. Toutefois, bien qu'elle ne s'écrive pas, comme nous l'avons expliqué dans la section précédente, au cœur de notre démarche, cette unité pourra nous éclairer dans certains cas. Comme le dit Fudge (Fudge, 1994 : 13):

> There is, however, important interplay between Stress Groups and Feet. Differences between Stress Group structure and Foot structure may be utilized to account for some complex rhythmic differences.

Nous n'écartons donc pas complètement ce concept du cadre de notre analyse, mais il n'est pas pris en compte systématiquement pour les mesures.

3.1.3.2.2 Les mesures

Comme nous l'avons expliqué sous la rubrique 3.1.3.1.1, « Choix de l'échantillonnage », les extraits analysés présentent des caractéristiques susceptibles de nous éclairer en ce qui concerne les paramètres acoustiques significatifs pour la perception de l'accent. Ils représentent des contextes divers : schémas accentuels variés pour les mêmes unités lexicales, accent d'insistance, différences au niveau de l'intonation, de la formation de l'unité lexicale, de la

position des unités lexicales dans le contexte. Dans les sections suivantes, nous évaluons à quel point ces mesures sont pertinentes et dans quels contextes. Nous étudierons le F_0, puis l'intensité, et enfin la durée et les mesures du temps.

3.1.3.2.2.1 Le F_0

Nous avons tout de suite constaté certaines anomalies qui faussent les résultats : des mesures de F_0 extrêmement élevées (plus de 400 Hz), parfois sur des syllabes inaccentuées, alors que la plupart des mesures ne dépassent pas 250 Hz. Nous avions pourtant déjà procédé à un lissage de la courbe, mais dans la suite de nos analyses, nous adopterons la position de Dujardin (2001 : 118) qui a créé la courbe F_0 en la plafonnant à 350 Hz au lieu de 600 Hz, qui est la valeur par défaut de PRAAT. Les remarques suivantes tiennent compte de ces faits car ces résultats affectent non seulement les valeurs maximales mais aussi les valeurs moyennes de F_0, l'écart type, et la pente.

En écartant les trois extraits ayant des valeurs anormalement élevées, nous constatons une assez bonne corrélation entre la valeur maximale de F_0 et la syllabe qui porte l'accent /1/, surtout pour les unités lexicales courtes. Ainsi, dans *Dark Side*, lorsque l'unité porte une accentuation simple, la valeur maximale de F_0 de *Dark* est plus grande que celle de *Side*, et inversement pour l'extrait où l'unité porte une accentuation double. Toutefois, dans les extraits comportant davantage de syllabes, la corrélation est moindre ; ceci peut s'expliquer par la déclinaison de F_0 (« F_0 declination ») à laquelle nous avons fait référence plus haut (voir 3.1.1.1.2.3). La valeur minimale de F_0 ne semble pas présenter un intérêt particulier ; cette valeur pourrait servir pour marquer le sens du contour intonatif (montant, descendant ou encore creusé ou circonflexe).

Les mesures correspondant aux valeurs moyennes, à l'écart type et à la pente de F_0 doivent être interprétées ensemble. En effet, la moyenne n'a de valeur que si l'on a conscience des écarts par rapport à cette valeur, ce qui est donné par l'écart type ; de même, la pente peut être forte ou douce, ce qui est montré par le graphique bien sûr, mais aussi par l'importance relative de l'écart type. Comme nous l'avons indiqué sous la rubrique précédente, les mesures prises sur une seule syllabe ne sont pas toujours très révélatrices : il a donc fallu revoir le choix des unités de référence mesurées. Cependant, nous avons pu utiliser l'échantillonnage, segmenté syllabe par syllabe, à plusieurs fins : nous rendre compte des valeurs habituelles de ces mesures afin d'être en mesure d'interpréter les résultats ; étudier la relation entre les trois mesures ; enfin, apprécier les tendances par rapport à la valeur accentuelle (/1/ ou/P/) de *dark*, et par rapport à la position de l'élément *dark* dans l'unité.

Si l'on écarte les extraits ayant des valeurs qui dépassent 350 Hz, la pente moyenne (F_0-MAS) se situe généralement entre 0 et 70 demi-tons, la moyenne de F_0 (F_0-Mean) entre –4 et 11 demi-tons, l'écart type (F_0-SD) entre 0 et 3 demi-tons. On constate que lorsque *dark* porte un accent /1/, les valeurs de F_0-MAS sont plutôt élevées (de 26 à 45 demi-tons) ; une des valeurs se situe à 12 demi-tons seulement, mais cela s'explique par le fait que la chute sur l'ensemble de l'unité se fait par paliers et il n'y a que peu de pente sur l'élément *dark*. Les valeurs de F_0-Mean sont aussi élevées (de 6.85 à 11.15 demi-tons) ce qui correspond aux valeurs maximales élevées. La valeur de F_0-SD est plus variée ce qui correspond aux types de courbes : l'une est douce, donc la valeur de F_0-SD est moyenne (1.53 demi-tons), l'une descend par paliers, la chute sur *dark* est donc faible ce qui

donne une valeur très basse, (0.53 demi-tons), enfin l'une monte puis descend avec une forte amplitude et la valeur est élevée (3 demi-tons). Nous pouvons donc en conclure que la pente moyenne de F_0 et celle de l'écart type représentent les deux mesures qui, interprétées ensemble, peuvent nous renseigner le mieux sur le contour intonatif.

3.1.3.2.2.2 L'intensité

L'échelle des mesures pour l'intensité semble acceptable[18]. Les maximales se situent généralement entre 67 et 89 dB, les minimales entre 38 et 59 dB. Pour les valeurs maximales de l'intensité, dans le cas d'accent d'insistance de cet échantillonnage (*dark magic* /10/ comparé à *dark magic* /P1/), l'intensité est un facteur déterminant. De même, comme pour le F_0, la valeur de l'intensité est plutôt élevée (entre 83 et 89 dB), lorsque *dark* porte un accent /1/, mais nous trouvons également des valeurs élevées lorsque *dark* porte un accent /P/ ; cette valeur semble donc moins déterminante que celle du F_0. Par ailleurs il y a souvent des cas où l'intensité sur la syllabe accentuée /P/ est légèrement plus importante que sur /1/ (2 ou 3 dB) ; cette différence ne paraît pas significative, et nous nous demandons, comme Dujardin (2001) si le phénomène de la déclinaison de F_0 ne s'étendrait pas aussi à l'intensité. La valeur minimale de l'intensité ne nous a pas paru pertinente.

Nous avons procédé à une analyse des valeurs, syllabe par syllabe, afin d'étudier la corrélation entre les valeurs de l'intensité moyenne et de l'écart type.

---

[18] Voir Johnson (1996. 50-54) qui explique que la relation entre la force (« loudness ») et l'amplitude n'est pas constante, d'où l'utilisation du décibel (dB) comme unité de mesure car elle tient compte de ceci.

Il y a 72 syllabes dans cet échantillonnage de dix-huit extraits sonores, dont 26 syllabes inaccentuées (les syllabes accentuées sont ou /1/ ou /R/ ou /P/). Ces valeurs devront être comparées à terme à des mesures d'unités de référence plus grandes, mais nous constatons déjà que pour l'écart type, dont les valeurs sont comprises entre 0 et 18 dB, les valeurs pour les syllabes accentuées sont élevées (toutes les valeurs qui dépassent 9 dB, à deux exceptions, sont des syllabes accentuées) alors que celles des syllabes inaccentuées sont comprises entre 0 et 8 dB (avec une à 10 dB et une à 12 dB).

Si nous comparons la courbe des valeurs de l'écart type de l'intensité à celle du F_0, elles sont très différentes : il y a des pics très nets à 12 dB et à 7 dB pour l'intensité alors que pour le F_0, la courbe monte et descend avec un seul pic situé au milieu à 1 et 1,5 demi-tons. Ce constat peut nous faire réfléchir sur les caractéristiques du F_0 et de l'intensité, cette dernière devant être interprétée en termes d'amplitude de pics plutôt qu'en valeurs absolues.

3.1.3.2.2.3 La durée

Cette mesure nous a permis de comparer deux types de sources : *Reporters* et la source littéraire. Nous constatons que la vitesse d'élocution de la source *Reporters* est beaucoup plus rapide que celle de la source littéraire, quel que soit le contexte : *dark* ne dure que 298 ms contre de 304 ms (dans *dark side*) à 498 ms (dans *anti-Dark Arts spell*), et même jusqu'à 516 ms dans le syntagme *dark wizards or witches*.

Ensuite, nous avons procédé à une étude des syllabes accentuées (portant un accent /R/, /P/, /2/ ou /1/) et des syllabes inaccentuées (portant un accent /0/) qui

révèle très nettement que la durée des syllabes inaccentuées tend à être inférieure à celle des syllabes accentuées. Ces résultats corroborent toutes les études que nous avons déjà citées au début de ce chapitre.

Enfin, nous avons étudié la durée relative du mot *dark* par rapport à la durée de l'unité lexicale (composé ou syntagme) mais les résultats ne sont pas concluants. Il ne semble pas qu'il y ait une corrélation entre la durée de *dark*, qu'il reçoive un accent « secondaire » (que nous appelons /P/) ou un accent primaire (/1/ susceptible de porter une chute intonative). Par contre, les différents débits sont évidents, mais la durée s'étale sur l'ensemble de l'extrait et ne porte pas sur *dark* seul. La durée la plus longue est dans *Anti-Dark Arts spell* (53% de *Dark Arts*) alors que la durée relative de *Dark* dans *Dark Forces* ne fait que 39%. Pour *dark wizard*, *dark magic*, les durées se situent entre 45 et 50% quel que soit le contexte (que *dark* soit accentué /P/ ou /1/, qu'il porte un accent d'insistance ou non, qu'il forme un syntagme ou un composé). Toutefois, il est intéressant de noter que *Dark* dans *Defence against the Dark Arts* ne représente que 35-39% de la durée de l'unité *Dark Arts* : la position de *dark* semble influencer sa durée.

3.1.3.2.2.4 Le temps du F_0 et de l'intensité

Les mesures du temps de F_0 ont été prises par Dujardin (2001) afin de servir d'indication sur le contour intonatif. Elle a tenu compte des pics des maxima et minima et de leur temps par rapport à la fin de l'unité tonale, ce qui lui a permis de déterminer s'il s'agissait d'une montée ou d'une descente ou bien d'un ton creusé ou circonflexe. Dans la mesure où nous ne procédons pas, comme elle l'a fait, à une analyse statistique des données, ces mesures ne sont pas très porteuses

car nous disposons du graphique ainsi que de l'extrait sonore lui-même pour apprécier le contour intonatif.

3.1.3.3 Conclusions

Cette étude nous a permis de nous rendre compte qu'il existe une corrélation entre les facteurs acoustiques mesurés (le F_0, l'intensité et la durée) et la valeur accentuelle ainsi que l'accent d'insistance. Quelques tendances ont été notées par rapport à la position de l'unité lexicale dans le contexte, mais des études plus poussées seront nécessaires pour les confirmer. L'étude a été moins probante en ce qui concerne la formation ; nous n'avons pas remarqué de facteurs acoustiques pouvant différencier les unités composées des syntagmes.

Dans l'analyse des différentes mesures du F_0, il est apparu clairement que nous sommes obligée de tenir compte simultanément de deux, voire plusieurs, facteurs afin de tirer des conclusions sur les contours intonatifs. Ce constat est vrai pour l'interprétation de toutes les mesures : c'est la combinaison des facteurs qui contribue à l'impression de l'accent perçu, et non un seul ; les conclusions que nous pourrons tirer devront tenir compte de ce fait.

De même, les mesures devront être interprétées à partir de plusieurs unités de référence. La syllabe ne suffit pas ; l'analyse se fera à d'autres niveaux. Sous la dernière rubrique de cette partie sur l'analyse acoustique, nous précisons la démarche que nous adopterons pour l'étude du corpus oral d'analyse dans son ensemble.

3.1.4 Les mesures et les unités de référence retenues pour le corpus oral d'analyse

L'étude de cet échantillonnage aboutit déjà à certaines conclusions préliminaires, mais elle nous a surtout permis de mieux définir quelles sont les mesures et les unités de référence à étudier dans l'analyse.

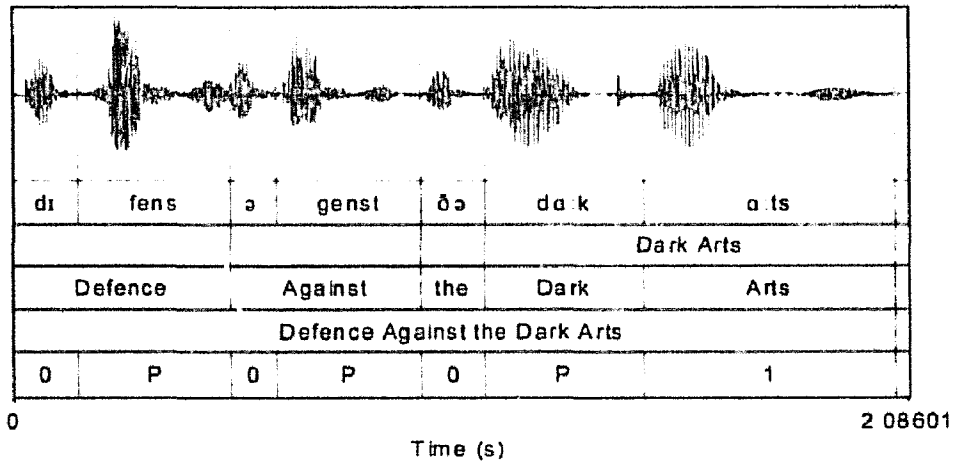
Les extraits seront segmentés par rapport à quatre types d'unités de référence : la syllabe, étiquetée d'après sa transcription phonétique (avec le schéma accentuel correspondant à chaque syllabe sur une ligne distincte), le mot, correspondant à l'élément potentiel de composition, l'unité lexicale (le composé ou le syntagme) et, enfin, l'extrait dans son ensemble. Ces trois dernières unités de référence sont étiquetées selon leur graphie, en respectant la casse. Dans plusieurs cas, l'unité lexicale correspond à l'extrait; dans d'autres, il y a deux unités lexicales et non une seule: par exemple, *Dark Arts* et *Dark Arts spell*, provenant de l'extrait *anti-Dark Arts spell*. Dans ce cas, la segmentation tient compte de ces niveaux supplémentaires.

Les mesures que nous avons décide de prendre correspondent à celles citées plus haut : les valeurs maximales et minimales du F_0 et de l'intensité, les valeurs moyennes et l'écart type du F_0 et de l'intensité, la pente moyenne du F_0. Nous n'avons écarté que les mesures des temps correspondant aux valeurs maximales et minimales du F_0 et de l'intensité. Plutôt que de calculer le type de contour intonatif d'après ces valeurs, les graphiques produits par PRAAT seront utilisés pour l'interprétation du contour intonatif.

Nous travaillons à partir de deux types de graphiques: la figure suivante illustre l'un d'entre eux qui montre l'oscillogramme avec le TextGrid comprenant

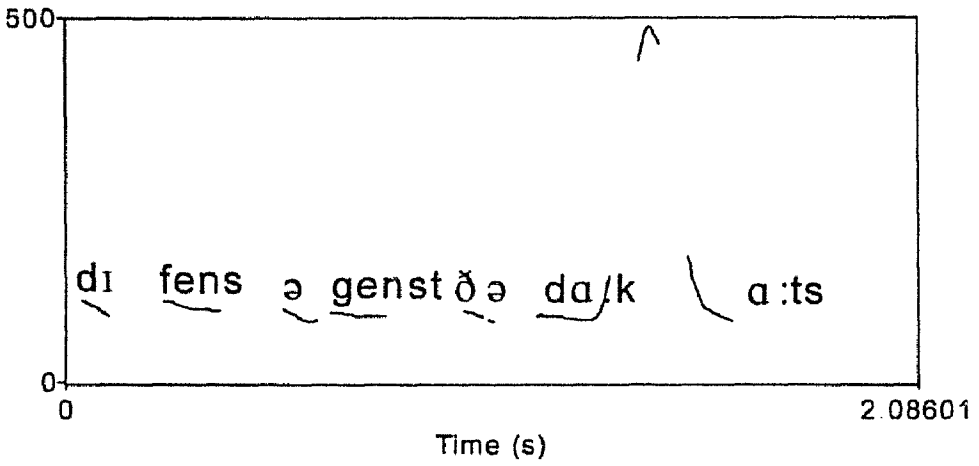
la segmentation de l'extrait selon les différentes unités de référence (la syllabe, le mot et l'unité lexicale), ce qui nous permet de comparer les durées relatives de ces ensembles. Le TextGrid comporte 5 niveaux : la phonie, segmentée par syllabe, l'unité, les mots, l'extrait entier, et enfin, le schéma accentuel.

Figure 3.5 : dark\defdark2 : Oscillogramme et TextGrid avec segmentation par syllabe



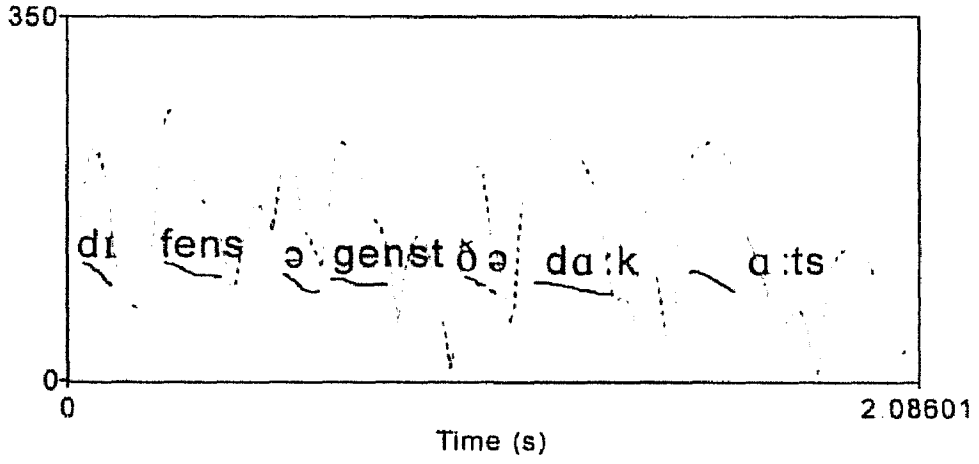
L'autre graphique est composé de la courbe de F_0 (ligne continue) et de celle de l'intensité (ligne brisée), associée à la phonie (étiquetée syllabe par syllabe). Ce graphique permet d'abord de déceler des erreurs : l'on voit des courbes brisées dans le cas d'une valeur de F_0 anormalement élevée, comme dans le cas suivant où la valeur de F_0 se décroche sans raison apparente:

Figure 3.6 : dark\defdark2 : Courbe de F_0 avec valeur erronée



C'est pourquoi nous avons plafonné les valeurs de F_0 à 350 Hz (voir 3.1.3.2.2.1, « Les mesures : Le F_0 »), ce qui nous permet d'obtenir la courbe suivante, qui montre également l'intensité (la courbe brisée) :

Figure 3.7 : dark\defdark2 : Courbe de F_0 avec plafond à 350 Hz et I de 40 à 100 dB



Ce graphique permet de visualiser l'extrait dans son ensemble : nous nous rendons compte des tendances qui sont plus révélatrices que les valeurs absolues

des maxima et minima, du sens du contour intonatif et de l'importance de la pente ; nous pouvons ainsi facilement comparer les unités de référence entre elles.

La plupart de ces analyses seront faites indépendamment de l'analyse des sources écrites, à l'aide de PRAAT et d'EXCEL. Toutefois, nous incorporons dans la base de données Access un certain nombre d'informations qui nous permettront de faire des rapprochements entre ces deux sources. Nous passons à l'étude de cet autre outil informatique d'analyse du corpus.

3.2 La base de données sous Access

Access est ce que l'on appelle un « SGBDR », c'est-à-dire un système de gestion de base de données relationnelles. Cet outil permet de réunir de nombreuses informations et d'établir des liens entre elles. Nous allons étudier la structure de la base de données qui doit être conçue afin de faciliter ces mises en relation.

3.2.1 Structure de la base de données

Nous avons décidé d'intégrer notre corpus d'analyse dans une base de données, ce qui nous permettra par la suite d'effectuer des traitements automatiques sur l'ensemble du corpus, surtout à partir des sources écrites. On doit distinguer trois termes clés. La « **table** » est le fichier qui contient l'ensemble des entrées ; une base de données peut comporter plusieurs tables différentes qui seront ensuite reliées entre elles. A l'intérieur d'une table, chaque entrée est associée à plusieurs « **champs** » ; chaque champ contient une information. La structure de la base dépend de ces deux notions : tables et champs. Ensuite, afin d'analyser les données, on peut formuler des « **requêtes** » qui permettent de faire

des tris et de sélectionner certaines entrées selon des critères précis. Ces trois notions sont interdépendantes.

Si une base de données permet de traiter un volume important d'informations, ce type de traitement n'est efficace que dans la mesure où la base est bien conçue. Cette condition est nécessaire, mais non suffisante, pour que l'on puisse formuler des requêtes claires qui fourniront des résultats pertinents. Sous les rubriques suivantes, nous allons d'abord nous intéresser au type de requêtes à formuler, afin de bien cerner nos objectifs ; puis, nous expliquerons la structure même de la base par rapport aux tables et aux champs qu'elle comporte. C'est à partir de cette structure que nous pourrons formuler des requêtes et nous ferons aussi état de certaines considérations techniques dont il faut tenir compte pour que cet outil de traitement automatique des données soient complètement cohérent.

3.2.1.1 Les requêtes

Une fois que la base de données est prête, nous pouvons formuler des requêtes afin de l'interroger. Les requêtes nous permettent de trier un très grand nombre de données (notre base comporte plus de vingt-et-un mille entrées), et nous obtenons ainsi une sélection d'entrées que nous pourrons analyser.

Les requêtes portent sur plusieurs domaines ; nous donnons le détail de celles utilisées pour nos analyses dans l'Annexe A du chapitre 4. Avant d'étudier les résultats fournis par les requêtes dans le chapitre suivant, nous allons donner un aperçu du type de recherche qui peut s'effectuer grâce à ce moyen informatique. Notre analyse s'appuie sur deux axes : la comparaison des sources entre elles, et la comparaison des données provenant d'une même source. Nous

voyons d'abord quel type de requête nous permettra de comparer les différentes sources entre elles.

3.2.1.1.1 Comparaison des données de sources différentes

Les données des sources écrites proviennent de trois dictionnaires lexico-phonétiques : LPD90, EPD et NPD. La taille des dictionnaires est très différente quant à leur inclusion des composés : il y a beaucoup plus d'entrées de LPD que des autres et notre premier souci est de bien étudier leur contenu respectif. Quelles sont les entrées communes (sans tenir compte des variantes graphiques) et quelles sont celles qui diffèrent ? Les différences entre les dictionnaires nous éclairent en ce qui concerne l'évolution du lexique. Les requêtes nous permettront d'établir une liste des entrées de EPD absentes de LPD et nous pourrons ainsi tester l'hypothèse selon laquelle les entrées de EPD qui ne figurent pas dans LPD sont archaïques. Un mot comme *motor-bus* ne figure plus dans LPD, ainsi que *motor-spirit* qui, lui, n'est même pas mentionné dans Chambers (1988). Il sera sûrement plus difficile de tirer des conclusions à propos des entrées de LPD ne figurant pas dans EPD à cause de l'importance du premier qui contient environ trois fois plus d'entrées. Nous nous attendons à trouver des néologismes, mais aussi beaucoup de composés existant à l'époque de EPD (voir 4.3.3, « Etude de l'évolution du lexique »).

A partir de la liste des entrées communes, quelle que soit leur orthographe (avec ou sans trait d'union, avec ou sans espace), nous étudierons l'évolution de la graphie et aussi celle du schéma accentuel. Par exemple, la graphie du mot *network* a changé depuis EPD où le mot s'écrit avec un trait d'union ; son schéma accentuel reste inchangé : /10/. De même, *motorbike* et *motorcycle* sont tous deux

écrits avec un trait d'union dans EPD et le schéma accentuel est /101/[19], c'est-à-dire une accentuation double, pour *motor-bike*, et /1020/ pour *motor-cycle*, c'est-à-dire une accentuation simple. Dans les deux cas, une variante est indiquée correspondant à l'autre type de schéma. Dans LPD, le schéma indiqué est un schéma à accentuation simple, sans aucune mention de variante. Nous ferons également des rapprochements entre les différentes graphies et le schéma accentuel ce qui nous permettra de tester si les deux sont liés (voir 4.3.2.3.2, « La corrélation entre le schéma accentuel et la graphie »).

Enfin, la base comporte non seulement les entrees des sources écrites mais aussi celles du corpus oral d'analyse. Les requêtes nous permettent de regrouper les entrées communes aux sources orales et une ou plusieurs sources écrites. Nous pourrons ainsi comparer les indications correspondant à l'accent lexical dans des composés comme *armchair* ou *bottle top* et leur comportement en discours dans des contextes différents. En ce qui concerne des néologismes tels que *trophy room*, il est possible de les comparer à des entrées des sources écrites ayant le même premier (ou deuxième) élément. Nous pourrons regrouper tous les composés en *-room* par exemple (voir 4.6.3, « Etude des entrées comportant un des éléments du corpus oral ».

[19] Notation de l'époque : les entrées à accentuation double sont marquées avec un accent 1 sur les deux éléments. Pour les entrées à accentuation simple où le deuxième élément est plurisyllabique, la syllabe accentuée du deuxième élément est marquée /2/ dans EPD qu'on retrouvera marquée 3 dans LPD90.

3.2.1.1.2 Comparaison des données d'une même source

Les requêtes permettent de sélectionner les entrées graphiques identiques qui se différencient par leur schéma. Nous pourrons ainsi regrouper tous les composés tels que *toy factory* qui ont deux entrées distinctes correspondant à des schémas et à des sens différents. Ces paires sont intéressantes car elles peuvent nous éclairer quant aux critères qui déterminent le type de schéma (voir 4.6.1, « Etude des entrées ayant deux types d'accentuation »).

Nous comparons des entrées ayant le même premier ou le même deuxième élément. Dans une consultation classique de dictionnaire, il est facile de rassembler les composés qui commencent de la même façon, mais pour étudier ceux qui se terminent de la même façon, il faut se servir d'un dictionnaire où la graphie inverse est indiquée. Les moyens informatiques sont tout à fait adaptés à ce type de recherche et nous pouvons formuler des requêtes à partir d'une séquence quelle que soit sa position dans le composé. Nous pourrons ainsi étudier la fréquence de ces éléments et les schémas qui leur sont associés (voir 4.6.3).

Nous voulons aussi regrouper des entrées ayant plusieurs caractéristiques communes. Si nous associons le même schéma accentuel et le même type de graphie, sans espace et sans trait d'union par exemple, nous pourrons tester l'hypothèse de la corrélation entre cette graphie et le schéma à accentuation simple. De même, nous pourrons voir quelles sont les corrélations entre une graphie à trait d'union et le type de schéma accentuel (voir 4.3.2.3.2).

Ces requêtes peuvent être affinées pour trier les données selon la catégorie syntaxique, ce qui nous permettra de relever les adjectifs composés et d'étudier

leur graphie, leur schéma et la tendance au déplacement de l'accent. D'autres types d'information peuvent servir de base pour les requêtes. Nous avons constaté une tendance à l'accentuation double qui augmente en fonction du nombre de syllabes. Si nous formulons une requête à partir du nombre de syllabes et le type de schéma, nous pourrons tester cette hypothèse (voir 4.3.2.1, « Etude du schéma accentuel »).

Cette rubrique nous a permis de définir un certain nombre d'objectifs. Dans les rubriques suivantes, nous expliquons les principes de la table et du champ, qui forment à eux deux la structure informationnelle de la base de données. C'est à partir de cette structure, qui doit rendre accessibles les informations qui nous intéressent, que nous pourrons formuler des requêtes pertinentes.

3.2.1.2 Les tables et les champs

3.2.1.2.1 Les tables de la base de données

La base de données Access est un ensemble de tables et les mêmes requêtes peuvent s'appliquer à toutes les tables d'une même base. C'est pourquoi nous avons décidé d'inclure toutes les sources de notre corpus sous cette forme. Pour nous, chaque source correspond à une table, donc pour les sources écrites (dictionnaires lexico-phonétiques), nous en avons quatre, correspondant aux composés de NPD (appelée NPD-Cp), de EPD12 (appelée EPD12-Cp), de la version électronique de LPD (appelée LPD-Cp) et de la version papier de LPD (appelée N-COMPACT). Pour les sources orales, nous en avons une seule, qui correspond aux extraits sonores de *Harry Potter and the Philosopher's Stone* et aux deux sources documentaires (*Reporters* et *Documentaires divers*), appelée ORAL-extraits.

Toutes les entrées d'une même table devraient être différentes. En ce qui concerne les sources électroniques, nous avons minimisé les risques de redondance en faisant des requêtes qui tenaient compte des extractions précédentes. Ainsi, nous avons d'abord extrait les entrées à trait d'union ; puis, lorsque nous avons extrait les entrées dissyllabiques, nous avons écarté celles ayant un trait d'union. Des fonctions d'Access permettent de repérer les entrées identiques, et nous les appliquons afin de nous assurer que la table LPD-Cp ne comporte pas de doublons (voir 4.3.1, « Vérification des différentes tables de la base de données : les doublons »). Ce genre d'erreur peut aussi exister dans la table des composés non compacts de la version papier de LPD90, saisie manuellement (N-COMPACT) ; elle est également vérifiée. Cependant, cette opération s'accompagne de beaucoup de précautions car si l'entrée graphique est identique, les autres informations peuvent être différentes. Par exemple, *toy factory*, que nous avons déjà cité et que nous voulons justement étudier, comporte deux entrées distinctes avec des schémas accentuels différents : *toy factory* en /1300/ signifie « une usine qui fabrique des jouets » alors qu'en /2100/ il signifie « une usine qui est un jouet pour enfants » (voir 4.6.1, « Etude des entrées ayant deux types d'accentuation »). Cet exemple est tiré de N-COMPACT. Dans les tables correspondant aux sources orales, il est fréquent de retrouver la même entrée graphique mais il s'agit d'un extrait différent, dit dans un contexte différent.

Chacune de ces informations est saisie dans un champ de la table. Sous la rubrique suivante, nous allons commenter la structure des tables en expliquant quelles sont les informations que nous avons incluses, champ par champ.

3.2.1.2.2 Les champs de la base de données

La base de données comporte ainsi cinq tables; ces cinq tables ont quatre structures différentes au niveau des champs (seules NPD-Cp et LPD90-Cp correspondent), ce qui implique que les informations contenues dans chacune sont différentes. Cette structure hétérogène résulte de la diversité des informations disponibles dans la source elle-même, du mode de saisie (à partir d'une source électronique ou d'une saisie manuelle), et du type de source (les informations associées aux sources orales sont en partie différentes de celles des sources écrites). Il est nécessaire que chaque information que nous voudrons isoler ou utiliser comme critère de sélection soit saisie dans un champ indépendant.

Pour les sources écrites électroniques, la saisie des informations s'est faite à partir d'une sélection des entrées, comme nous l'avons expliqué dans le chapitre 2 (2.2, Sélection des données: sources écrites»). Les informations associées à ces entrées dans le fichier électronique ont donc été intégrées directement, sans qu'il y ait risque d'erreur. Pour les sources orales, nous avons inclus des informations relatives à l'identification de l'extrait et de son contexte en discours, en plus des indications relatives au schéma accentuel. Pour la saisie manuelle des composés non compacts du LPD, nous avons sélectionné les informations qui nous semblent les plus pertinentes, et que l'on peut intégrer sans que le risque d'erreur soit trop important. Ensuite, afin de disposer d'une structure pour chaque table qui permette de les comparer entre elles, et de sélectionner les entrées d'après des critères pertinents, d'autres informations ont été saisies.

Le tableau suivant récapitule les différentes informations que contiennent les tables, de sorte que la structure de chacune apparaît clairement. Les

typographies correspondent à des types de saisie différents. Les cases marquées « OUI » en petites majuscules correspondent aux informations qui sont intégrées directement dans la table à partir d'une source écrite électronique. Les cases marquées « oui » en minuscules correspondent à celles qui sont saisies manuellement. Lorsque « (oui) » est indiqué entre parenthèses, la saisie est le résultat d'une fonction opérée à partir d'un autre champ (par exemple, la graphie de N-COMPACT a été saisie grâce à une fonction permettant de rassembler les champs contenant les éléments graphiques individuels). Là où nous avons pris la décision de ne pas saisir l'information, la case est vide. Enfin, les cases barrées correspondent aux informations non pertinentes (le nom du fichier son pour une source écrite, par exemple). Le nom des champs communs à toutes les sources figure en gras. Pour ORAL-extraits, lorsque figure la mention « Oui* », la saisie correspond aux extraits de la source littéraire seulement.

Figure 3.8 : Les champs associés aux différentes tables

TABLES / CHAMPS	EPD12-Cp	NPD-Cp	LPD90-Cp	N-COMPACT	ORAL-extraits
CHAMPS CONCERNANT L'IDENTIFICATION DE L'ENTREE					
Source (table ou critère)	Oui	Oui	Oui	Oui	Oui
Nom du répertoire\fichier son	———	———	———	———	Oui
(N° page*[20]) / concordance	———	———	———	———	Oui
CHAMPS POUR LA GRAPHIE					
Graphie	OUI	OUI	OUI	(Oui)	Oui[21]
Graphie inverse		OUI	OUI		
élément 1				Oui	Oui*
élément intermédiaire				Oui	Oui*
élément final				Oui	Oui*
CHAMPS CONCERNANT LE SCHEMA ACCENTUEL					
Nombre de syllabes	Oui	OUI	OUI	Oui	Oui
Schéma accentuel (schéma de l'accent lexical ou schéma perçu des extraits sonores)	OUI	OUI	OUI	(Oui)	Oui
Schéma accentuel avec indications particulières[22]				Oui	Oui
Type de schéma (s / d) (correspondant au schéma de l'accent lexical ou au schéma perçu des extraits sonores)	Oui	Oui	Oui	Oui	Oui
Type de schéma (s / d) prévu (accent lexical)					Oui
Type de schéma (s / d) anglo (accent prévu par l'anglophone)					Oui*
Variantes du schéma accentuel				Oui	———
Stress shift (tendance)			(Oui)[23]	Oui	———
CHAMPS CONCERNANT LA PRONONCIATION DES SEGMENTS					
Phonie	OUI	OUI	OUI		(oui)[24]
Variantes phonie		OUI (PEU)	OUI		
Valeur de la voyelle non accentuée (dissyllabes)			Oui		
CHAMPS INDIQUANT DES INFORMATIONS COMPLEMENTAIRES					
Type de composé CR, OPC, C			Oui	Oui	Oui
Catégorie syntaxique du composé	OUI	OUI	OUI		Oui
Formation					Oui

[20] Toutes les entrées marquées d'un * concernent la source *Harry Potter* uniquement

[21] Pour *Harry Potter*, la saisie s'est faite à partir d'une fonction qui concatène les éléments individuels

[22] N-COMPACT indique la borne # entre chaque élément. ORAL-extraits indique le schéma selon les indications proposées dans le chapitre 1, en incluant la borne #.

[23] Cette indication apparaît dans le champ variation, sous une forme codée.

[24] Ces indications sont dans un fichier EXCEL à part, avec les mesures acoustiques. Voir Annexe A de ce chapitre et Annexe D du chapitre 4. Le codage utilisé est un codage propre à PRAAT. Voir Annexe B de ce chapitre.

Ce tableau montre clairement la diversité des structures des tables. Nous constatons qu'il existe cinq champs communs ; sinon, le nombre de champs pour chaque table varie de sept à dix-huit sur un total de vingt-trois champs proposés. Cependant, l'objectif d'une base de données est de pouvoir comparer les informations ; il est donc essentiel de disposer d'une structure rigoureuse qui comporte toujours les mêmes données essentielles. C'est pourquoi les cinq champs communs ont une telle importance. Nous allons expliquer les raisons pour lesquelles nous les avons choisis, les contraintes techniques dont il faut tenir compte, puis nous commenterons la nécessité des champs spécifiques à certaines tables.

3.2.1.2.2.1 L'identification de la source

Sous la rubrique « Requêtes », nous avons cité la comparaison des différentes sources comme un des objectifs principaux de notre recherche, au niveau de la graphie ainsi que sur le plan du schéma accentuel. Il est donc nécessaire que, dans chaque table, la source de l'entrée soit indiquée. Il existe donc un champ qui contient cette information, ce qui nous permet de l'identifier lorsque nous comparons des listes d'entrées sélectionnées à partir de critères graphiques ou accentuels.

L'identification des sources orales est d'autant plus nécessaire que nous retenons plusieurs fois les mêmes unités à dessein. C'est pourquoi deux autres champs sont prévus pour les sources orales : le nom du fichier son et son répertoire, puis la concordance qui replace l'entrée dans son contexte de discours. Il faut aussi pouvoir les comparer aux entrées des sources écrites qui leur correspondent mais cela nécessite des manipulations particulières car, pour un

même composé, nous trouverons des graphies différentes. Nous nous tournons vers ce type d'information.

3.2.1.2.2.2 La graphie

Nos requêtes aboutissent toutes à des listes d'entrées où la graphie va figurer en première place ; c'est ainsi que l'on identifiera une entrée. Ce champ est donc essentiel, et c'est pourquoi il a été créé là où il n'avait pas été saisi directement (dans N-COMPACT et les entrées provenant de *Harry Potter* dans ORAL-extraits où la saisie a été faite élément par élément).

Comme nous l'avons rappelé sous la rubrique précédente, nous voulons sélectionner les entrées communes à deux ou à plusieurs tables. Pour ce faire, nous devons neutraliser les distinctions graphiques. Des tables parallèles ont été créées, en retirant les traits d'union et les espaces, là où il y en avait. Ces nouvelles tables ne contiennent que des entrées écrites comme des composés compacts dans le champ graphie. Toutefois, la graphie d'origine peut être retrouvée d'après la source, ce qui montre à nouveau l'intérêt de ce champ : EPD-Cp ne contient que des composés à trait d'union, NPD-Cp ne contient que des composés compacts et pour N-COMPACT, les éléments sont séparés par l'espace. En ce qui concerne LPD90-Cp, où l'on trouve au moins deux graphies, le champ de la graphie inverse nous permet de retrouver la graphie d'origine. Pour ORAL-extraits, elle se retrouve dans le champ du contexte du discours (Voir 4.3.2.2, « Etude de la graphie »).

Il faut rappeler que, dans la table originelle, les variantes graphiques correspondant aux entrées des dictionnaires ne doivent en aucun cas être

modifiées. Toutefois, la version électronique de EPD12 a des caractéristiques particulières. Guierre notait toutes les entrées graphiques en minuscules car les majuscules étaient réservées pour la phonie. Nous avons donc modifié la table et introduit des majuscules dans la graphie là où elles figurent dans la version papier de EPD12 : *easter day* est écrit *Easter Day* (voir 2.2.1.3.1).

En ce qui concerne les champs spécifiques à certaines tables, la graphie inverse est incluse pour NPD-Cp et LPD-Cp. Ces données figurent dans le fichier électronique d'origine. Même s'il est possible d'inclure cette information pour les autres sources grâce à une fonction, nous n'avons pas jugé utile de le faire. Par contre, lors de la saisie manuelle de N-COMPACT, nous avons saisi la graphie élément par élément, ce qui nous permet de disposer des différents éléments de composition dans des champs indépendants, donc nous pourrons sélectionner des entrées en nous servant de ce critère. Pour la saisie des entrées de *Harry Potter*, la saisie de la graphie a été effectuée de la même manière. Cette saisie est pertinente parce qu'il s'agit d'une source qui est aussi écrite ; par contre, la graphie est la nôtre, établie à partir de dictionnaires, pour les sources documentaires entièrement orales.

3.2.1.2.2.3 Le schéma accentuel

Trois champs ayant trait au schéma accentuel sont communs à toutes les tables : « Nombre de syllabes », « Schéma accentuel » (accent lexical ou accent perçu des extraits sonores), « Type de schéma » (correspondant aux schémas du champ « Schéma accentuel »). Certaines requêtes qui testent des hypothèses concernant la corrélation entre la longueur du composé et son schéma accentuel nécessitent le critère du nombre de syllabes. C'est pourquoi cette information a été

saisie pour toutes les sources, y compris pour les sources écrites électroniques là où elle n'est pas disponible dans le fichier d'origine (EPD-Cp).

En ce qui concerne le schéma accentuel, cette information est incluse sous une forme codée qui reprend les notions d'accent primaire, d'accent secondaire et de la non-accentuation (ou désaccentuation). Toutefois, bien que ce champ soit commun à toutes les tables, il ne peut pas être utilisé afin de sélectionner des schémas semblables à cause des différents codages. En effet, nous avons conservé dans ce champ la notation d'origine des fichiers électroniques, et pour ORAL-extraits, nous avons saisi cette information avec le codage proposé dans le chapitre 1 (voir 1.7, « Nomenclature et notation adoptée »). Pour les sources orales, il s'agit du schéma effectivement perçu ; pour les autres sources, il s'agit de l'accent lexical, c'est-à-dire de la prévision de l'accent hors contexte. EPD12 utilise un schéma /11/ pour des composés à accentuation double alors que les autres utilisent /21/ pour le même schéma, ORAL-extraits propose /P1/ ; LPD est le seul, avec notre proposition de notation, à employer l'accent /3/ ; nous trouverons /2/ à la place dans NPD et EPD12 ; enfin, le schéma accentuel de NPD s'accompagne d'informations sur les variantes introduites par « β » qui empêchent la comparaison de ce champ avec celui des autres tables. Ce champ présente un intérêt parce qu'il donne des informations détaillées sur le schéma, mais il ne peut pas servir de critère de recherche, sauf peut-être à l'intérieur de la table.

Si nous gardons ces codages différents, c'est parce qu'ils sont représentatifs de certains partis pris théoriques (l'accent /3/ de LPD90) et d'une certaine époque. Par contre, certaines anomalies de EPD12 ont été corrigées, pour que la table reflète bien la version papier du dictionnaire. Ces modifications concernent

surtout les mots dérivés qui doivent porter un accent /2/ sur le deuxième élément. Ainsi, *knick-knackery* a le schéma /1200/ et non /1000/[25]. Jones écrit dans ses explications au début de EPD12 (Jones, 1963 : xxiv) :

> When a *head-word* is a compound of which the second element is a weakly stressed monosyllable, and a termination for forming a derived word adds yet another syllable, it is to be understood that the derived word has secondary stress on the first syllable of the second element, except when otherwise indicated.

Par ailleurs, nous remarquons que le schéma d'un mot tel que *dinner-hour* doit être /1020/ car dans EPD12 l'on considère que *hour* comporte deux syllabes (comme *power*, *fire*, *iron*, *royal*). Là où les schémas indiquent que ces éléments ne comportent qu'une seule syllabe, ils ont dû être corrigés. Par contre, nous avons laissé des schémas où /2/ est indiqué après /1/ sur certains deuxièmes éléments monosyllabiques tels que *caraway-seed*, mais cela témoigne d'un manque de cohérence dans la notation du dictionnaire.

Enfin, afin de disposer d'un champ qui permet de comparer des schémas accentuels entre des tables différentes, nous avons créé le champ « Type de schéma ». Le schéma accentuel est saisi sous une forme simplifiée, indiquant seulement s'il s'agit d'un composé à accentuation simple (codé « s ») ou d'un composé à accentuation double (codé « d »). Nous avons choisi ce codage car, si on devait simplifier le débat, on pourrait dire que nous nous posons toujours la même question : l'accent est-il à gauche ou à droite (selon la nomenclature de Boisson (Boisson, 1980)), l'accent est-il simple ou double ? Ce champ correspond aux prévisions de l'accent lexical pour les sources écrites, et au schéma effectivement réalisé pour les extraits sonores.

---

[25] L'entrée principale (« head word ») est *knick-knack*, qui est suivi de *–ery* en sous-entrée.

De plus, il y a plusieurs champs spécifiques pour le schéma accentuel. D'abord, pour les sources orales, nous avons inclus « Type de schéma prévu » codé « s » ou « d », correspondant a l'accent lexical des extraits sonores, le cas échéant. Les mêmes codes sont utilisés afin de garder une cohérence et de pouvoir les comparer. Pour ceux provenant de la source littéraire, et qui ne sont pas lexicalisés, il y a également un champ « Type de schéma anglo » qui correspond aux prévisions de l'accent par l'anglophone lors de la lecture du livre. Dans quelques cas, ce champ est différent du schéma effectivement perçu, ce qui mérite d'être étudié.

Pour les tables où la saisie est manuelle, nous avons aussi saisi les schémas, élément par élément, en indiquant la borne morphologique #. Les variantes accentuelles ainsi que la tendance au déplacement de l'accent sont indiquées dans des champs spécifiques pour N-COMPACT, et ces informations sont présentes pour LPD-Cp dans le champ des variantes phonie. NPD fournit peu d'indications[26] ; EPD12 n'en fournit aucune. Pour les sources orales, il n'y a pas non plus de référence à la variation puisqu'il s'agit, pour chaque entrée, d'une occurrence unique.

3.2.1.2.2.4 La prononciation des segments

Il n'y a pas de champ commun à toutes les tables en ce qui concerne la prononciation car elle n'est pas indiquée pour N-COMPACT parce que le travail de saisie, et le risque d'erreur, sont prohibitifs. De plus, nous ne disposons

---

[26] Quelques variantes sur le schéma sont indiquées. Les deux variantes sont séparées par ß mais les informations ne sont pas très lisibles.

d'aucun moyen pour indiquer la phonie sous une forme de prévision, si ce n'est le dictionnaire LPD dans son ensemble, car le dictionnaire n'indique que la graphie du composé dans la sous-entrée avec l'accentuation sous la forme *ˌyelloˌ ˌhammer*. Cela suppose que la prononciation des éléments correspond à celle des éléments pris isolément, et que seul le schéma peut changer. Cette supposition implique que tout deuxième élément monosyllabique, inaccentué, ne subit pas de réduction vocalique.

La phonie est indiquée par ailleurs dans toutes les tables correspondant à une source électronique. Toutefois, la version électronique de EPD12 comporte une notation de la phonie qui n'est pas dans le code SAMPA (voir les remarques en 2.2.1.3.1 «English Pronouncing Dictionary, 12ème édition»). En ce qui concerne les sources orales, la phonie se trouve dans le fichier EXCEL avec les mesures acoustiques (voir Annexe A), sous une forme codée qui est propre à PRAAT (voir Annexe B).

3.2.1.2.2.5 Les informations complémentaires

Le champ «Type de Cp» est présent dans les tables LPD-Cp, N-COMPACT et ORAL-extraits; il fournit au départ des indications simples permettant de sélectionner les composés noms propres de la base, ou bien les composés réduplicatifs. Ce codage pourra être affiné pour inclure d'autres catégories.

Le champ «Catégorie syntaxique» concerne les trois tables provenant de sources électroniques. Le codage se présente sous la forme de N (nom), V (verbe), Q (adjectif)... Pour les sources orales, nous avons inclus cette information ainsi

qu'un champ « Formation » qui récapitule la formation du composé, en utilisant ces mêmes codes (voir aussi 4.3.2.4, « Etude de la categorie syntaxique »).

3.2.2 L'indexation et la présentation des données

La base de données a été constituée pour faciliter les recherches sur les composés ; deux considérations sont à prendre en compte: l'accès aux données et leur présentation. Nous allons d'abord étudier de quelle manière elles sont présentées par le dictionnaire LPD sous sa forme papier.

3.2.2.1 Accès et présentation des données dans LPD

Tout dictionnaire adopte un classement alphabétique des entrées, habituellement à partir du début de celle-ci. Quant aux composés compacts ou ceux écrits avec un trait d'union, nous les retrouvons également par ordre alphabétique, classés à partir du début de la graphie, donc sous le premier élément, quel que soit le type du composé. Pour les composés non compacts, ils seront généralement mis dans une sous-entrée, classés sous le premier élément. Toutefois, ce classement ne semble pas toujours adapté aux composés

Dans LPD90, lorsqu'un composé est formé à partir d'un premier élément dérivé, si cet élément existe isolément le composé est classé sous ce mot ; s'il n'existe pas en tant que mot isolé, alors le composé sera classé sous le radical. Ainsi, *licensed premises* et *licencing laws* se trouvent tous les deux sous *licence, license*, alors que *writing desk* se trouve sous *writing*. Parfois, cette méthode de classement porte à confusion. Les composés *training college* où *training* vient du verbe *train* dans le sens de « formation », et *train spotter* qui est formé à partir du mot *train* au sens de « train ferroviaire » sont classés tous les deux sous *train*, sans

qu'il n'y ait aucune indication des différences de sens. De même, il n'y a pas de distinction entre majuscules et minuscules, donc sous l'entrée *miss, Miss*, nous trouvons deux sous-entrées : *Miss Abbott* et *missing link*. Ces confusions résultent de l'homonymie des entrées et posent la question de leur traitement dans un dictionnaire de prononciation. Comme leur prononciation n'est pas différenciée, dans ce genre d'ouvrage, il n'est pas non plus nécessaire de les y distinguer sous la forme d'entrées séparées.

Le classement interne des sous-entrées n'est pas non plus complètement satisfaisant. Pour reprendre l'exemple de *licence, license*, nous avons cinq sous-entrées, ordonnées comme suit : *licensed premises, licensed practical nurse, licensed victualler, license plate* et enfin *licensing laws*. C'est-à-dire que le composé formé à partir du radical se trouve au milieu des composés formés à partir d'un dérivé. Le classement se fait selon un ordre alphabétique strict qui ne tient pas compte de l'espace. Nous aimerions regrouper les composés selon leur formation : c'est-à-dire, d'abord les composés formés à partir du radical, puis ceux formés à partir d'un dérivé, puis ceux formés à partir d'un autre dérivé. Dans le cas de *licence*, notre ordre serait le suivant : *license plate / licensed premises, licensed practical nurse, licensed victualler / licensing laws*.

Notons que nous avons également décelé quelques cas où un composé se trouve en entrée principale, écrit en un seul mot, et aussi dans une sous-entrée, écrit en deux mots. C'est le cas de *teargas, tear gas*.

La méthode de classement utilisée dans LPD s'explique aisément parce que c'est un dictionnaire de prononciation et, dans la mesure où il n'y a aucune

différence de prononciation, il n'est pas gênant de rassembler ces mots, comme nous l'avons déjà signalé. Par contre, dans une étude des composés, nous nous intéressons et à leur prononciation et à leur formation, nous voulons trouver une formule plus adaptée. De plus, le système d'entrées principales et de sous-entrées classées sous le premier élément (s'il existe isolément, sinon l'entrée principale correspond au radical) ne favorise pas les recherches sur les composés, car en anglais ce que l'on peut considérer comme étant l'élément clé (le déterminé) est fréquemment celui qui se trouve à la fin du composé. Nous allons voir si, en utilisant la base de données, nous pouvons trouver une solution plus satisfaisante.

3.2.2.2 Accès aux données de la base

La base de données est un support électronique qui permet de consulter des données et de faire des listes d'entrées sélectionnées selon certains critères précis. Les deux problèmes d'accès et de présentation subsistent, mais avec des exigences différentes d'un simple support papier.

Pour faciliter l'accès, il serait opportun de prévoir un système d'indexation pour permettre des recherches sur tel ou tel élément. La question est posée clairement par Tournier (1988 : 183) :

> Doit-on les placer à l'adresse du déterminant ou du déterminé ? Le plus logique serait de les placer à l'adresse du déterminé - qui est l'élément principal, mais en prévoyant un renvoi dans l'entrée concernant le déterminant.

Access permet une recherche sur un champ, donc l'accès est possible à partir d'un élément dans les tables conçues ainsi (N-COMPACT, ORAL-extraits). De même, nous pouvons rechercher des suites de caractères dans un champ ; donc nous avons la possibilité de faire des recherches sur le champ graphie dans les autres tables à condition que l'élément soit suffisamment identifié. La recherche

de la suite « fire », par exemple, est pertinente car cette séquence en début ou en fin de mot correspond généralement à un élément de composition alors que la séquence « age » risque d'inclure dans la sélection un grand nombre d'entrées qui se terminent par cette suite de trois caractères et non par l'élément de composition « age ». Nous ne pouvons sélectionner les éléments de cette manière parce que nous n'avons pas pu inclure la borne # pour les composés compacts. A terme, il est souhaitable que ce travail soit réalisé, mais il n'est pas possible de le faire par un traitement automatique, comme nous l'avons déjà expliqué au chapitre 2 (2.2.2.6). La sélection des entrées par élément est donc sujette aux contraintes liées à la saisie des éléments, telle que nous l'avons décrite ci-dessus.

3.2.2.3 Présentation des données de la base

Nous nous tournons vers l'autre considération, d'ailleurs très liée à cette question, qui concerne la présentation des données : c'est-à-dire, l'ordre dans lequel elles apparaîtront. Habituellement, les mots d'un dictionnaire sur support papier sont classés par ordre alphabétique, et lorsqu'il y a des composés avec deux mots séparés, cet ordre peut être modifié dans la mesure où ces mots sont classés dans une sous-entrée. Cependant, ce classement n'aboutit pas au résultat que nous attendons. En effet, notre base de données ne contient que des composés, et notre objectif est de rassembler visuellement ceux ayant les mêmes éléments sans que des composés formés à partir d'autres éléments s'y intercalent.

Pour illustrer ces difficultés, nous prenons comme exemple les composés commençant par l'élément *bull* que nous voulons rassembler ; l'ordre alphabétique du dictionnaire ne suffit pas. En effet, deux problèmes se posent : l'un concerne la notation graphique des composés en un seul mot, avec trait d'union ou en deux

mots séparés. Nous devons décider si nous présentons ces composés en les rassemblant par type graphique (qui présente l'inconvénient de séparer les mêmes composés avec des variantes graphiques, mais qui est satisfaisant sur le plan esthétique et de la clarté). L'autre concerne l'intrusion possible d'autres éléments – en l'occurrence, les composés formés à partir de l'élément *bullet* et de l'élément *bulletin* qui s'intercalent entre les différents composés formés à partir de *bull*. Nous montrons dans le tableau ci-après l'ordre de la version papier de LPD, ainsi que deux autres possibilités de tri. Pour bien cerner le problème, nous incluons les mots non composés correspondant aux entrées principales (*bull*, *Bull* et *bulletin*) sous lesquelles sont classés les composés non compacts (en sous-entrée, et donc indiqués par * dans la colonne concernant LPD et présentés avec retrait d'alinéa) ainsi que les mots simples sans sous-entrée, mais formant le premier élément d'un composé (*bullet*). Les composés avec *bull* sont en caractères gras, les autres composés sont en caractères normaux, et les éléments non composés sont en italiques.

Figure 3.9 : Présentation des composés selon des tris différents

LPD 1990	composés : tri alphabétique	composés : tri alphabétique et graphique par élément
Bull, Bull	**bull terrier**	*bull, Bull*
bull terrier*	*bull, Bull*	**bull terrier**
Bulldog	**Bulldog**	**bull-nosed**
bulldog clip*	**bulldog clip**	**bull's-eye**
bulldoze	**Bulldoze**	**bulldog**
bulldozer	**Bulldozer**	**bulldog clip**
bullet	*Bullet*	**bulldoze**
bullet-headed	bullet-headed	**bulldozer**
bulletin	*bulletin*	**bullfight**
bulletin board*	bulletin board	**bullfighter**
bu''etproof	bulletproof	**bullfinch**
bullfight	**bullfight**	**bullfrog**
bullfighter	**bullfighter**	**bullhead**
bullfinch	**bullfinch**	**bullhorn**
bullfrog	**bullfrog**	**bullnecked**
bullhead	**bullhead**	**bullpen**
bullhorn	**bullhorn**	**bullring**
bullnecked	**bullnecked**	**bullroarer**
bull-nosed	**bull-nosed**	**bullshit**
bullpen	**bullpen**	**bullwhip**
bullring	**bullring**	*bullet*
bullroarer	**bullroarer**	bullet-headed
bull's-eye	**bull's-eye**	bulletproof
bullshit	**bullshit**	*bulletin*
bullwhip	**bullwhip**	bulletin board

L'astérisque * indique que le composé fait partie d'une sous-entrée, classé sous le terme qui précède dans LPD (version papier). Nous l'avons également fait figurer avec un retrait d'alinéa.

Dans le cas d'un tri alphabétique, le traitement électronique ne tient pas compte du trait d'union, donc les mots avec et sans ce séparateur sont classés indifféremment. En revanche, il repère d'abord les mots isolés et les classe avant les autres. Comme la virgule a un code ASCII inférieur à l'espace, le composé *bull terrier* est classé devant *bull, Bull*. A cette exception près, nous constatons que le

tri alphabétique génère les mêmes résultats que ceux que nous trouvons dans LPD. Le classement de la troisième colonne est assez différent. Les entrées sont triées d'après trois critères : un critère de logique interne (par élément) d'abord, puis un critère alphabétique, et enfin un critère graphique qui rassemble les mêmes graphies. Néanmoins, la question du classement du composé formé à partir d'un autre composé, comme *bulldog clip*, reste posée.

Il semble qu'il soit souhaitable de prévoir, comme pour l'indexation, un système souple afin que celui qui consulte la base de données soit en mesure de présenter les données comme il l'entend, selon les recherches qu'il souhaite entreprendre.

3.3 Conclusion

Dans ce chapitre, notre démarche a été très explicite car les données du corpus sont de nature différente et elles exigent des méthodes d'analyse spécifiques. Les extraits sonores impliquent le maniement d'outils qui demandent des connaissances dans le domaine de la phonétique acoustique. C'est pourquoi nous avons consacré la première partie de ce chapitre à ces considérations. Dans la seconde partie, nous avons exposé la mise en place de la base de données, sa structure interne ainsi que les moyens d'accéder aux données et de les présenter. Ces explications sont nécessaires comme préalables aux traitements automatiques, afin d'être en mesure de mener à bien les analyses et d'évaluer la portée des résultats.

Dans le chapitre suivant, nous allons étudier les règles associées aux composés en mettant en œuvre les méthodes d'analyse que nous avons proposées.

Notre corpus est constitué de composés lexicalisés, mais aussi de nouvelles formations. Comme le dit Jensen (Jensen, 1993 : 21), la productivité a non seulement besoin de règles morphologiques mais aussi de règles phonologiques afin que le locuteur soit en mesure de prédire la prononciation de ces créations.

> Speakers can generally produce a correct pronunciation of unfamiliar words (often, in fact, several acceptable pronunciations) and furthermore create new words in accordance with the morphological patterns of the language, to which an acceptable pronunciation can be given. The creation of new words to correspond to new ideas and technological advances is the norm in most societies. Thus, the set of words is no more finite than the number of utterances. Even if creativity in the lexical arena is rarer than in sentence construction, it is still clear that we need phonological rules to predict the pronunciation of new words. We can assume that these same rules play a rôle, even if only a passive one, in the pronunciation of familiar old ones as well.

En effet, nous étudierons le corpus d'analyse afin de tirer des conclusions en ce qui concerne les règles de l'accentuation des composés et de mieux comprendre les règles qui nous permettent de prédire l'accentuation des nouveaux mots.

ANNEXES DU CHAPITRE 3

Annexe A : Analyse acoustique : tableaux récapitulatifs des mesures pour l'échantillon des extraits sonores en *dark* **357**

Annexe B : Codes utilisés par PRAAT pour la phonie **365**

Annexe A: Analyse acoustique: tableaux récapitulatifs des mesures pour l'échantillon des extraits sonores en dark

Les tableaux suivants récapitulent les mesures prises pour les extraits sonores en *dark*. Nous avons scindé les résultats afin qu'ils soient plus lisibles. Pour cet échantillonnage, il y a trois tableaux correspondant à l'étude qui est faite dans le chapitre 3 : les mesures du temps ; les mesures du F_0 ; les mesures de l'intensité. Il n'y a pas de récapitulatif par unité.

Les temps sont tous donnés en secondes (s) : le début (T-début), la fin (T-fin), la durée (Durée) et les temps des maxima et minima de F_0 (T- F_0Max, T-F_0Min) et de l'intensité (T-Imax, T-Imin).

Les mesures du F_0 sont en hertz (Hz), en nombres entiers pour les maxima et minima (F_0Max et F_0Min), et en demi-tons (dt), avec deux décimales pour la moyenne (F_0Mean) et l'écart type (F_0SD) ; la pente de F_0 (F_0MAS) est en demi-tons/seconde (dt/s). Nous avons également inclus dans ce tableau les mesures du temps des maxima et minima.

Les mesures de l'intensité sont en décibels (dB), en nombres entiers pour les maxima et minima (Imax et Imin) et avec deux décimales pour la moyenne (I-Mean) et l'écart type (I-SD). Nous avons également inclus dans ce tableau les mesures du temps des maxima et minima.

Le symbole « # » indique que la mesure n'a pas pu être prise, ou parce que le segment est trop brève (intensité) ou parce que la valeur n'est pas comprise dans l'échelle spécifiée.

Dans la colonne « Phon-syll », la phonie est donnée en utilisant le codage spécifique à PRAAT qui permet d'incorporer les étiquettes sur les graphiques en IPA . Ces codages se trouvent en Annexe B. Par exemple, schwa correspond à « \sw »

Les mesures pour la syllabe correspondant à l'accent primaire /1/ sont en caractères gras.

Figure III.1 : Echantillonnage des extraits sonores en *dark* : mesures du temps

Fichier son	Schéma	Phon-syll	T-début en s	T-fin en s	Durée en s	T-F_0Max en ms	T-F_0Min en s	T-Imax en s	T-Imin en s
antidark1	R	\aent	0,000	0,322	0,322	0,156	0,241	0,121	0,036
	0	i	0,322	0,472	0,150	0,347	0,447	0,357	0,437
	P	d\as:k	0,472	0,970	0,498	0,734	0,632	0,617	0,849
	1	\as:ts	0,970	1,417	0,447	1,102	1,203	1,088	1,246
	0	spel	1,417	1,957	0,541	1,627	1,837	1,646	1,922
antidark2	R	\aent	0,054	0,283	0,228	0,141	0,229	0,134	0,239
	0	i	0,283	0,417	0,134	0,305	0,395	0,328	0,385
	P	d\as:k	0,417	0,722	0,304	0,469	0,510	0,520	0,689
	1	f\ct:s	0,722	1,096	0,375	0,973	0,876	0,937	0,802
	0	spel	1,096	1,504	0,408	1,150	1,321	1,245	1,458
darkart1	P	d\as:k	0,000	0,449	0,449	0,065	0,208	0,098	0,351
	1	\as:ts	0,449	1,094	0,645	0,573	0,762	0,607	0,500
darkart2	P	d\as:k	0,000	0,476	0,476	0,239	0,023	0,190	0,397
	1	\as:ts	0,476	1,117	0,641	0,624	0,806	0,602	0,891
darkcupb	P	d\as:k	0,000	0,320	0,320	0,190	0,065	0,121	0,218
	1	k\vtb	0,320	0,604	0,284	0,432	0,582	0,486	0,368
	0	\swd	0,604	0,788	0,183	#	#	0,641	0,751
darkforce1	0	\dh\sw	0,000	0,274	0,274	0,252	0,172	0,226	0,037
	1	d\as:k	0,274	0,657	0,383	0,635	0,505	0,369	0,598
	3	f\ct s	0,657	0,981	0,324	0,759	0,839	0,772	0,707
	0	\ics	0,981	1,261	0,280	1,006	1,036	1,016	1,226
darkmag1	1	d\as:k	0,000	0,431	0,431	0,256	0,106	0,157	0,374
	3	m\aed\zh	0,431	0,727	0,297	0,461	0,614	0,563	0,633
	0	\ick	0,727	0,948	0,220	0,788	0,808	0,896	0,852
darkmag2	P	d\as:k	0,000	0,374	0,374	0,098	0,232	0,113	0,306
	1	m\aed\zh	0,374	0,693	0,319	0,536	0,608	0,532	0,639
	0	\ick	0,693	0,797	0,104	#	#	0,725	0,765
darknge1	1	d\as:k	0,000	0,298	0,298	0,034	0,224	0,106	0,251
	3	e\ed\zh	0,298	0,539	0,241	0,319	0,489	0,348	0,504
	0	\icz	0,539	0,744	0,205	0,561	0,601	0,571	0,641
darksid1	P	d\as:k	0,000	0,346	0,346	0,213	0,123	0,051	0,258
	1	sa\ed	0,346	0,823	0,477	0,799	0,709	0,405	0,754
darksid2	P	d\as:k	0,000	0,304	0,304	0,042	0,162	0,086	0,272
	1	sa\ed	0,304	0,808	0,504	0,459	0,459	0,475	0,757
darksid3	1	d\as:k	0,000	0,332	0,332	0,041	0,171	0,079	0,267
	0	sa\ed	0,332	0,806	0,474	0,464	0,534	0,461	0,719
darkwiz1A	P	d\as:k	0,000	0,415	0,415	0,254	0,092	0,153	0,334
	3	w\icz	0,415	0,653	0,239	0,543	0,629	0,538	0,619
	0	\swd	0,653	0,869	0,216	0,676	0,846	0,698	0,836
	1	gr\end	0,869	1,200	0,331	1,060	1,180	0,998	0,911
	0	\swl	1,200	1,328	0,128	1,224	1,304	1,234	1,294
	0	w\ct:ld	1,328	1,739	0,411	1,636	1,490	1,462	1,703
darkwiz1	1	d\as:k	0,000	0,415	0,415	0,254	0,092	0,153	0,334
	3	w\icz	0,415	0,653	0,239	0,543	0,629	0,538	0,619
	0	\swd	0,653	0,869	0,216	0,676	0,846	0,698	0,836

Fichier son	Schéma	Phon-syll	T-début en s	T-fin en s	Durée en s	T-F_0Max en s	T-F_0Min en s	T-Imax en s	T-Imin en s
darkwiz1B	P	d\as:k	0,000	0,516	0,516	0,317	0,185	0,203	0,050
	P	w\icz	0,516	0,813	0,296	0,600	0,649	0,685	0,581
	0	\swdz	0,813	1,045	0,232	0,834	0,984	0,849	0,991
	0	\ct	1,045	1,201	0,156	1,120	1,178	1,104	1,168
	1	w\ict\sh	1,201	1,453	0,252	1,281	1,222	1,283	1,358
	0	\icz	1,453	1,713	0,260	1,473	1,623	1,488	1,678
darkwiz2	P	d\as:k	0,000	0,333	0,333	0,062	0,232	0,115	0,297
	1	w\icz	0,333	0,542	0,209	0,460	0,357	0,420	0,507
	0	\swd	0,542	0,664	0,122	0,563	0,643	0,578	0,628
defdark1	0	d\ic	0,000	0,131	0,131	0,060	0,110	0,073	0,035
	P	fens	0,131	0,666	0,536	0,384	0,444	0,310	0,181
	0	\sw	0,666	0,778	0,111	0,687	0,757	0,702	0,742
	P	genst	0,778	1,155	0,377	0,943	0,873	0,851	1,121
	0	\dh\sw	1,155	1,351	0,196	1,198	1,318	1,227	1,318
	P	d\as:k	1,351	1,744	0,393	1,562	1,437	1,440	1,649
	1	\as:ts	1,744	2,399	0,656	1,896	2,036	1,909	2,172
defdark2	0	d\ic	0,000	0,155	0,155	0,053	0,103	0,066	0,123
	P	fens	0,155	0,511	0,356	0,228	0,378	0,251	0,188
	0	\sw	0,511	0,619	0,108	0,535	0,595	0,545	0,585
	P	genst	0,619	0,957	0,338	0,653	0,793	0,673	0,815
	0	\dh\sw	0,957	1,106	0,149	0,982	1,052	1,012	1,072
	P	d\as:k	1,106	1,476	0,370	1,451	1,281	1,214	1,389
	1	\as:ts	1,476	2,058	0,582	1,517	1,647	1,587	1,839
defdark3	0	d\ic	0,000	0,125	0,125	0,073	0,103	0,086	0,040
	P	fens	0,125	0,570	0,444	0,372	0,296	0,303	0,209
	0	\sw	0,570	0,661	0,091	0,590	0,640	0,605	0,625
	P	genst	0,661	0,923	0,262	0,789	0,822	0,717	0,833
	0	\dh\sw	0,923	1,068	0,145	0,945	1,015	0,966	1,035
	P	d\as:k	1,068	1,442	0,374	1,110	1,169	1,130	1,328
	1	\as:ts	1,442	1,991	0,549	1,506	1,636	1,130	1,948

Figure III.2 : Echantillonnage des extraits sonores en *dark* : mesures du F_0

Fichier son	Schéma	Phon-syll	F_0Max en Hz	T-F_0Max en s	F_0Min en Hz	T-F_0Min en s	F_0-MAS en dt/s	F_0-Mean en Hz	F_0-SD en dt
antidark1	R	\aent	183	0,156	143	0,241	38,81	9,29	1,41
	0	i	170	0,347	122	0,44	57,68	6,77	2,10
	P	d\as:k	156	0,734	147	0,632	9,59	7,18	0,37
	1	\as:ts	162	1,102	145	1,203	17,80	7,62	0,64
	0	spel	124	1,627	77	1,837	34,94	-1,74	2,93
antidark2	R	\aent	144	0,141	120	0,229	28,40	5,40	1,06
	0	i	130	0,305	104	0,395	42,39	2,88	1,35
	P	d\as:k	141	0,469	137	0,510	5,68	5,70	0,14
	1	f\ct:s	139	0,973	129	0,876	9,56	4,95	0,50
	0	spel	98	1,150	81	1,321	17,47	-2,41	0,89
darkart1	P	d\as:k	125	0,065	110	0,208	10,93	2,37	0,73
	1	\as:ts	123	0,573	78	0,762	39,47	0,41	2,71
darkart2	P	d\as:k	317	0,239	183	0,023	34,93	18,00	2,52
	1	\as:ts	304	0,624	203	0,806	31,30	17,60	2,00
darkcupb	P	d\as:k	93	0,190	83	0,065	14,30	-2,26	0,72
	1	k\vtb	98	0,432	79	0,582	25,30	-2,05	1,29
	0	\swd	#	#	#	#	#	#	#
darkforce1	0	\dh\sw	185	0,252	160	0,172	31,00	9,65	0,97
	1	d\as:k	557	0,635	138	0,505	185,49	21,48	13,73
	3	f\ct:s	106	0,759	79	0,839	62,88	-1,69	1,77
	0	\ics	83	1,006	78	1,036	39,22	-3,75	0,51
darkmag1	1	d\as:k	209	0,256	133	0,106	45,12	8,95	3,00
	3	m\aed\zh	96	0,461	77	0,614	23,11	2,51	1,55
	0	\ick	395	0,788	393	0,808	4,15	23,75	0,04
darkmag2	P	d\as:k	109	0,098	101	0,232	8,38	0,96	0,49
	1	m\aed\zh	101	0,536	89	0,608	22,43	-0,68	0,81
	0	\ick	#	#	#	#	#	#	#
darkage1	1	d\as:k	163	0,034	121	0,224	26,93	6,85	1,53
	3	e\ed\zh	128	0,319	99	0,489	25,65	1,94	1,36
	0	\tcz	115	0,561	110	0,601	16,94	1,99	0,27
darksid1	P	d\as:k	476	0,213	117	0,123	269,57	17,76	11,87
	1	sa\ed	119	0,799	98	0,709	37,51	0,80	1,38
darksid2	P	d\as:k	107	0,042	83	0,162	35,55	-1,41	1,44
	1	sa\ed	296	0,459	88	0,459	116,05	5,68	7,37
darksid3	1	d\as:k	196	0,041	178	0,171	12,37	11,15	0,53
	0	sa\ed	110	0,464	94	0,534	39,32	0,00	0,97
darkwiz1A	P	d\as:k	160	0,254	108	0,092	38,20	5,31	2,50
	3	w\icz	123	0,543	107	0,629	22,02	2,55	0,77
	0	\swd	135	0,676	99	0,846	31,21	2,81	1,80
	1	gr\iend	195	1,060	110	1,180	66,66	8,22	3,01
	0	\swl	101	1,224	86	1,304	34,87	-1,28	1,00
	0	w\ct:ld	139	1,636	84	1,490	31,14	0,27	3,50
darkwiz1	1	d\as:k	160	0,254	108	0,092	38,20	5,31	2,50
	3	w\icz	123	0,543	107	0,629	22,02	2,55	0,77
	0	\swd	135	0,676	99	0,846	31,21	2,81	1,80

Fichier son	Schéma	Phon-syll	F_0Max en Hz	T-F_0Max en s	F_0Min en Hz	T-F_0Min en s	F_0-MAS en dt/s	F_0-Mean en Hz	F_0-SD en dt
darkwiz1B	P	d\as:k	141	0,317	127	0,185	15,26	5,17	0,65
	P	w\icz	143	0,600	126	0,649	31,96	4,85	0,64
	0	\swdz	169	0,834	117	0,984	42,03	6,46	2,18
	0	\ct	157	1,120	148	1,178	14,93	7,48	0,33
	1	w\ict\sh	139	1,281	121	1,222	42,72	4,67	0,85
	0	\icz	108	1,473	85	1,623	26,71	-0,33	1,21
darkwiz2	P	d\as:k	138	0,062	131	0,232	5,56	5,30	0,26
	1	w\icz	161	0,460	115	0,357	51,00	6,04	1,95
	0	\swd	[illegible]	0,563	77	0,643	64,58	-2,20	1,90
defdark1	0	d\ic	107	0,060	86	0,110	77,62	-0,61	1,50
	P	fens	157	0,384	132	0,444	26,85	6,75	0,81
	0	\sw	128	0,687	105	0,757	50,33	2,55	1,29
	P	genst	124	0,943	116	0,873	15,66	3,16	0,42
	0	\dh\sw	111	1,198	79	1,318	49,05	-0,62	2,08
	P	d\as:k	124	1,562	116	1,437	10,87	3,14	0,38
	1	\as:ts	117	1,896	85	2,036	39,22	0,15	2,00
defdark2	0	d\ic	111	0,053	89	0,103	75,31	0,09	1,45
	P	fens	123	0,228	95	0,378	29,39	0,86	1,33
	0	\sw	101	0,535	85	0,595	49,84	-1,29	1,12
	P	genst	102	0,653	87	0,793	20,41	-1,33	0,84
	0	\dh\sw	100	0,982	82	1,052	48,82	-1,46	1,31
	P	d\as:k	527	1,451	86	1,281	113,58	5,09	12,28
	1	\as:ts	110	1,517	82	1,647	38,86	-0,65	1,66
defdark3	0	d\ic	96	0,073	92	0,103	28,11	-1,08	0,36
	P	fens	130	0,372	112	0,296	28,43	3,10	0,99
	0	\sw	107	0,590	93	0,640	46,06	-0,03	0,89
	P	genst	106	0,789	103	0,822	9,79	0,80	0,17
	0	\dh\sw	104	0,945	80	1,015	65,22	-1,33	1,91
	P	d\as:k	109	1,110	103	1,169	7,72	0,69	0,26
	1	\as:ts	105	1,506	79	1,636	29,51	-1,99	1,62

Figure III.3 : Echantillonnage des extraits sonores en *dark* : mesures de l'intensité

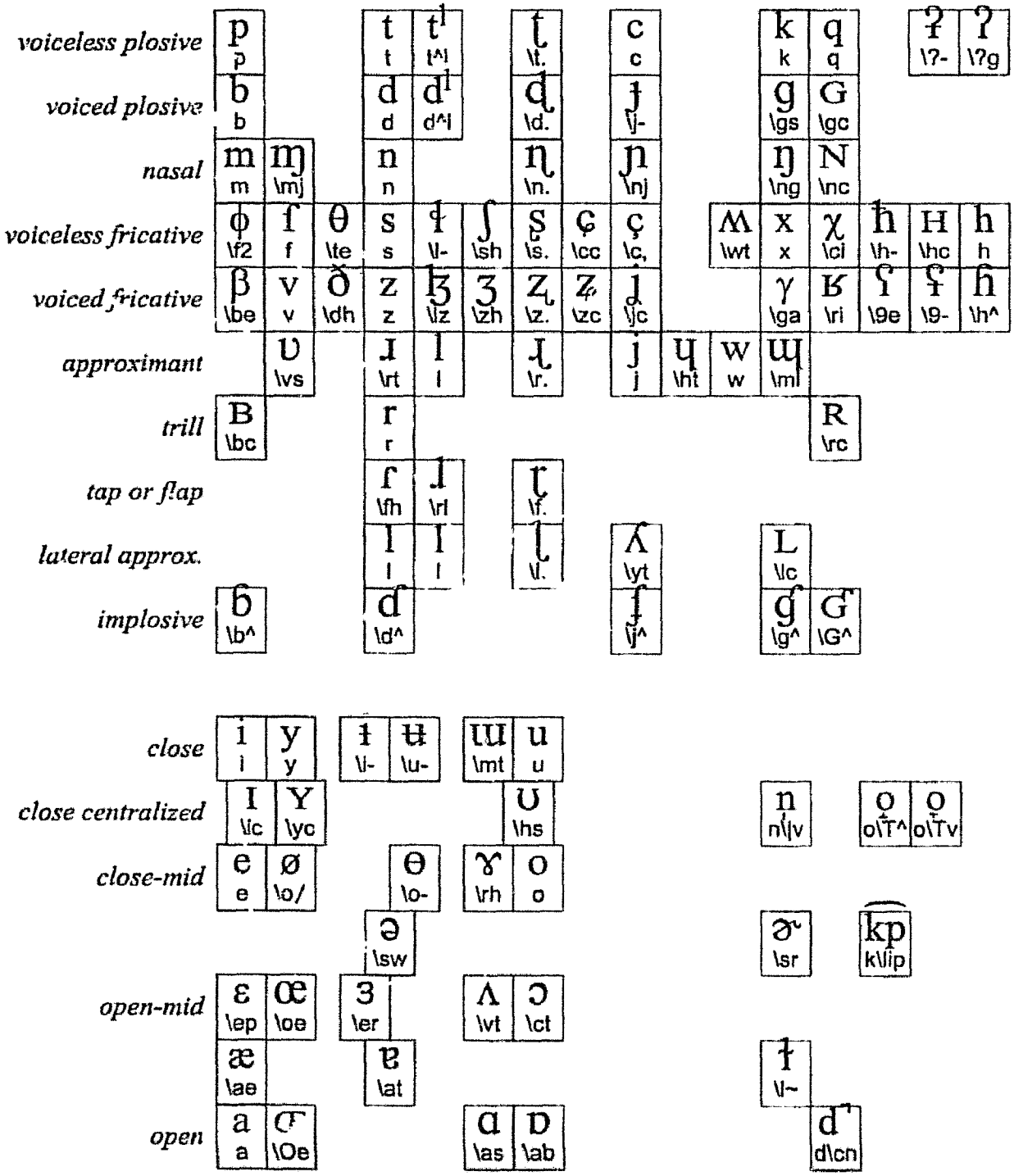
Fichier son	Schéma	Phon-syll	Imax en dB	T-Imax en s	Imin en dB	T-Imin en s	I-Mean en dB	I-SD en dt
antidark1	R	\aent	83	0,121	51	0,036	71,33	9,15
	0	i	73	0,357	66	0,437	70,11	2,10
	P	d\as:k	87	0,617	42	0,849	69,70	15,14
	1	**\as:ts**	**83**	**1,088**	**48**	**1,246**	**71,17**	**10,92**
	0	spel	76	1,646	41	1,922	63,61	7,57
antidark2	R	\aent	81	0,134	58	0,219	72,43	6,54
	0	i	76	0,328	64	0,385	72,48	4,97
	P	d\as:k	84	0,520	49	0,689	72,10	13,36
	1	**f\ct:s**	**80**	**0,937**	**54**	**0,802**	**70,21**	**8,60**
	0	spel	76	1,245	42	1,458	64,55	10,31
darkart1	P	d\as:k	84	0,098	48	0,351	72,07	12,04
	1	**\as:ts**	**80**	**0,607**	**47**	**0,500**	**66,51**	**10,29**
darkart2	P	d\as:k	89	0,190	47	0,397	75,18	14,14
	1	**\as:ts**	**87**	**0,602**	**36**	**0,891**	**71,45**	**14,45**
darkcupb	P	d\as:k	80	0,121	49	0,218	67,27	11,84
	1	**.\vtb**	**76**	**0,486**	**51**	**0,368**	**64,98**	**7,68**
	0	\swd	70	0,641	52	0,751	63,07	6,60
darkforce1	0	\dh\sw	85	0,226	40	0,037	57,27	15,08
	1	**d\as:k**	**89**	**0,369**	**41**	**0,598**	**70,87**	**15,49**
	3	f\ct:s	76	0,772	57	0,707	66,31	6,32
	0	\ics	70	1,016	53	1,226	61,69	4,09
darkmag1	1	d\as:k	85	0,157	46	0,374	73,62	12,05
	3	m\aed\zh	76	0,563	56	0,633	64,84	7,40
	0	\ick	65	0,896	43	0,852	54,18	8,13
darkmag2	P	d\as:k	83	0,113	46	0,306	71,18	13,19
	1	m\aed\zh	77	0,532	52	0,639	67,63	7,66
	0	\ick	63	0,725	52	0,765	58,32	4,22
darkage1	1	d\as:k	83	0,106	59	0,251	76,04	7,98
	3	e\ied\zh	81	0,348	62	0,504	73,64	7,22
	0	\icz	71	0,571	63	0,641	65,40	2,33
darksid1	P	d\as:k	73	0,051	43	0,258	61,48	9,20
	1	sa\icd	71	0,405	51	0,754	61,51	5,27
darksid2	P	d\as:k	78	0,086	42	0,272	66,15	11,90
	1	sa\icd	71	0,475	50	0,757	63,52	5,81
darksid3	1	d\as:k	86	0,079	54	0,267	74,19	12,36
	0	sa\icd	73	0,461	38	0,719	58,77	9,77
darkwiz1A	P	d\as:k	86	0,153	45	0,334	73,58	12,32
	3	w\icz	81	0,538	67	0,619	74,48	5,07
	0	\swd	80	0,698	57	0,836	69,04	8,13
	1	gr\t:nd	84	0,998	47	0,911	74,49	9,74
	0	\swl	78	1,234	71	1,294	75,22	2,50
	0	w\ct:ld	77	1,462	61	1,703	72,74	4,23
darkwiz1	1	d\as:k	86	0,153	45	0,334	73,58	12,32
	3	w\icz	81	0,538	67	0,619	74,48	5,07
	0	\swd	80	0,698	57	0,836	69,04	8,13

Fichier son	Schéma	Phon-syll	Imax	T-Imax	Imin	T-Imin	I-Mean	I-SD
darkwiz1B	P	d\as:k	86	0,203	38	0,050	68,93	17,35
	P	w\icz	83	0,685	53	0,581	69,93	9,17
	0	\swdz	78	0,849	61	0,991	68,98	5,76
	0	\ct	84	1,104	79	1,168	82,04	1,42
	1	w\ict\sh	78	1,283	44	1,358	64,42	10,55
	0	\icz	72	1,488	56	1,678	66,82	4,80
darkwiz2	P	d\as:k	87	0,115	54	0,297	77,79	9,62
	1	w\icz	83	0,420	69	0,507	77,23	5,55
	0	\swd	76	0,578	63	0,628	67,51	4,99
defdark1	0	d\ic	72	0,073	57	0,035	67,71	5,43
	P	fens	82	0,310	52	0,181	65,24	8,92
	0	\sw	74	0,702	63	0,742	68,26	4,90
	P	genst	81	0,851	38	1,121	63,68	12,01
	0	\dh\sw	75	1,227	55	1,318	66,12	7,43
	P	d\as:k	83	1,440	48	1,649	69,35	12,71
	1	\as:ts	80	1,909	34	2,172	60,03	12,81
defdark2	0	d\ic	78	0,066	58	0,123	70,64	6,96
	P	fens	85	0,251	54	0,188	69,05	9,39
	0	\sw	78	0,545	64	0,585	71,64	5,80
	P	genst	79	0,673	49	0,815	65,08	9,48
	0	\dh\sw	75	1,012	57	1,072	67,76	7,23
	P	d\as:k	80	1,214	50	1,389	68,91	10,80
	1	\as:ts	79	1,587	40	1,839	61,17	11,23
defdark3	0	d\ic	73	0,086	55	0,040	65,31	7,09
	P	fens	83	0,303	53	0,209	65,56	9,48
	0	\sw	75	0,605	67	0,625	71,24	4,36
	P	genst	80	0,717	57	0,833	69,82	7,40
	0	\dh\sw	74	0,966	60	1,035	67,08	5,59
	P	d\as:k	83	1,130	46	1,328	69,48	12,35
	1	\as:ts	83	1,130	43	1,948	64,99	11,03

Annexe B : Codes utilisés par PRAAT pour la phonie

Phonetic symbol chart

Phonetic symbol chart



CHAPITRE 4

Chapitre 4 : Règles d'accentuation des composés : analyses et résultats

Introduction

Nous avons expliqué les méthodes qui seront utilisées afin d'analyser les données que nous avons rassemblées, mais avant de procéder à cette étape, nous souhaitons d'abord rappeler la nomenclature que nous avons adoptée, puis commenter brièvement les principes généraux qui gouvernent l'accentuation des composés. Ensuite, nous passerons à l'étude détaillée de notre base de données, d'abord en ce qui concerne l'accent lexical, puis à partir des extraits sonores. Les résultats de ces analyses nous permettront de tirer des conclusions sur les règles déjà énoncées et leur application dans un contexte de discours.

4.1 Rappel de la nomenclature

Dans le chapitre 1, nous avons passé en revue les différentes appellations et notations de l'accent. En ce qui concerne les composés, la question se pose de manière simple, comme nous le rappelle Roach (1991 : 99) : « As far as stress is concerned, the question is quite simple. When is primary stress placed on the first word of the compound and when on the second » ? Nous avons adopté la nomenclature suivante pour distinguer ces deux types d'accentuation : « accentuation simple » lorsque l'accent primaire porte sur le premier élément, (noté /10/ dans les dictionnaires) puis « accentuation double » lorsque les deux éléments portent un accent (noté /21/ dans les dictionnaires).

Pour la notation du schéma accentuel, nous avons préservé les notations d'origine des fichiers électroniques des dictionnaires, qui ne sont pas identiques

les unes aux autres (voir 3.2.1.2.2.2.3 « Les champs de la base de données : Le schéma accentuel »). Ce manque de cohérence entre les dictionnaires est la raison pour laquelle le champ « Type de schéma » a été introduit. Il contient le code « s » ou « d » : le code « s » correspond à l'accentuation simple et le code « d » à l'accentuation double. Grâce à ce codage, nous sommes en mesure d'isoler toutes les entrées ayant une accentuation simple, par exemple.

Pour la table LPD saisie manuellement (les composés non compacts), nous avons repris la même notation que celle du LPD électronique afin de garder la cohérence entre les données d'une même source, car les deux tables proviennent de la première édition de LPD. En revanche, une notation du schéma accentuel bornée a été rajoutée pour la table saisie manuellement : le schéma de *potato cake*, /0100/ est également noté /010#0/ pour indiquer clairement les bornes morphologiques. Grâce à cette indication, il est possible de connaître le nombre de syllabes de chaque élément dans cette table, alors qu'il n'est pas possible de disposer d'informations sur les éléments individuels dans les autres tables, du moins pour tout composé de 3 syllabes ou plus.

Enfin, pour les extraits sonores, comme nous l'avons expliqué dans le chapitre 3 (3.1.3.1.2), nous avons adopté la notation proposée à la fin du chapitre 1 afin de tester son efficacité :

/1/ : syllabe susceptible de porter la chute intonative /1/

/2/ : accent secondaire, rythmique, sur un élément plurisyllabique (sur un élément portant l'accent /1/)

/3/ : syllabe accentuée du deuxième élément d'un composé à accentuation simple

/P/ : syllabe accentuée du premier élément d'un composé à accentuation double

/R/ accent secondaire, rythmique sur un élément plurisyllabique (sur un élément portant l'accent /P/ ou l'accent /3/)

Toutefois, dans la discussion qui suit, nous nous référerons à la notation utilisée par le dictionnaire LPD qui utilise /1/, /2/, /3/ et /0/. En revanche, nous reprendrons systématiquement la nomenclature que nous avons rappelée ci-dessus.

4.2 Règles pour l'accentuation des composés : principes généraux

Dans cette section, nous ferons un bref rappel des différentes manières qui existent pour classer la formation des composés, car les règles de l'accentuation sont parfois reliées à ces partis pris théoriques. Puis, nous commenterons les règles qui ont été élaborées pour prédire et expliquer les schémas de l'accent lexical. Enfin, nous nous intéresserons aux règles concernant le comportement des composés en discours et au phénomène du déplacement de l'accent.

4.2.1 La formation des composés

Dans les rubriques suivantes, nous tentons de résumer les travaux des lexicologues en ce qui concerne la formation des composés. Nous avons choisi de les traiter sous deux angles : syntaxique et sémantique, que nous retrouverons dans la partie suivante qui concerne plus précisément les règles de l'accentuation. Nous tenons à préciser que beaucoup de chercheurs tiennent compte de ces deux axes, mais nous préférons simplifier ce débat extrêmement riche qui est un préalable à notre analyse, mais dont les implications théoriques dépassent le cadre de nos travaux.

4.2.1.1 L'approche syntaxique

Toute référence à la catégorie syntaxique des éléments du composé implique une approche syntaxique. Ainsi, la théorie générative s'en inspire largement. Les travaux de Liberman & Prince (1977), qui s'intéressent aux composés formés à partir de verbes, expliquent la formation des composés à l'aide d'arbres étiquetés selon la catégorie qui peuvent ensuite être utilisés afin d'expliquer les schémas accentuels. Ces mêmes outils sont adoptés par d'autres (Downing, 1977, Roeper & Siegel, 1978).

Bauer, dans son ouvrage de 1983 sur la formation des mots en anglais, fait l'inventaire de toutes les combinaisons existant entre les différentes catégories des éléments qui peuvent s'associer pour former des composés. Il souligne également la difficulté de déterminer de quelle catégorie relèvent certains éléments à cause du processus de conversion entre nom et verbe. Dans son analyse, il ne se limite pas à ce travail, mais démontre la très grande variété de ces combinaisons.

Un classement des composés qui se limite à la catégorie syntaxique du composé lui-même, ou à un sous-classement selon la catégorie des éléments, ne peut pas rendre compte des relations plus générales qui existent et que l'on peut tenter d'exprimer en adoptant une autre approche en tenant compte du sens des relations. Passons à l'étude de ce que nous avons appelé « l'approche sémantique » et qui inclut les classements autres que purement syntaxiques.

4.2.1.2 L'approche sémantique

La plupart des lexicologues se réfèrent, du moins en partie, à la catégorie syntaxique des éléments afin d'établir un classement des composés, mais ils font

référence à d'autres types de catégorisations qui sont plus ou moins complexes. Le plus simple est celui qui distingue les composés radicaux (« root compounds ») des composés formés à partir de verbes (« verbal compounds » ou « synthetic compounds »). C'est la distinction, nous l'avons vu, que font Roeper & Siegel (1978), mais aussi Marchand (1969)[1].

Un autre classement qui tient davantage compte des catégories sémantiques est celui que nous rappelle Kuiper (1999 : 411). Il cite les quatre catégories qui sont généralement considérées comme fondamentales : « root compounds, co-ordinated (or *dvandva*) compounds, exocentric (or *bahuvrihi*) compounds and synthetic compounds ». La première catégorie des composés radicaux (« root compounds ») comprend les unités dont le référent correspond au deuxième élément du composé : par exemple, *wallpaper* ; la seconde, les composés co-ordonnés (« co-ordinated or *dvandva* compound ») comprend les formations où les deux éléments sont d'égale importance : *boy-king*, par exemple où le référent est à la fois *boy* et *king* (aussi appelé « appositional compound » par Adams (1973)) ; la troisième catégorie (les *bahuvrihis*) comprend les composés dont le second élément n'est pas le référent : *redskin*, par exemple, qui ne désigne pas une sorte de peau mais une personne ayant ce type de peau ; enfin, la quatrième catégorie, comme nous l'avons vu, comprend tous les composés qui sont formés à partir

---

[1] Marchand (1969) part du principe qu'un composé peut se gloser par un syntagme et il en distingue deux types : les composés primaires, formés à partir de relations sémantiques, et les composés synthétiques, formés à partir de relations syntagmatiques, comme *watchmaker* où le composé comprend le sujet, le prédicat et l'objet. Dans son analyse, il utilise les rapports sujet / objet afin d'expliquer les différentes formations, mais il fait référence à bien d'autres facteurs, dont l'accentuation.

d'un verbe, donc en anglais les composés en *-ed*, *-ing*, *-er* tels que *troubleshooter / troubleshooting*. Ces distinctions ont été utilisées par Jespersen (1961)[2], Bauer (1983)[3] et Tournier (1985)[4], qui y ont également rajouté d'autres catégories. Adams (1973)[5], elle, fait un classement encore plus complet qui tient compte et des catégories syntaxiques et des relations sémantiques.

Malgré l'intérêt que présentent ces travaux, nous nous limiterons à des analyses comprenant les quatre éléments traditionnels, en y ajoutant la catégorie des composés réduplicatifs, citée également par Jespersen (1961), Marchand (1969), Tournier (1980), et étudiée de manière détaillée par Thuns (1963), ainsi que les formations en 's. Nous tiendrons compte également des catégories syntaxiques du composé lui-même.

4.2.2 L'accentuation des composés

Nous verrons que les deux approches que nous avons abordées sous la rubrique précédente ont des implications pour les règles de l'accentuation elles-mêmes, mais avant de les étudier, nous voulons d'abord rappeler quelques généralités en ce qui concerne l'accentuation des composés.

Comme le rappelle Wells (Wells, 1990: ix) dans l'introduction de son dictionnaire, il est notoirement difficile de prédire des règles d'accentuation pour

---

[2] Voir Jespersen, 1961: 107-138.

[3] Voir Bauer, 1983: 201-236.

[4] Voir Tournier, 1985: 134-186.

[5] Voir Adams, 1973: 57-135.

les composés : « The pronunciation of compound words can often be derived from their component parts. Stress patterns, however, are not always easy to predict and so important compounds and their patterns are listed after the main entry ». Nous reviendrons ultérieurement sur les problèmes de la sélection des composés à intégrer dans le dictionnaire de prononciation. Dans cette section, nous allons tenter d'appréhender les principes généraux qui gouvernent ces règles afin de pouvoir mieux les évaluer dans les analyses.

Ces règles font souvent appel aux cinq critères de définition des composés dont nous avons discuté dans le chapitre 2 (2.1.1.1). Ces critères s'appuient sur des considérations sémantiques, syntaxiques, graphiques, phonétiques et phonologiques, et enfin, accentuelles. Nous avons déjà fait largement référence à ces considérations dans le chapitre 2, mais il sera parfois utile d'y revenir dans la discussion des règles. Dans l'analyse qui suivra nous évaluerons de manière systématique leur importance relative pour expliquer les caractéristiques de l'accentuation des composés.

D'après Boisson (1980), les deux types d'accentuation que nous rencontrons dans les composés sont le reflet du comportement accentuel mixte de la langue dans son ensemble[6]. L'accentuation simple dans les composés correspond à l'accent de mot, et l'accentuation double, elle, correspond à l'accent de phrase. Il explique (Boisson, 1980 : 235) :

---

[6] Voir aussi les remarques en 1.2 « Accent et compréhension » où Boisson oppose le comportement de l'accentuation et celui de l'intonation.

> Le fait que l'anglais soit devenu une langue mixte parce que sa composition appartient à un type et sa syntaxe à un autre type, explique sans doute son comportement accentuel [...] En effet, dans la majorité des composés (spécialement nominaux), l'accent est à gauche, sur le premier élément du composé. Au contraire, « l'accent de phrase » (le noyau du groupe phonologique, qui est à la phrase ce que l'accent est au mot), lui, est situé tout à droite. On a :
>
> blackbird / the bird is black.
>
> L'anglais a donc un comportement de sa proéminence qui est « incohérent » parce que son comportement pour l'ordre des éléments est incohérent.

Ces remarques ont une importance certaine en ce qui concerne l'analyse de l'accent en discours, et nous reviendrons à ces propos dans l'analyse acoustique ; mais elles sont également significatives pour les règles élaborées pour l'accent lexical des composés, et notamment la manière dont les génératives abordent la question.

4.2.2.1 L'approche syntaxique : le mécanisme de l'accent

Chez les génératives, ce comportement double est directement traduit par le fait d'avoir deux règles d'accentuation différentes. Dès Chomsky & Halle (1968), la règle appelée *Compound Stress Rule* (CSR) est opposée à la *Nuclear Stress Rule* (NSR) (voir 1.5.1 « L'approche de *The Sound Pattern of English* » pour une explication précise du fonctionnement de ces règles). Elles s'appliquent dans des contextes différents et ont des effets contraires en ce qui concerne l'attribution de l'accent : la CSR conserve l'accent sur le premier élément, alors que la NSR attribue l'accent principal sur le dernier élément. C'est donc la CSR qui s'applique aux composés à accentuation simple et la NSR qui s'applique aux composés à accentuation double. La question de savoir comment les distinguer n'est pas résolue, ni même étudiée, dans cet ouvrage. Des contrastes de type *Fifth Avenue*, /21/, *Fifth Street*, /12/, sont cités, et les auteurs suggèrent que pour les différencier, il faudrait un étiquetage différent, mais ils reconnaissent que les

critères de cet étiquetage ne sont pas déterminés (Chomsky & Halle, 1968 : 156) (souligné par nous) :

> [...] a feature specification boundary between the constituents, in which case the rule can be limited to boundaries not containing this feature. This, obviously, does not solve the general problem, but serves only to eliminate it from the domain of phonology. The problem remains of determining under what syntactic conditions this feature is or is not present.

Finalement, Chomsky & Halle se contentent de considérer les composés à accentuation double comme des syntagmes ; seuls les composés à accentuation simple sont considérés effectivement comme des composés. Ils expliquent leur parti pris théorique (Chomsky & Halle, 1968 : 156) :

> In fact, the general problem certainly belongs in part to syntax, in part to the readjustment component, rather than to phonology proper, and it can be clarified and resolved only by an investigation of the conditions, syntactic and other, under which the Compound Rule is applicable. For this reason, we will make no attempt to go more deeeply into the question here. We have throughout been limiting ourselves arbitrarily to problems of phonological interpretation, and are making no attempt in the present study to investigate the processes by which the syntactic component of the grammar forms the surface structures that are phonetically interpreted by the rules we have been discussing here. Because of this limitation of scope, we will simply leave this question in its present unsatisfactory state.

Cette méthode d'analyse est très différente de la nôtre. En effet, dans la constitution de notre base de données, nous avons adopté une démarche beaucoup moins restrictive car notre travail s'appuie d'abord sur les unités lexicales elles-mêmes, en tenant compte de leur valeur sémantique, avant de savoir quelle en est leur accentuation. En revanche, l'école générative considère que des unités lexicales ayant une accentuation différente appartiennent à des catégories syntaxiques différentes, car la théorie prend en compte des considérations syntaxiques et accentuelles pour déterminer l'étiquetage des unités, ce qui, à son tour, détermine laquelle des deux règles s'applique.

Leur approche s'appuie donc largement sur la syntaxe et ne tient aucun compte du sémantisme des éléments. Fudge (1984 : 136) étudie cette hypothèse, et tente de déterminer l'association qui existe entre un type de construction et un type d'accentuation. Il étudie des paires de composés tels que *Christmas cake / Christmas pudding*, où le schéma accentuel est différent, mais la formation est apparemment identique ; il s'agit d'un composé N + N dans les deux cas :

> This [...the fact that there is a second type of compound with final stress] means that it is important to investigate whether there are any general principles governing which stress patterns are assigned to various construction-types; if both initial stress and final stress can be associated with certain constructions [...], we must if possible say which is normal and which exceptional.

Finalement, il la rejette, et conclut que l'accentuation des composés ne peut pas s'expliquer uniquement en termes de structure syntaxique (Fudge, 1984 : 136) : « We have little alternative but to recognize a second type of compound whose nuclear stress falls on the compound final, and whose stress pattern is therefore identical with the phrasal pattern ».

Bauer (1983) arrive à la même conclusion et il plaide pour une prise en compte de considérations sémantiques. Pour lui, comme pour Fudge, le critère accentuel est trop restrictif : des unités lexicales ayant les deux types d'accentuation peuvent être des composés, et la syntaxe n'explique pas tout (Bauer, 1983 : 109) :

> All this suggests that the difference between single and double stressed collocations is not a distinction between two very different syntactic structures, but an accidental surface structure division in a unitary group of compounds, and that it makes more sense to talk of single and double stressed compounds than of compounds as opposed to noun + noun syntactic phrases.

Ces mêmes problèmes sont soulevés par Giegerich (1992) qui s'appuie sur les règles génératives (largement affinées depuis Chomsky & Halle) d'attribution

de l'accent. Il explique qu'il est parfois difficile de savoir laquelle règle doit s'appliquer. Il explicite les raisons de ces difficultés (Giegerich, 1992 : 258) :

> [...] in purely syntactic terms the distinction between compound and phrase is often hard to draw. But it does seem clear that the prominence pattern alone is not in itself an indicator of compound or phrase status (so that every compound with final stress is a phrase and every one with nonfinal stress a compound). If that were so then *ˌLondon ˈRoad* would be a syntactic phrase and *ˈLondon ˌStreet* a compound, *ˌChristmas ˈpudding* a phrase and *ˈChristmas ˌcake* a compound – hardly a categorisation that could be justified on syntactic grounds! We are left with the rather unsatisfactory situation, then, that the prominence patterns of compound words idiosyncratically conform with either those of words or those of phrases.

Il soulève ainsi les problèmes posés par le critère syntaxique et rejette le critère accentuel seul, lui aussi. En revanche, sa conclusion est loin d'être satisfaisante car il nous laisse avec l'impression qu'il ne peut y avoir de règles qui expliqueraient le comportement accentuel des composés. Dans la suite de cette discussion, nous nous intéressons plus particulièrement aux avis de chercheurs qui ont effectivement réussi à élaborer des règles et des principes qui tiennent compte de ces différences accentuelles.

4.2.2.2 L'approche sémantique : le rôle de l'accent

Si les chercheurs qui adoptent l'approche syntaxique attribuent un rôle de structuration du discours à l'accent, depuis les travaux de Liberman & Prince (1977), ce rôle fonctionne au niveau de la structure syntaxique exclusivement. Dans la mesure où les composantes phonologiques et sémantiques sont distinctes, il ne peut y avoir de lien entre elles. D'autres chercheurs, au contraire, attribuent à l'accent un rôle très important sur le plan sémantique afin d'expliquer le comportement accentuel des composés, ainsi que le système accentuel tout entier : l'accent est porteur de sens, il n'est pas simplement le résultat de règles syntaxiques.

L'idée sous-jacente de toute hypothèse sémantique est celle résumée par Lyons (Lyons, 1977 : 503) : « In any question that we might put relating to the components or circumstances of a situation, there is something that is presupposed and something that is in focus ». Lorsque nous appliquons ce principe à une étude de l'accent, cela implique non seulement les éléments accentués mais aussi ceux qui ne le sont pas. C'est-à-dire que la non-accentuation, ou l désaccentuation, joue un rôle aussi important que l'accentuation elle-même. Ce phénomène concerne non seulement l'accent lexical, qui porte sur une syllabe d'un mot, mais aussi, et surtout, l'accent nucléaire de la phrase. Il a donc été largement discuté dans le cadre de l'intonation (Kingdon, 1958b, Bolinger, 1972, Selkirk, 1984, Nicaise & Gray, 1998, Gray, 2001, Huart, 2002) mais au niveau de la composition, il est aussi significatif. Il y a désaccentuation lorsque des mots qui peuvent porter un accent n'en portent pas : c'est le cas des composés à accentuation simple. Nous allons d'abord étudier le rôle de l'accent dans ces cas de désaccentuation, puis son rôle lorsque l'un ou l'autre élément porte l'accent.

4.2.2.2.1 Désaccentuation d'un élément

4.2.2.2.1.1 Considérations accentuelles

Dans le cas de l'accentuation simple, l'élément final est désaccentué. D'après la règle générative de la *Compound Stress Rule*, cette désaccentuation est le résultat du placement de l'accent primaire, et non d'une opération significative en soi. La règle fonctionne d'une manière mécanique et l'élément se voit attribuer un accent de degré moindre en fonction de la réaction en chaîne qui est déclenchée par le placement de l'accent primaire. L'avis d'Huart (Huart, 1989 et 2002) est

tout autre : d'après elle, cette désaccentuation marque le fait qu'il y ait un élément qui manque et qui peut être reconstitué. La relation entre les éléments est pré-établie, et l'interlocuteur retrouvera ce qui manque soit dans le contexte immédiat soit dans le savoir partagé avec le locuteur. Il nous semble en effet qu'une hypothèse qui tient compte du contexte sémantique fournit des explications plus satisfaisantes des types d'accentuation que nous trouvons dans les composés.

4.2.2.2.1.2 Considérations phonétiques et phonologiques

Comme nous l'avons signalé dans les chapitres précédents, une des caractéristiques des composés est que l'élément non accentué ne subit pas de réduction vocalique, contrairement à ce qui se passe très souvent dans le lexique simple. Certains auteurs (Lutstorf, 1960 : 152, Marchand, 1969 : 24) considèrent que là où il y a une réduction vocalique, il s'agit d'une grammaticalisation de l'élément qui devrait plutôt être considéré comme un suffixe : c'est le cas des mots en *-man*, par exemple, ou *-man* serait devenu un suffixe désignant un agent (masculin). Nous verrons dans la suite de nos analyses la portée de ces remarques.

Fudge (Fudge, 1984) fait remarquer le caractère paradoxal de ce type de phénomène, car même si le second élément n'est pas proéminent, il est central du point de vue sémantique. « Note that although the compound final in these cases is less prominent phonetically, it is central from a semantic point of view ». De la mesure où la proéminence porte sur le premier et non le second élément, nous concluons qu'il n'y a pas de corrélation directe entre information et accent. En effet, Huart nous met en garde contre une utilisation trop simpliste de la théorie

informative selon laquelle la proéminence serait toujours associée à la « nouvelle » information.

Pour Duchet (Duchet, 1993 : 247), la non réduction de l'élément qui n'est pas proéminent est un signe de motivation. « Le maintien de l'identité sémantique trouve son parallèle dans celui de l'identité vocalique, marqueur de la motivation ». En effet, pour interpréter la relation entre les deux éléments, nous avons besoin de disposer de la phonie complète de l'élément désaccentué ; c'est ce que Duchet entend par « identité ».

Nous pouvons ainsi facilement concilier la non-accentuation du deuxième élément et sa non-réduction, si nous considérons que la non-réduction d'un élément correspond à la désaccentuation, plutôt qu'à la non-accentuation. Toutefois, ce phénomène soulève à nouveau la question de la notation des degrés d'accent (voir 1.4). Faudrait-il marquer l'élément désaccentué par /3/ plutôt que par /0/ ? Cette question sera étudiée dans l'analyse acoustique.

4.2.2.2.2 Accentuation simple ou accentuation double

Nous avons rappelé au début de ce chapitre que les composés ont deux types d'accentuation : l'accentuation simple lorsque l'accent est porté par le premier élément et l'accentuation double lorsqu'il porte sur les deux éléments. Nous allons étudier quelles sont les différentes explications proposées pour ces deux types d'accentuation, en commençant par l'accentuation simple.

4.2.2.2.2.1 L'accentuation « normale »

L'approche générative, qui consiste à ne considérer comme unité lexicale composée que les unités dont le premier élément porte l'accent, trouve une certaine justification dans le fait que les auteurs d'ouvrages pédagogiques comme les dictionnaires de prononciation et les manuels de phonétique et de phonologie s'accordent pour dire que l'accentuation « normale » d'un composé est l'accentuation simple (Gimson, 1980, Ginésy, 1995, Deschamps et al., 2000, Wells, 1990 et 2000). Cependant, ils ne se contentent pas de cette généralisation, et citent ensuite les classes d'unités qui ont une accentuation double ; mais ces classes ont aussi leurs exceptions avec une accentuation simple ! Ainsi, Wells (1990 et 2000) donne les composés où le premier élément désigne le matériel ou l'ingrédient du second élément comme formant une classe qui déroge à la règle ; puis il donne comme écart à cette règle le cas des composés dont le deuxième élément est l'un des suivants : *cake*, *juice*, *water*. De même, les noms de lieu ont une accentuation double sauf ceux en *street*. C'est dire la complexité de la question.

4.2.2.2.2.2 L'accentuation « exceptionnelle »

L'accentuation double correspond à l'accentuation utilisée dans les syntagmes, et donc à l'accentuation générée par la *Nuclear Stress Rule* des génératives. Dans les règles d'accentuation des composés, elle est présentée comme étant un type d'accentuation exceptionnelle, mais qui s'applique à des catégories importantes de composés, donc elle est utilisée assez fréquemment (voir 4.3). Dans les manuels et dictionnaires de prononciation, nous avons vu sous

la rubrique précédente qu'elle s'applique aux composés dont le premier élément se réfère au matériel ou à l'ingrédient qui compose le second. Elle s'applique aussi aux noms propres et aux noms de lieux.

Pour expliquer le type de schéma d'un composé, les règles les mettent souvent en opposition en se servant de critères sémantiques. Nous allons donc étudier les différentes manières d'expliquer ces deux types d'accentuation par rapport aux théories qui tiennent compte de ces facteurs.

4.2.2.2.2.3 L'accentuation : reflet d'une relation syntaxique entre les éléments

Certains chercheurs ont tenté d'expliquer le type d'accent par rapport à la relation entre les éléments sur le plan syntaxique. Pour Kingdon (1958b), l'accentuation simple correspond à la relation objet/sujet du premier élément par rapport au second (*fault-finding*). Quant à l'accentuation double, elle s'applique lorsque le premier élément fonctionne comme épithète du second

Les explications de Boisson (1980) sont structurées par rapport à la formation syntaxique. Cependant, cette méthode d'analyse est adoptée à des fins de clarification et pour étudier les tendances accentuelles en fonction des catégories. Nos méthodes seront semblables dans la mesure où nous nous intéresserons aux catégories syntaxiques des composés provenant des sources écrites et nous étudierons ces tendances à partir de notre base de données; toutefois, notre analyse ne se limitera pas aux simples catégories syntaxiques.

4.2.2.2.2.4 L'accentuation : reflet d'une relation sémantique entre les éléments

Les explications de Boisson (1980) peuvent être résumées dans cette section. Il cite la spécialisation du sens comme un des phénomènes qui entraînent l'accentuation simple. C'est-à-dire que, lorsqu'une unité comportant deux éléments est lexicalisée, le sens du composé ne peut pas s'interpréter à partir des éléments. Ainsi, pour les composés en Adj + N tels que *greenhouse* ou *blackbird*, l'accentuation est simple: *greenhouse* n'est pas une maison verte, et pour *blackbird*, les femelles ne sont pas noires, mais marron.

En plus de la spécialisation du sens, Boisson explique que nous trouverons une accentuation simple dans les composés où le second élément a un sens très général. L'explication repose sur une théorie informative, car dans ces cas, c'est le premier élément qui est porteur de sens précis et non le second. Un des exemples donnés est *dramatic society*. Si ce composé avait une accentuation double, ce ne serait plus un composé mais un adjectif qualificatif, *dramatic*, épithète du nom *society*, et cela voudrait dire que la société (ou l'association) en question est "dramatique" (ce qui ne veut pas dire grand-chose, d'ailleurs). Il semblerait que cette explication ne soit pas suffisante. Dans nos analyses, nous testerons l'observation que les composés formés à partir d'un deuxième élément général sont accentués sur le premier, et nous tenterons de formuler d'autres hypothèses en fonction de nos résultats.

Enfin, dans le cas d'un contraste implicite de type *backhand / forehand*, le premier élément sera accentué. Il ne faut pas confondre ce phénomène avec l'accent de contraste qui appartient, lui, au discours et où l'on veut faire ressortir

un mot par rapport aux autres (voir 1.3 «Les différents types d'accent»). Ici, Boisson parle de contraste implicite et il s'agit de deux composés (ou davantage) qui se distinguent grâce à leur premier élément. Là encore, nous testerons cette hypothèse, qui est également liée à la question de la sous-catégorisation. Ainsi, des types différents d'arbres se distinguent de par leur premier élément et l'accentuation est simple : *oak tree / cherry tree*, par exemple.

4.2.2.2.2.5 L'accentuation : reflet d'une relation de fonction / nature

Pour Ginésy (Ginésy, 1995), l'accentuation simple correspond aux cas où le composé définit une fonction, donc *cream jug* est accentué sur le premier élément car il s'agit d'un objet (second élément) ayant une fonction (au niveau du premier) Le sens de l'ensemble s'éloigne du sens des éléments, ce qui rejoint Boisson qui l'appelle la spécialisation du sens. En revanche, lorsqu'il s'agit de sa «nature» (Ginésy, 1995 : 130), comme dans *cream cheese*, le composé aura une accentuation double. Le sens du composé correspond au sens de la somme de ses éléments, lui conservant ainsi sa compositionnalité.

4.2.2.2.2.6 L'accentuation : reflet d'une relation préconstruite (ou non) entre les éléments

Les travaux d'Huart (Huart, 1979, 1984, 1989, 2002) s'inscrivent dans le cadre de la théorie de l'énonciation ; les composés y sont analysés comme faisant partie d'une «mini-grammaire», qui est le sous-titre de la partie concernant les composés dans son ouvrage le plus récent (Huart, 2002).

Ainsi, elle considère que l'accentuation simple est un marqueur de la relation entre les éléments qui indique qu'il s'agit ou d'une relation primitive ou

d'une relation qui peut être reconstruite par l'interlocuteur. En revanche, l'accentuation double correspond à l'identification ou à la localisation, c'est-à-dire à des opérations de mise en relation directe. La relation est donc simple, et non complexe comme dans les cas où le premier élément porte l'accent (ce que nous appelons accentuation simple). Dans le cas de l'accentuation double, les deux éléments fonctionnent comme une « chaîne » d'information. C'est ce que Ginésy appelle une relation de « nature ». Elle résume les principes fondamentaux de son approche (Huart, 1989 : 88)

> Les bases du système prosodique se résument ainsi : la chute intonative est placée sur le dernier élément porteur de "sens" lorsque les relations entre les différents termes en présence dépendent directement de ceux-ci (identification, localisation, prédication d'une propriété) ; la désaccentuation d'un ou de plusieurs éléments finaux fonctionne comme un signal indiquant que ces relations reposent sur un préconstruit et doivent donc être reconstruites. Il faut chercher ailleurs que dans l'énoncé les repères qui permettent de rétablir les liens : parfois dans le contexte immédiat ("contraste") mais tout aussi souvent dans une relation primitive, une notion prédicative présentée "en bloc", ou un enchaînement d'opérations non explicitées.
>
> Le non-accent est donc la marque d'un "non-dit", qui comprend le "déjà dit" mais il n'y a pas de correspondance terme à terme entre ce qui est désaccentué et ce que l'interlocuteur est censé pouvoir reconstruire. Dans un cas, il s'agit d'éléments de la chaîne, dans l'autre de relations.

Chez Huart, l'accentuation est expliquée comme étant un phénomène qui s'intègre à l'énoncé dans son ensemble et non comme un phénomène s'appliquant aux seuls mots isolés. C'est pourquoi elle emploie le terme « mini-grammaire ». La ressemblance entre le comportement accentuel des composés et le discours a déjà été relevée par Ladd (1984) qui rapproche les deux comportements en faisant remarquer que si le deuxième élément est général, il est donc attendu ; par conséquent, il n'est pas accentué. Nous trouvons ce même phénomène en discours. Mais les explications d'Huart du schéma accentuel des mots composés à accentuation simple sont reprises de manière bien plus précise par Nicaise & Gray

(1998: 61) dans leur ouvrage sur l'intonation. Ils soulignent eux aussi la ressemblance entre ces comportements, mais ils insistent également sur l'interprétation semblable de la relation :

> On peut considérer le placement de la proéminence nucléaire sur l'élément de gauche dans une phrase intransitive énoncée en début de discours comme le signal que la relation entre l'argument (A) et le prédicat (P) est préétablie – un peu comme la relation entre *coffee* et *shop* est préétablie dans le composé COFfee shop (accent à gauche). La relation est donc saisie en bloc et repérée directement par rapport à la situation d'énonciation.

Nous ne manquerons pas de mettre ces hypothèses à l'épreuve dans notre analyse acoustique.

Sous cette rubrique, les règles pour les composés ont été situées par rapport au contexte général du discours. Dans la section suivante, nous nous intéresserons à un phénomène du discours qui est tellement répandu que les dictionnaires en tiennent compte dans leurs indications sur l'accent lexical: il s'agit du déplacement de l'accent. C'est également un domaine qui a été largement étudié dans le cadre de la phonologie métrique.

4.2.3 Le déplacement de l'accent

Le déplacement de l'accent concerne les composés à accentuation double qui peuvent subir un inversement des valeurs accentuelles des éléments, voire une désaccentuation complète du deuxième élément[7]. Il est cité comme une occurrence très fréquente par les dictionnaires de prononciation, et par bien des chercheurs (Boisson (1980), Duchet (1991), Gimson (1980), Giegerich (1992), Montreuil,

---

[7] Voir aussi nos commentaires en 1.6, « L'accent dans le discours continu ».

(2001)). Il est également cité comme un test de la composition chez Roeper & Siegel, (1978). Le déplacement de l'accent est aussi une des causes principales de la variation (Lutstorf, 1960, Sampson, 1980, Shattuck-Hufnagel, 1994).

Les manières de l'expliquer et de le représenter sont diverses. Nous avons déjà montré l'intérêt de la grille métrique (voir 1.5.2, « La phonologie métrique », et 3.1.2.2, « Le pied ») qui a été introduite par Liberman & Prince (1977) et qui est utilisée dans la phonologie métrique depuis. Giegerich (1992 : 272) parle d'ajustement rythmique : « eurythmy: the rhythmic adjustment of stress ». Selkirk (Goldsmith, 1995) tout en utilisant la même représentation, plaide plutôt pour une prise en compte de « *Phrase Edge Prominence* » (voir 4.5.3.2, « *Phrase Edge Prominence* ou le déplacement de l'accent »).

Pour nos analyses, le déplacement de l'accent est significatif car il s'agit d'un phénomène de discours qui remet en cause dans une certaine mesure l'accent lexical. En effet, Kingdon (1958b) fait remarquer l'extrême difficulté de marquer l'accentuation (surtout des adjectifs composés) dans un dictionnaire ; il insiste sur les problèmes d'interprétation posés par ces notations. Bauer (1978 : 48) pose le problème très clairement :

> However, all of the works that have been discussed so far - and standard dictionaries as well – make the assumption that it is possible to state what the stress pattern on a given noun + noun configuration is. If this is not the case then the rules proposed in the literature become totally vacuous. It is, of course, widely recognized, that stress patterns on individual items frequently vary as a function of the prosodic environments in which they occur.

L'accent lexical peut donc servir d'indication relativement fiable, mais il faut avoir conscience de ses limites.

Enfin, nous citons Boisson (1980) qui, même s'il est très critique à l'égard du manque de systématicité des dictionnaires, prend une position qui justifie pleinement l'étude de l'accent lexical (Boisson, 1980 : 306) :

> Bien sûr ceci ne condamne pas du tout l'étude linguistique des mots à l'état isolé, parce que de leur structure il est possible de prévoir leur aspect en contexte grâce à des règles simples. L'inverse n'est pas possible : on ne peut pas remonter d'un mot en contexte à son allure isolé. Du reste, l'allure en contexte dépend ... du contexte, aussi d'autres facteurs comme la vitesse d'élocution.

Malgré les problèmes que posent la notation et l'étude des mots à l'état isolé, ce genre d'analyse permet de dégager des règles ; mais ces règles devront ensuite être testées en fonction du discours. C'est notre démarche dans les sections suivantes. Nous commençons notre analyse par l'accent lexical, puis nous étudierons les composés du corpus oral.

4.3 Analyse de l'accent lexical dans la base de données

Notre base de données contient quatre tables correspondant aux composés provenant de trois dictionnaires différents : EPD-Cp (de la 12ème édition du *English Pronouncing Dictionary*), NPD-Cp (du *New Pronouncing Dictionary* de Guierre), LPD-Cp (de la première édition de 1990 du *Longman Pronouncing Dictionary*, version électronique) et enfin N-COMPACT (également du *Longman Pronouncing Dictionary*, table saisie manuellement, et contenant les entrées absentes de la version électronique de ce dictionnaire).

L'analyse lexicale s'est faite à partir de deux types de requêtes : celles qui permettent de comparer les dictionnaires entre eux, car il y a une mise en relation entre les tables ; puis celles qui portent sur une seule et même table. Nous avons

déjà expliqué dans le chapitre 3[8] les différences de codages et de notations dont nous tenons compte dans la formulation des requêtes. Dans la première partie de cette analyse, nous vérifions le contenu de chaque table, surtout en ce qui concerne les « doublons ». Ensuite, nous nous intéresserons au contenu de chaque table, en étudiant les différents critères de sélection d'un composé afin de tirer des conclusions sur une éventuelle corrélation entre ces critères et le schéma accentuel. Enfin, nous étudierons l'évolution du lexique en comparant les entrées provenant de EPD12, paru en 1963, qui est notre source la plus ancienne, et celles de LPD90, paru presque trente ans plus tard, et qui contient près de quatre fois plus d'entrées.

4.3.1 Vérification des différentes tables de la base de données : les doublons

A l'aide des requêtes **DOUBLON-EPD** puis **EPD-DOUBLE** (voir Annexe A de ce chapitre), nous avons pu réunir les entrées ayant la même graphie dans la table EPD-Cp. Des requêtes semblables ont été créées afin d'obtenir ce resultat pour les autres tables. Nous constatons que les « doublons » se divisent en deux catégories : celles qui sont réellement des entrées dupliquées et celles qui sont différenciées par au moins une information. Le tableau suivant récapitule les résultats de cette étude :

---

[8] Voir 3.2.1.2.2, « Les champs de la base de données ».

Figure 4.1 : Tableau récapitulatif des « doublons »

TABLE	Nb entrées communes	Nb vrais doublons	Nb faux doublons
EPD-Cp	18	3	15
NPD-Cp	29	0	29
LPD-Cp	71	6	65
N-COMPACT	12	0	12

Pour les entrées qui sont de « vrais doublons », il s'agit d'une erreur ; nous en constatons le très petit nombre : seulement 9 entrées en double sur plus de 21 000 au total. Les « faux doublons » se différencient de plusieurs manières. La graphie en majuscules ou en minuscules explique 3 des cas de N-COMPACT (*new deal* et *New Deal*), alors que les 9 autres se différencient à la fois par leur schéma accentuel et par leur sens (par exemple, *toy factory* que nous avons déjà cité[9]). De même, dans EPD-Cp, nous avons un cas similaire (*counter-action*), et un qui s'explique par la différence de sens des éléments : *drawing-room*, qui a le même schéma, mais deux sens différents, selon s'il s'agit d'une « pièce où l'on dessine » ou bien du « salon ». Nous reviendrons à ces considérations sémantiques dans la dernière partie de ce chapitre (4.6.1).

Toutefois, la différence la plus fréquente porte sur la catégorie grammaticale, très souvent associée à une différence de schéma ou à une variante du schéma (sinon, le dictionnaire ne prévoit pas d'entrée séparée) : c'est le cas de 9 entrées de EPD-Cp, 18 entrées de NPD-Cp, et 59 entrées de LPD-Cp. Le nom correspond à un schéma « s » (accentuation simple) alors que les autres catégories syntaxiques correspondent à un schéma « d » (accentuation double). Ces résultats

[9] Voir 1.2, « Accent et compréhension » et 3.2.1.2, « Les tables et les champs ».

posent la question du comportement du reste du lexique en fonction de la catégorie syntaxique, ce qui sera étudié dans la section suivante.

Pour les entrées restantes, une différence de la catégorie syntaxique est associée à une différence de la phonie (6 cas de LPD-Cp), et nous constatons des différences au niveau de la phonie seule (3 cas dans NPD-Cp et 2 dans EPD-Cp), de la phonie et du schéma (3 cas dans NPD-Cp) et enfin des cas où seule la catégorie syntaxique les différencie (1 dans NPD-Cp, 2 dans EPD-Cp). En ce qui concerne la phonie, nous avons des entrées dont la prononciation de « s » est /z/ pour les verbes et /s/ pour les noms (*ill-use*), et des entrées où la voyelle finale du nom peut se réduire : c'est le cas des entrées en *-man* ou *-men* et de celles qui se terminent par « ate » qui se prononcent /eɪt/ pour les verbes et /ət/ pour les noms (*underestimate*), tout comme *homestead* qui a une variante en /ɪ/ pour le nom. Ceci n'entraîne pas de changement du schéma accentuel, car la réduction se fait sur la syllabe qui est déjà inaccentuée. En revanche, cette tendance révèle l'opacification du composé dans le cas de *homestead* et repose la question du statut des terminaisons en *-man*.

4.3.2 Etude de la corrélation entre le schéma accentuel et les autres critères de définition d'un composé

Dans le chapitre 2, nous avons passé en revue les différents critères de définition d'un composé : le critère syntaxique, le critère sémantique, la graphie, les caractéristiques phonétiques et phonologiques, enfin, le schéma accentuel. Ces critères ont été utilisés afin de sélectionner les composés de la base de données à partir des dictionnaires. Dans cette section, nous souhaitons étudier la corrélation

éventuelle entre le schéma accentuel indiqué par le dictionnaire (l'accent lexical) et ces autres critères.

Nous commencons cette étude par le schéma accentuel lui-même, puis nous l'étudierons par rapport aux trois critères suivants: la graphie, la phonie et la catégorie syntaxique. Cette dernière information donne des indications au niveau syntaxique mais sa portée est restreinte car l'accent lexical donne des prévisions hors contexte de discours. En ce qui concerne le critère sémantique, la base de données ne fournit que peu d'indications (seulement celles qui ont été citées dans la section précédente concernant quelques doublons de N-COMPACT qui se distinguent par leur sens).

4.3.2.1 Etude du schéma accentuel

Nous voulons d'abord étudier la structure de la base de données en ce qui concerne les indications de l'accent; puis nous serons en mesure de tirer des conclusions sur la corrélation entre ces différents critères et le schéma accentuel.

4.3.2.1.1 La structure des tables

Les presque vingt-deux mille composés de la base de données se divisent en deux catégories selon le type d'accentuation. Nous nous intéressons aux prononciations principales[10], et nous proposons le tableau suivant qui a été établi à

---

[10] En ce qui concerne les variantes, voir **VAR-SCH-LPD**, Annexe A. Pour extraire les variantes des entrées accentuées « s », nous avons pu rechercher la séquence {*%*} dans le champ « Variation ». Ce symbole indique la présence d'un accent secondaire (donc forcément une accentuation double, parfois sur la prononciation US). En revanche, il n'est pas possible de faire une requête simple pour extraire les variantes en « s » des entrées dont l'accentuation est de type « d » car LPD marque l'accent de plusieurs manières différentes (avec des tirets indiquant les syllabes, sur la syllabe concernée avec la phonie complète); l'entrée dans son ensemble n'est pas

partir de la requête **SYLL-SCHEMA**[11], et qui récapitule, table par table, le nombre et le pourcentage des entrées qui ont une accentuation simple « s » et celles ayant une accentuation double « d ».

Figure 4.2 : Tableau récapitulatif du type d'accentuation par table des entrées de la base de données (« s » accentuation simple / « d » accentuation double)

Table / d'accentuation	EPD-Cp	NPD-Cp	LPD-Cp	N-COMPACT[12]	base de données
Nb entrées total	4221	4126	7154	6261	21762
Accentuation simple « s » : nb d'entrées	3002	3907	5186	3042	15137
% « s » de la table	71,12%	94,69%	72,49%	48,59%	69,56%
Accentuation double « d » : nb d'entrées	1219	219	1968	3219	6625
% « d » de la table	28,88%	5,31%	27,51%	51,41%	30,44%

Les pourcentages sont très différents selon les tables. NPD-Cp a une très forte proportion de composés à accentuation simple, près de 95% ; EPD-Cp et LPD-Cp ont une proportion de plus de 70% ; en revanche, la table N-COMPACT en a moins de 50%. Si nous faisons la moyenne des deux dernières tables qui proviennent du même dictionnaire (LPD-Cp et N-COMPACT), 61,33% des entrées sont à accentuation simple. Ces chiffres nous incitent à approfondir cette étude, table par table, car de telles différences ne peuvent pas être dues au hasard.

Ainsi, nous avons voulu obtenir les mêmes informations, par table, mais aussi par nombre de syllabes. Les résultats font l'objet du tableau suivant[13] :

---

reprise. Une tentative d'extraire ces entrées a échoué : beaucoup de variantes ne concernent que la prononciation des segments. Il faut donc étudier chaque entrée individuellement.

[11] Voir Annexe A de ce chapitre.

[12] Ces chiffres sont légèrement faussés par la présence dans NCOMPACT d'entrées comportant des majuscules. Si on les exclut, la proportion de « s » et de « d » s'inverse : 53,45% de « s » et 46,55% de « d ». Nous en avons fait une étude particulière, voir partie 6 de ce chapitre.

Figure 4.3 : Accentuation simple ou double des composés par table et par nombre de syllabes[14] : « s » accentuation simple / « d » accentuation double

table % nb entrées / type d'accentuation	EPD-Cp	NPD-Cp	LPD-Cp	N-COMPACT
Nb entrées au total	4221	4126	7154	6261
Entrées de 2 syllabes en % du dictionnaire	**41,84%**	**81,77%**	**55,16%**	**20,06%**
% « s » des entrées de 2 syllabes	74,63%	95,94%	82,94%	58,52%
% « d » des entrées de 2 syllabes	25,37%	4,06%	17,06%	41,48%
Entrées de 3 syllabes en % du dictionnaire	**41,67%**	**15,83%**	**34,37%**	**35,30%**
% « s » des entrées de 3 syllabes	71,52%	91,42%	66,82%	55,48%
% « d » des entrées de 3 syllabes	28,48%	8,58%	33,18%	44,52%
Entrées de 4 syllabes en % du dictionnaire	**13,76%**	**2,23%**	**8,28%**	**25,70%**
% « s » des entrées de 4 syllabes	64,89%	75,00%	42,06%	45,31%
% « d » des entrées de 4 syllabes	35,11%	25,00%	57,94%	54,69%
Entrées de 5 syllabes en % du dictionnaire	**2,20%**	**0,17%**	**1,68%**	**10,81%**
% « s » des entrées de 5 syllabes	45,16%	57,14%	17,50%	33,09%
% « d » des entrées de 5 syllabes	54,84%	42,86%	82,50%	66,91%
Entrées de 6 syllabes en % du dictionnaire	**0,52%**	**0,00%**	**0,46%**	**5,02%**
% « s » des entrées de 6 syllabes	31,82%		0,00%	27,71%
% « d » des entrées de 6 syllabes	68,18%		100,00%	72,29%
Entrées de 7 syllabes en % du dictionnaire			**0,04%**	**1,90%**
% « s » des entrées de 7 syllabes			0,00%	24,37%
% « d » des entrées de 7 syllabes			100,00%	75,63%

Nous constatons que la structure de ces tables, en fonction du nombre de syllabes, est très différente : NPD-Cp comporte 80% d'entrées de deux syllabes alors que N-COMPACT n'en comporte que 20%. De manière générale, les entrées de N-COMPACT ont tendance à être sensiblement plus longues que celles des autres tables : 44% des entrées sont de quatre syllabes ou plus, alors que des entrées de cette longueur ne représentent que 16% des entrées de EPD-Cp, 10% des entrées de LPD-Cp, et un peu plus de 2% de NPD-Cp. Le fait que N-

---

[13] Ce tableau ne reprend que les pourcentages. En annexe B se trouve le détail, syllabe par syllabe et table par table.

[14] La table NCOMPACT contient des entrées de plus de sept syllabes, mais elles représentent moins de 1% du total (61 entrées en tout) ; c'est la raison pour laquelle nous les avons exclues du tableau car elles ne sont pas représentatives.

COMPACT comporte plus d'entrées de plusieurs syllabes explique vraisemblablement la plus grande proportion d'entrées à accentuation double dans cette table.

Afin de s'assurer du bien-fondé de cette hypothèse, il suffit de regarder les pourcentages de « s » par rapport aux « d » dans le tableau ci-dessus. Pour toutes les tables, plus le nombre de syllabes augmente, moins la proportion de « s » est grande. Il s'agit bien de pourcentages, donc de proportions et non de chiffres absolus. Ainsi, pour les entrées de deux syllabes, les « s » représentent de 60% et 95% des entrées, selon les tables, alors que pour les entrées de quatre syllabes, cette proportion se situe entre 42% et 65% seulement.

4.3.2.1.2 Conclusions

L'étude de la structure des tables par rapport aux schémas accentuels a permis de vérifier l'hypothèse selon laquelle l'accentuation simple est celle que nous devons associer aux composés; l'accentuation double étant considérée comme exceptionnelle (voir ce chapitre, 4.2.2.2, « Accentuation simple ou accentuation double »). Toutefois, la proportion de composés ayant ce schéma « exceptionnel » est non négligeable, sauf dans NPD-Cp : environ 30% de LPD-Cp et EPD-Cp, et 50% pour N-COMPACT. L'importance de ces chiffres nous oblige à poursuivre notre analyse.

Cette étude a également permis de dégager une première corrélation entre le nombre de syllabes et le type d'accentuation : les composés ayant le moins de syllabes ont tendance à avoir une accentuation simple ; plus le composé est long, plus il aura tendance à porter une accentuation double. Dans la partie 5 de c.

chapitre, l'analyse acoustique permettra de donner d'autres éclairages sur cette hypothèse.

Toutefois, si ces tendances sont les mêmes pour toutes les tables, il subsiste des différences de grandeur significatives entre elles. Nous pourrons peut-être les expliquer grâce à d'autres critères ; nous passons à l'étude de la graphie.

4.3.2.2 Etude de la graphie

Tout d'abord, nous rappelons les trois graphies possibles pour un composé : graphie sans séparateur (appelée « compacte » dans nos travaux), graphie avec un trait d'union, et graphie avec un espace (appelée « non compacte » dans nos travaux). Les règles d'accentuation rappellent souvent que la graphie est une indication peu fiable du type d'accentuation[15]. Nous allons étudier cette question.

4.3.2.2.1 Caractéristiques des tables

Les tables n'ont pas du tout les mêmes caractéristiques graphiques ; nous rappelons les spécificités de chacune. EPD ne comporte que des entrées à trait d'union car c'était le critère selon lequel Guierre les avait écartées de son fichier principal (et les a restituées dans le fichier séparé que nous avons utilisé), comme nous l'avons expliqué dans le chapitre 2 (2.2.1.3.1). En revanche, NPD ne comporte aucun séparateur : ni trait d'union, ni espace (l'apostrophe n'est pas un séparateur au même titre que les autres). N-COMPACT contient toutes les entrées du dictionnaire LPD qui ne figuraient pas dans la version électronique dont nous

---

[15] Voir 2.1.1.1.3, « Le critère graphique ».

disposions, c'est-à-dire les sous-entrées du dictionnaire qui s'écrivent toutes avec un espace, parfois aussi avec un trait d'union ou sans séparateur en plus de l'espace dans le cas de surcomposés (*grandfather clock*). Enfin, LPD comporte des entrées à trait d'union et des entrées sans séparateur, mais très peu d'entrées avec espace.

4.3.2.2.2 La structure des tables

Le tableau suivant permet de préciser la structure graphique des tables.

Figure 4.4 : Tableau récapitulatif de la structure graphique des tables

Table & Graphie	EPD-Cp	NPD-Cp	LPD-Cp[16]	N-COMPACT[17]	Base de données
Nb entrées	4221	4126	7154	6261	21762
Graphie compacte	0	4126	5045	0	9171
% Compacte	**0,00%**	**100,00%**	**70,52%**	**0,00%**	**42,14%**
Graphie à trait d'union	4221	0	2108	0	6329
% Trait d'union	**100,00%**	**0,00%**	**29,47%**	**0,00%**	**29,08%**
Graphie non compacte	0	0	43	6261	6304
% Non compacte	**0,00%**	**0,00%**	**0,60%**	**100,00%**	**28,97%**
Graphie mixte non compacte /trait d'union	0	0	17	215	232
Graphie mixte non compacte /compacte	0	0	0	99	99
% Mixte	**0,00%**	**0,00%**	**0,24%**	**1,02%**	**1,52%**

Pour les tables EPD-Cp, NPD-Cp et N-COMPACT, les pourcentages obtenus sont de 100%, ce qui s'explique par le fait que chacune de ces tables correspond à un type de graphie, donc toutes les entrées s'écrivent de la même

---

[16] Les résultats pour LPD-Cp ont été obtenus à partir de la requête **RECH-GRAPHIE**, voir Annexe A de ce chapitre. Nous avons recherché les caractères du trait d'union et de l'espace, en tenant compte du codage. La virgule indique la présence d'une variante graphique, l'espace finale est présent dans toutes les entrées donc nous avons dû rechercher des séquences de type {* * }, graphique. Les informations relatives aux variantes graphiques se trouvent en Annexe B.

[17] Si nous écartons les entrées avec majuscules, ce chiffre est ramené à 5474. Voir l'étude des majuscules en 4.6.2, « Etude des entrées ayant une majuscule ».

manière. En revanche, pour LPD-Cp, les trois graphies existent, et certaines entrées comportent deux graphies (ou davantage). Les pourcentages sont calculés à partir du nombre total d'entrées; c'est pourquoi, si on additionne les pourcentages, ceux qui correspondent à une graphie mixte (avec au moins deux graphies données pour l'entrée) dépassent 100%. Nous constatons que LPD-Cp contient environ 30% d'entrées à trait d'union et 70% d'entrées sans séparateur; le nombre d'entrées ayant une graphie avec espace est non significatif (0,60%).

Il est évident que certaines des entrées de ces tables sont communes avec d'autres: le composé *greenhouse* se trouve-t-il dans une seule table ou dans plusieurs? Afin d'être en mesure de tirer des conclusions sur la graphie des entrées et la corrélation éventuelle avec le schéma accentuel, il faut pouvoir comparer les entrées des différentes tables. Pour ce faire, nous devons tenir compte de deux facteurs, le premier étant le plus important. Les entrées de dictionnaires différents peuvent s'écrire de manière différente. Par exemple, toutes les entrées de EPD-Cp s'écrivent avec un trait d'union alors que toutes celles de NPD-Cp s'écrivent sans séparateur. Il est extrêmement improbable qu'il n'y ait aucune entrée commune, donc il faut trouver le moyen de les comparer, sans tenir compte du type de séparateur, tout en conservant le moyen de retrouver la graphie d'origine.

Le second facteur concerne le codage des entrées graphiques. Pour les autres tables, il y a une entrée graphique unique, mais pour LPD-Cp, il existe trois symboles qui peuvent fausser les résultats: la présence de la virgule là où il y a une variante graphique (207 entrées) ; le signe « / » (60 entrées), et la barre

verticale qui indique une coupe morphologique (214 entrées). De plus, toutes les entrées graphiques de LPD-Cp se terminent par un espace (ce qui permet d'identifier la fin de la graphie si on fait une recherche sur un élément).

Nous avons donc créé des tables parallèles où les séparateurs ont été enlevés de la graphie ainsi que l'espace final de LPD-Cp ; ces tables s'appellent : EPD-grx, NPD-grx, LPD-grx et NCOMP-grx. La terminaison « grx » représente la graphie « x », c'est-à-dire sans séparateurs, comme nous l'avons expliqué dans le chapitre 3 (3.2.1.2.2.2.2, « Les champs de la base de données : La graphie »). Pour retrouver la graphie d'origine, nous nous sommes servie de deux indications : la source, car EPD, NPD et N-COMPACT correspondent chacune à un type de graphie ; la graphie inverse pour LPD et aussi NPD. Il faut se rappeler que la graphie inverse correspond à la graphie principale seulement[18].

Les tableaux suivants récapitulent les résultats de la comparaison des tables entre elles[19] ; nous avons d'abord comparé les tables d'origine, puis celles avec la graphie neutralisée, en recherchant la graphie d'origine d'après les critères cités ci-dessus.

---

[18] Cet exercice a également révélé quelques erreurs, sans doute lors du transfert des données (de Mac vers WINDOWS ou d'EXCEL vers ACCESS) ; par exemple, les apostrophes de la graphie principale ne figurent pas toujours dans la graphie inverse ; et là où il y a un espace dans la graphie principale, la graphie inverse s'arrête (graphie : *Eng. Lit.* → graphie inverse : *til.*). Ces erreurs ne représentent que quelques entrées.

[19] La requête **COMP-*LPD-EPD***, voir Annexe A.

Figure 4.5: Comparaison des entrées graphiques des tables, graphie d'origine

graphie d'origine	EPD-Cp	NPD-Cp	LPD-Cp	N-COMPACT	base de données
Nb entrées	4221	4126	7154	6261	21762
Entrées communes EPD/LPD	585		585		
Entrées communes LPD/N-COMPACT			0	0	
Entrées communes NPD/LPD		3035	3035		
Entrées communes EPD/NPD	0	0			
Entrées communes EPD/N-COMPACT	0			0	
Entrées communes NPD/N-COMPACT		0		0	

Il est normal qu'il n'y ait aucune correspondance entre les trois tables qui correspondent à des graphies différentes (EPD-Cp, tiret; NPD-Cp, sans séparateur; et N-COMPACT, espace). De même, nous nous attendons à ce qu'il n'y ait pas de correspondance entre LPD-Cp :. N-COMPACT car il s'agit de la même source.

Si nous comparons les chiffres relatifs aux deux tables de taille comparable, EPD-Cp et NPD-Cp (qui contiennent chacune environ 4000 entrées) avec les chiffres de la table LPD-Cp qui comporte les deux graphies, le premier constat qui s'impose est le très petit nombre d'entrées communes entre EPD-Cp et LPD-Cp alors qu'il y a une forte proportion d'entrées communes entre NPD-Cp et LPD-Cp (les trois-quarts de la table NPD-Cp). Sans doute ceci s'explique-t-il par la différence de graphie puisque LPD-Cp a deux fois plus d'entrées sans séparateur (ce qui correspond à la graphie de NPD-Cp) que d'entrées avec un trait d'union (qui est la graphie de EPD-Cp). Cependant, l'importance de l'écart est étonnante. Nous pourrons vérifier ces résultats à partir du tableau suivant qui récapitule les résultats de la comparaison de ces tables avec la graphie neutralisée.

Figure 4.6: Comparaison des entrées graphiques des tables, graphie neutralisée

graphie neutralisée	EPD-grx	NPD-grx	LPD-grx	NCOMP-grx
entrées communes NPD/LPD		3078	3078	
entrées communes EPD/LPD	1683		1683	
entrées communes EPD/N-COMPACT	693			693
entrées communes EPD/NPD	602	602		
entrées communes EPD/NPD/LPD	589	589	589	
entrées communes LPD/N-COMPACT			76	76
entrées communes NPD/N-COMPACT		32		32
entrées communes EPD/NPD/LPD/N-COMPACT	5	5	5	5

Nous constatons que les chiffres des deux tableaux diffèrent peu en ce qui concerne la comparaison entre NPD et LPD. En ajoutant les entrées communes de NPD-Cp /N-COMPACT (car LPD-Cp et N-COMPACT sont de la même source) nous retrouvons environ 80 entrées communes de plus, qui correspondent à des variantes graphiques entre les dictionnaires. En revanche, les chiffres relatifs à la comparaison entre EPD-Cp et LPD-Cp/N-COMPACT sont largement supérieurs: ils passent de moins de 600 entrées communes à 2376 (en additionnant les entrées communes de LPD-Cp et N-COMPACT). Cependant, si cette augmentation s'explique par la variation graphique, ces résultats montrent que seulement 56% des entrées provenant de EPD se trouvent dans les tables de LPD. Nous reviendrons à ces considérations dans la section suivante, où nous étudierons l'évolution du lexique entre EPD12 et LPD90.

Si nous comparons les chiffres de EPD-Cp et de NPD-Cp, de taille semblable, nous constatons le petit nombre d'entrées communes (seulement 600 sur 4000). Ce résultat est très important car il nous oblige à poser la question du choix des entrées composées et des critères de sélection.

Il est aussi intéressant de commenter les 76 entrées communes de LPD-Cp/N-COMPACT car il s'agit de la même source, ce qui implique que ces entrées sont à la fois dans l'entrée principale du dictionnaire et dans une sous-entrée. Il est vrai que le nombre est très petit par rapport au nombre d'entrées au total (76 sur 13 415), mais l'existence de ces entrées doubles pose la question de l'indexation et de la présentation des composés dans un dictionnaire (voir 3.2.2). Le souci d'économie voudrait que chaque entrée ne figure qu'une seule fois, mais il est évident que la recherche d'un composé peut se faire par éléments. Nous reviendrons à ces considérations dans notre conclusion finale.

4.3.2.2.3 Conclusions

Ces comparaisons nous permettent d'étudier non seulement la graphie en tant que choix de représentation, mais aussi par rapport aux enseignements qu'elle nous fournit sur le contenu même des tables. Sous cette rubrique, nous nous intéresserons d'abord à ces enseignements, ce qui nous permettra de tirer des conclusions sur la graphie en tant que critère de sélection. Puis, pour terminer, nous récapitulerons les informations relatives à la variation graphique.

4.3.2.2.3.1 Le contenu des tables : critères de sélection des composés

La comparaison des tables fait apparaître des différences au niveau informationnel : si LPD-Cp et N-COMPACT contiennent beaucoup plus d'entrées que les autres tables, c'est que la sélection des entrées est moins restrictive. Cependant, il y a environ 1000 entrées de NPD-Cp et 1800 de EPD-Cp qui n'y figurent pas. Nous reviendrons sur la comparaison entre EPD12 et LPD90 dans la

section suivante lorsque nous étudierons l'évolution du lexique, mais en ce qui concerne NPD, nous devons faire quelques remarques.

4.3.2.2.3.1.1 Entrées manquantes de LPD-Cp/N-COMPACT par rapport à NPD-Cp

NPD a été conçu par Guierre qui a marqué «Cp» comme catégorie syntaxique. Nous ne savons pas comment ce marquage a été fait; il est peu probable qu'il ait utilisé un critère graphique quelconque. Toutefois, il subsiste dans cette table des entrées qui ne sont pas des composés, telles que *badminton* et *vodka*, mais dont la structure phonologique rappelle celle d'un composé: la suite /dm/ ou /dk/ étant «illicite», il est possible qu'une programmation visant à marquer les composés d'après ce type de critère ait pu générer ce genre d'erreur.

Il subsiste également un certain nombre de composés étrangers que nous avons volontairement écartés de la base de données (*bratwurst*, *autobahn*), ainsi que des composés historiques devenus opaques (*bonfire*, *boatswain*). NPD inclut aussi des formes dérivées telles que *blackmailer*, où LPD ne donne que *blackmail*. Certaines entrées de deux syllabes accentuées /10/ et dont la seconde voyelle est réduite ne figurent pas non plus dans la table LPD. Nous y reviendrons sous la rubrique suivante lorsque nous étudierons le critère de la phonie. Les 203 entrées en *-man* ou *-men* ne sont pas toutes reprises.

Cependant, les manques ne peuvent pas tous s'expliquer et nous avons conscience qu'il y a réellement un problème de choix des entrées.

4.3.2.2.3.1.2 Entrées manquantes de EPD-Cp par rapport à NPD-Cp

Nous avons déjà souligné le très petit nombre d'entrées communes entre ces deux tables. S'il n'y a que 600 entrées communes sur plus de 4000 pour chacune de ces tables, le critère de sélection peut être remis en cause. Le critère de sélection de EPD-Cp est le trait d'union, ce qui explique l'absence dans cette table de composés qui s'écrivent sans séparateur. Or, cette graphie représente les deux-tiers de la table LPD-Cp, soit plus de 5000 entrées et c'est la seule graphie représentée dans NPD. Le composé *greenhouse* que nous avons cité ci-dessus se trouve dans NPD-Cp et dans LPD-Cp mais il est absent de EPD-Cp[20] : pourtant, il s'agit bien d'un composé d'après les critères sémantiques que nous avons cités au chapitre 2. Il est clair que nous devons tenir compte d'autres critères que le critère graphique pour sélectionner les composés.

4.3.2.2.3.2 La graphie des tables : la variation

Pour évaluer l'étendue de la variation, nous devons comparer les résultats des deux tableaux comparatifs, puis soustraire les chiffres du premier tableau (qui représentent des graphies communes) du deuxième (qui rassemble les entrées ayant la même graphie et celles dont la graphie varie). Ainsi, près de 1800 des 2300[21] entrées communes de EPD et LPD/N-COMPACT ont des variantes

---

[20] Il faut rappeler que le fichier des composés de EPD a été conçu pour rassembler les mots composés car leur comportement accentuel, selon Guierre, ne correspond pas à celui du lexique dans son ensemble. Peut-être que les nombreux composés subsistants (souvent de deux syllabes) ne présentent pas le genre d'anomalie typique de certains composés (notamment la double accentuation ou les accents adjacents).

[21] Ce chiffre tient compte du fait qu'il y a 76 entrées communes dans LPD-Cp/NCOMPACT qu'il ne faut pas compter deux fois.

graphiques. En revanche, nous constatons qu'il y a peu de variantes entre NPD et LPD/N-COMPACT. Est-ce que nous pouvons expliquer ce phénomène par une évolution de la graphie ? Dans ce cas, l'évolution se fait dans les deux sens car les deux tables provenant du *Longman Pronouncing Dictionary*, LPD et N-COMPACT, ont des graphies différentes : N-COMPACT correspond à un éclatement de la représentation graphique puisqu'il y a un espace, alors que LPD contient des entrées sans séparateur. Nous étudierons cette éventualité dans la section sur l'évolution du lexique.

Pour conclure cette rubrique sur le critère de la graphie, nous allons étudier la corrélation entre le type de graphie et le schéma accentuel.

4.3.2.2.4 La corrélation entre le schéma accentuel et la graphie

Comme nous l'avons rappelé au début de cette rubrique, la graphie est considérée en général comme une indication peu fiable du schéma accentuel. Nous avons voulu étudier cette question ; le tableau suivant nous permettra de tirer certaines conclusions.

Figure 4.7 : Structure des entrées graphiques / schémas accentuels, table par table

Table *type de graphie* Nb et % s/d	EPD-Cp *Trait d'union*	NPD-Cp *Compact*	LPD-Cp *Toutes graphies*[22]	LPD-Cp[23] *Compact*	LPD-Cp *Trait d'union*	LPD-Cp *Non compact*	N-COMPACT *Non compact*
Nb entrées	4221	4126	7154	*5045*	*2108*	*44*	6261
Nb avec schéma «s»	3002	3907	5186	*4390*	*796*	*18*	3042
% «s» du dictionnaire (% des entrées avec cette graphie)	71,12%	94,69%	72,49%	*61,36% (87%)*	*11,13% (38%)*	*0,25% (41%)*	48,59%
Nb avec schéma «d»	1219	219	1968	*656*	*1311*	*26*	3219
% «d» du dictionnaire (% des entrées avec cette graphie)	28,88%	5,31%	27,51%	*9,17% (13%)*	*18,33% (62%)*	*0,36% (59%)*	51,41%
% Graphie de LPD-Cp				*70,52%*	*29,47%*	*0,70%*	

La graphie sans séparateur est généralement associée à une accentuation simple ce qui semble être tout à fait corrélé par ces résultats : près de 95% des entrées de NPD-Cp (donc graphie en compact) et 87% des entrées avec cette graphie dans LPD-Cp portent ce schéma. En revanche, la graphie en non compact, avec espace, indiquerait une tendance à l'accentuation double. Nos résultats montrent effectivement cette tendance, car entre 50% et 60% des entrées de N-COMPACT et les entrées de LPD-Cp ayant cette graphie ont une accentuation double ; mais contrairement aux entrées sans séparateur, il n'y a qu'une tendance.

---

[22] Ce chiffre ne tient compte que des entrées principales alors que les chiffres correspondant aux différentes graphies comprennent aussi les variantes. Ainsi, si on les additionne, on dépasse le nombre d'entrées.

[23] Les colonnes avec les caractères en italiques correspondent aux variantes graphiques de LPD-Cp.

Environ la moitié des entrées sont concernées, l'autre moitié s'accentue toujours sur le premier élément.

Nous sommes en mesure de formuler une première hypothèse selon laquelle la graphie en compact est un bon indicateur du type d'accentuation ; nous sommes plus réservée en ce qui concerne les tendances accentuelles de la graphie non compacte. Par ailleurs, ces résultats démontrent une nouvelle fois que l'accentuation « normale », quelle que soit la graphie, est l'accentuation simple.

Néanmoins, ce sont les résultats concernant les entrées avec un trait d'union qui sont les plus surprenants. Si EPD a une proportion de 70% d'entrées ayant une accentuation simple, cette proportion n'est pas la même pour les entrées à trait d'union de LPD-Cp où le chiffre atteint à peine 40%. De telles inversions exigent un commentaire. Nous rappelons que 1100 entrées de LPD-Cp communes à EPD-Cp s'écrivent sans séparateur, ce qui correspond à peu près au nombre d'entrées communes ayant une accentuation simple[24]. Cela implique une préférence de LPD-Cp pour la graphie sans séparateur : comme le schéma accentuel ne change pas, cette préférence aboutit à des proportions très différentes en ce qui concerne la relation entre la graphie des entrées à trait d'union et les schémas. Si ces entrées avaient la même graphie à trait d'union, la proportion serait d'environ 60% au lieu de 40%. De plus, il faut noter que près de 70% des entrées communes entre EPD-

[24] Voir Annexe B pour le détail de ces résultats, ainsi qu'un récapitulatif du schéma accentuel de LPD-Cp, graphie par graphie et syllabe par syllabe.

Cp et N-COMPACT ont une accentuation simple, ce qui représente une forte proportion par rapport aux résultats de l'ensemble de cette table.

Ces faits appellent deux commentaires. Premièrement, le choix même de la graphie peut être un choix délibéré. Nous rappelons les propos de Fowler (Fowler, 1952) qui fait du schéma accentuel un critère orthographique Le choix d'une graphie en compact alors que l'entrée existe sous d'autres formes, correspond peut-être à une volonté de refléter le schéma dans la graphie ; ce choix reflète sans doute également le processus de lexicalisation car, sans séparateur, l'identité des éléments est moins distincte. Cependant, la variation graphique existe aussi dans le sens inverse ; il serait imprudent de tirer des conclusions trop hâtives de ces variantes.

Le second commentaire reprend la question du choix des entrées. Si près de 700 entrées sont communes à EPD-Cp et à N-COMPACT, près de 500 ont une accentuation simple, ce qui démontre que les entrées de EPD ne sont pas du tout représentatives de la table N-COMPACT dont la proportion d'entrées à accentuation simple est bien plus basse. De plus, la table N-COMPACT comprend beaucoup d'entrées qui ne figurent ni dans EPD-Cp ni dans NPD-Cp. Nous devons nous poser la question de savoir pour quelles raisons ces entrées en sont écartées ; lesquelles sont incontournables ; selon quels critères de sélection.

Certaines des considérations que nous avons relevées dans l'étude de la graphie seront reprises dans la section concernant l'évolution du lexique, mais nous allons d'abord étudier d'autres critères qui ont pu servir pour définir les

composés. Sous la rubrique suivante, nous étudierons la phonie et la corrélation entre le schéma accentuel et celle-ci.

4.3.2.3 Etude de la phonie

Nous savons qu'une des caractéristiques des composés à accentuation simple est la non-réduction du deuxième élément (monosyllabique) désaccentué (voir 4.2.2.2.1.2, « Considérations phonétiques et phonologiques : Désaccentuation d'un élément »). Nous en avons fait un critère d'extraction pour obtenir les entrées supposées être des composés à partir de la version électronique de LPD[25] : nous avons écarté toutes les entrées en /10/ dont la deuxième voyelle est réduite à schwa.

Afin de tester le bien fondé de cette hypothèse, et sa portée sur les statistiques concernant le type d'accentuation, nous avons recherché les entrées de deux syllabes des autres dictionnaires dont la voyelle du deuxième élément est schwa. Nous avons pu évaluer ainsi le nombre et le type d'entrées que nous n'avons pas extraites de LPD en faisant cette hypothèse[26].

Les requêtes **SCHWA-NPD** et **SCHWA-EPD** (voir annexe A) nous ont permis d'interroger la base de données. Pour NPD-Cp, 286 entrées sont concernées, c'est-à-dire près de 7% de la table, ce qui paraît être une proportion très importante. Cependant, ces entrées se concentrent surtout sur les mots en *-man* (110) et *-men* (109), soit 219 entrées. Il est certain que toutes ces entrées ne

---

[25] Voir 2.2.1.2.4.1, « La non-réduction de la voyelle du deuxième composant ».

[26] Si ces entrées figuraient dans la table, cela augmenterait le taux d'entrées à accentuation simple.

figurent pas dans LPD qui n'en a que 30 en *-man* et 2 en *-men* (N-COMPACT en a 9 en *-man*). Cela explique en partie les entrées manquantes de LPD par rapport à NPD que nous avons commentées sous la rubrique précédente. Cela pose aussi la question de l'inclusion ou non des mots au pluriel dans le dictionnaire. C'est en effet ce qui se passe souvent dans NPD et EPD, qui incluent des entrées distinctes, mais LPD indique le pluriel d'une autre façon dans la même entrée.

Il faut rappeler que la phonie de *-man* ne se trouve pas toujours avec une voyelle réduite, comme on le constate en étudiant la table EPD où la prononciation principale est toujours /mæn/. La voyelle réduite est alors une variante qui figure dans la version papier. Dans la mesure où il ne s'agit que de variantes, il est difficile de se prononcer sur le statut de *-man* qui est peut-être celui d'un suffixe plutôt que d'un élément de composition. La question a déjà été soulevée (voir 4.2.2.2.1.2 «Considérations phonétiques et phonologiques»). En tant que suffixe, il désignerait un agent (masculin). Nous pensons que le cas de *-man* est un peu plus compliqué. La réduction de la voyelle fait disparaître l'identité de l'élément, ce qui sous-entend son statut de suffixe, mais dans le cas de *-man* l'identité se perd au point d'abandonner aussi dans une certaine mesure son genre. Si le mot *chairman* a été remplacé aujourd'hui par *chairperson*, c'est parce qu'il est politiquement correct de ne pas utiliser d'élément pouvant être interprété comme discriminatoire; jusqu'alors, nous utilisions le mot *chairman* indifféremment pour les hommes et les femmes. Il est peu probable que tous les mots en *-man* seront remplacés par *-person*.

Grâce à cet exemple anecdotique, nous voudrions souligner deux points qui paraissent contradictoires. Si des mots en *-man* sont utilisés indifféremment pour des hommes et des femmes, c'est qu'une partie de l'identité s'est perdue, sans doute à cause de la réduction vocalique, et il ne subsiste que l'indication d'agent. Si nous comparons *Superman*, qui se prononce obligatoirement avec une voyelle pleine, nous savons qu'il s'agit d'un homme ; en revanche, *postman* peut très bien désigner une femme. Pour distinguer la femme de l'homme, nous pourrons faire comme pour les médecins : *woman doctor*, *woman postman*. En revanche, si nous changeons les mots du lexique pour les rendre plus acceptables dans notre société, ce qui est une forme d'hypercorrection, c'est que l'élément *-man* est bien considéré comme un élément de composition, quelle qu'en soit la prononciation. Ceci est certainement dû à l'influence de l'écrit qui permet de conserver encore cette identité.

Si nous écartons les entrées en *-man* et *-men*, il reste à peine 70 cas de NPD ayant une voyelle réduite sur la deuxième syllabe inaccentuée. Une quarantaine sont des erreurs ; nous en avons déjà relevé : *vodka*, et d'autres comme *boycott*, *lobster*, *neighbour*, *happen*, *röntgen*, *hoodlum*, *jetsam*, *kincob*. Il y a également quelques composés historiques que nous n'avions pas retenus[27] tels que *stedfast*, *breakfast*, justement parce que l'identité des éléments n'est plus reconnaissable. Dans le cas de *breakfast*, cette perte d'identité est due à la réduction de *fast*, et aussi peut-être à des raisons culturelles : le jeûne est devenu une coutume moins

[27] Voir 2.1.2.3.3 « Les composés historiques ».

courante et moins centrale dans la périodisation de la journée et des repas. Pour *stedfast*, le premier élément n'est pas connu, ce qui fait que, même si le second ne se réduit pas, il n'est pas perçu comme un composé.

Il subsiste 27 entrées que nous devons encore commenter; le second élément en est: *pence* (10), *land* (8), *board* (3), *yard* (3), *pan* (un cas: *saucepan*), *worth* (un cas: *pennyworth*), *stone* (un cas: *gladstone*). Le cas de *gladstone* est particulier: bien qu'il n'y ait pas de majuscules dans NPD, il s'agit bien d'un nom propre. Nous l'aurions donc écarté de notre base de données; de plus, puisque les éléments ne sont pas ressentis comme étant compositionnels, nous ne le considérons pas comme un composé. Il y a 31 entrées de deux syllabes en *stone* dans LPD-Cp (6 dans N-COMPACT) toutes avec une voyelle pleine.

En ce qui concerne *pennyworth*, la prononciation principale de LPD est /ˈpenəθ/, c'est-à-dire que la réduction de l'élément désaccentué s'accompagne d'une compression du premier élément. Il n'y a plus aucun lien entre les éléments et leur prononciation en tant que mots isolés; pour cette raison, cette entrée ne pourra pas être considérée comme un composé.

Le cas de *saucepan* est un peu différent. Lorsque nous parlons de composition, nous citons souvent la spécialisation du sens[28] comme étant un critère de définition d'un composé. Ainsi, *blackbird* désigne un type d'oiseau et pas tous les oiseaux noirs. En revanche, le premier élément de *saucepan* désigne

---

[28] Voir la discussion en 4.2.2.2.2.4.

une seule utilisation de la casserole, alors que cet ustensile s'emploie pour bien d'autres plats. Peut-être que la réduction vocalique est le reflet de cette perte de spécificité et le signe d'une unification totale du mot qui s'écrit effectivement en un seul mot, et dont la prononciation occulte complètement l'identité du deuxième élément. Toutefois, il reste la trace écrite qui est peut-être encore suffisamment forte pour que l'on puisse considérer cette entrée comme un composé, par opposition aussi à *frying pan*. En revanche, la trace écrite de la composition pour les entrées de NPD en *-yard* (3) et en *-board* (3) ne suffit pas pour préserver leur identité[29]. Elles sont toutes devenues opaques. La réduction du second élément de *cupboard* ou de *vineyard* fait que son identité est perdue. Nous ne pouvons plus les considérer comme des composés.

Pour les mots en *-pence*, il s'agit aujourd'hui de mots qui n'ont plus cours car avec la venue du système de monnaie décimal, le *p* (prononcé /pi:/) a remplacé *pence/penny*: il est donc difficile d'évaluer à quel point nous restions conscients de l'identité des éléments malgré la réduction vocalique. La désignation linguistique des pièces de monnaie britannique d'aujourd'hui n'a pas du tout tendance à réduire; au contraire, leur prononciation adopte un schéma à accentuation double, comme s'il s'agissait d'un syntagme. Sans doute la réduction vocalique du deuxième élément désignant les anciennes pièces était-elle le résultat de longues années d'utilisation où l'on ne considérait plus la pièce mais

---

[29] Il s'agit de *halyard, lanyard* et *vineyard* pour *-yard*, et *cupboard, larboard* et *starboard* pour *-board*.

simplement la valeur. Cette hypothèse est corroborée par le fait que dans certains cas, comme pour l'exemple de *pennyworth* plus haut, il y a également une modification phonétique du premier élément : *twopence* qui se prononce *tuppence*, avec réduction sur *pence* en plus, de même que *threepence* dont la première syllabe se prononce /θrep/ d'après LPD, ou bien avec l'une des voyelles suivantes : /ʌ/, /ɪ/, ou /ʊ/ ; en aucun cas, cet élément ne se prononce /i:/. La note dans LPD (Wells, 1990) nous rappelle que le sens correspond à l'ancienne monnaie, qu'il peut toutefois être employé pour désigner la nouvelle, mais que d'habitude nous disons *ˌthree ˈpence* (en deux mots avec la prononciation des éléments comme les mots isolés). Nous en concluons que la réduction vocalique semble bien signaler une perte d'identité, et le mot résultant sera considéré comme un mot simple du lexique.

Enfin, pour les mots en *-land*, nous pourrons émettre la même hypothèse que pour les mots en *-man/men* ; cet élément serait suffixal (voir Marchand, 1969). Si nous regardons les entrées LPD en *-land* sans réduction vocalique, nous différencions d'abord celles dont le second élément est un nom et celle où il s'agit d'un verbe. Les premiers éléments peuvent être des noms (*wasteland*), des adjectifs (*wetland*), des verbes (*ploughland*). Une telle diversité ne nous paraît pas très caractéristique de la suffixation. L'identité sémantique de l'élément reste forte, même dans les cas où LPD note une variante en schwa (*farmland, homeland*). Nous ne retiendrons donc pas cette hypothèse.

4.3.2.3.1 Conclusions

Nous avons pu constater que la réduction vocalique est le reflet d'une perte d'identité sémantique du deuxième élément. Dans ce cas, nous ne pourrons plus parler de désaccentuation (voir 4.2.2.2.1, « Désaccentuation d'un élément ») mais d'élément inaccentué. Dès lors que la réduction s'accompagne d'autres modifications phonétiques sur le premier élément, le mot ne peut plus être considéré comme un composé. L'hypothèse a été avancée que ces éléments doivent être considérés comme des suffixes : dans le cas de *-man/men* le statut suffixal de l'élément semble une hypothèse convaincante[30] ; en revanche, dans d'autres cas, la simple trace écrite suffit pour préserver le statut d'élément de composition, malgré la réduction. La préservation de ce statut dépend de la fonction de l'élément qui sert à créer des catégories, comme *-pan* dans *saucepan*, ou bien à créer du sens, comme pour les mots en *-land*.

4.3.2.3.2 La corrélation entre le schéma accentuel et la phonie

Il est évident que les entrées de deux syllabes dont le deuxième élément contient une voyelle réduite à schwa sont à accentuation simple, car schwa ne porte jamais d'accent. Nous avons recherché dans LPD-Cp et N-COMPACT[31] les entrées correspondant aux éléments susceptibles de se réduire à schwa. Quasiment

---

[30] Toutefois, il faudrait étudier ce phénomène dans les dialectes comme l'anglais gallois où *man* ne se réduit pas. Ce problème a été soulevé par Marc Fryd, université de Poitiers (communication personnelle).

[31] Voir Annexe A, la requête **RECH-ELEMENT**.

toutes portent une accentuation simple, ce qui explique que la réduction soit possible.

Néanmoins, il y a quelques exceptions sur les 150 entrées concernées. Le tableau suivant les récapitule.

Figure 4.8 : Eléments susceptibles de réduire : entrées à accentuation double

land	man	stone	yard
dry land	old man	philosopher's stone	back yard
	straw man	precious stone	backyard
	young man	rolling stone	

Certaines de ces entrées sont des constructions Adj+N où l'adjectif joue un rôle purement qualificatif ; nous retrouvons ce que Ginésy (Ginésy, 1995) appelle la « nature ». Le premier élément indique la nature du second. C'est le cas de *dry land, old man, young man, precious stone*, et également de *rolling stone* (c'est ce que fait la pierre, sa nature). Huart (Huart 1989) expliquerait que la relation entre les éléments est directe.

Dans le cas de *straw man*, le premier élément indique la matière dont le second est fait ; c'est aussi sa nature (par opposition à sa fonction). Ces cas sont même cités par Wells (Wells, 1990 & 2000) comme formant une catégorie de composés à accentuation double. Le dictionnaire *Chamber's* donne comme définition « man of straw ». Nous pourrons rapprocher cet exemple des composés en *of* que nous étudierons dans la dernière partie de ce chapitre.

Les deux graphies de *back yard* (de N-COMPACT et de LPD-Cp) correspondent à d'autres exemples de type *front door*. Comme l'explique Huart,

l'accentuation double indique la localisation du deuxième élément par rapport au premier (voir 4.2.2.2.2.6)

Enfin, *philosopher's stone*, qui nous intéresse d'autant plus qu'il fait partie du titre du livre, *Harry Potter and the Philosopher's Stone*, représente un exemple de composé en *'s* que nous étudierons également dans la dernière partie de ce chapitre. En fait, son statut de composé peut être débattu, mais sa présence dans le dictionnaire s'explique par le fait que cette unité désigne un concept bien précis. L'accentuation double indique que la relation entre les éléments est une relation d'appartenance (« stone of the philosopher ») ; la relation est directe.

Sous la rubrique suivante, qui est la dernière de cette section, nous étudierons la catégorie syntaxique et la corrélation entre elle et le schéma accentuel.

4.3.2.4 Etude de la catégorie syntaxique

4.3.2.4.1 Structure des tables

Nous avons d'abord étudié les trois tables provenant de sources électroniques, EPD-Cp, NPD-Cp et LPD-Cp, afin de savoir quelles sont les catégories indiquées, et quelle proportion de la table la catégorie représente. Nous n'avons pas inclus la table N-COMPACT dans cette étude pour deux raisons : premièrement, cette table comporte un certain nombre d'entrées qui ne se conforment pas aux sources électroniques : des noms propres, des expressions et des exemples du déplacement de l'accent, qui ne permettraient pas d'interpréter les résultats selon les mêmes critères que pour les autres. Deuxièmement, la saisie de cette table s'est faite manuellement et l'étiquetage n'a pas été prévu La très

grande majorité des entrées sont des noms, mais beaucoup d'entre eux peuvent parfaitement fonctionner comme des adjectifs. Nous n'avons pas voulu nous heurter à ce problème d'ambiguïté. Toutefois, dans l'analyse d's entrées provenant de cette table, nous ne manquerons pas d'y faire référence.

Le tableau suivant récapitule les résultats de cette recherche.

Figure 4.9 : Nombre d'entrées par catégorie syntaxique et par table : EPD, NPD, LPD

Table	EPD	EPD	NPD	NPD	LPD	LPD
Nb entrées au total	*4221*		*4126*		*7154*	
Catégorie syntaxique	Nb entrées	% Entrées de la table	Nb entrées	% Entrées de la table	Nb entrées	% Entrées de la table
N	3493	82,75%	3735	90,52%	5473	76,50%
V	83	1,97%	213	5,16%	573	8,01%
Q	677	16,04%	279	6,76%	1527	21,34%
A	79	1,87%	138	3,34%	226	3,16%
Pn			25	0,61%	24	0,34%
Ex			3	0,07%	42	0,59%
tdmk					152	2,12%
TOTAL	*4332*		*4393*		*8017*	
Plurifonctionnelles	111		267		863	
% Plurifonctionnelles	2,56%		6,08%		10,76%	

Le premier constat qui s'impose concerne la variété des catégories représentées. EPD ne contient que les quatre premières catégories : « N », nom, « V », verbe, « Q », adjectif et « A », adverbe. En revanche, NPD et LPD contiennent toutes les deux des entrées étiquetées « Pn » (pronom) et « Ex » (exclamation). De plus, LPD inclut la catégorie « tdmk » (*trademark* pour les noms de marques) qui représente un pourcentage non négligeable des entrées (plus de 2%). Ces différences soulignent l'importance du problème du choix des entrées. Dans un dictionnaire de prononciation, il est souhaitable d'inclure des entrées dont l'utilisateur aura besoin, d'où l'inclusion des noms de marque dans le

dictionnaire le plus récent[32]. L'exclusion des catégories «pronom» et «exclamation» de EPD peut aussi être le résultat non d'un choix délibéré mais du choix de la graphie qui n'inclut que des entrées avec un trait d'union; mais cette hypothèse est peu convaincante dans la mesure où, sur les 42 entrées étiquetées «Ex» de LPD, 28 ont un trait d'union.

Le second constat concerne le pourcentage que représente chaque catégorie par rapport à la table. Il est vrai que les tendances sont les mêmes pour toutes: une très forte proportion de noms (entre 76% et 91%); puis un nombre conséquent d'adjectifs (de 7% à 21%). La proportion de verbes et d'adverbes est similaire dans EPD (moins de 2%) alors que pour LPD et NPD il y a davantage de verbes (LPD en a 8% contre 3% d'adverbes; NPD en a 5% contre un peu plus de 3%).

Les chiffres révèlent également l'existence d'un certain nombre d'entrées plurifonctionnelles, puisque le nombre total des entrées par catégorie dépasse le nombre d'entrées de la table. Ce fait n'est pas étonnant mais il est intéressant de noter les différences entre les tables: EPD a relativement peu d'entrées plurifonctionnelles: 2,5% de la table; NPD en a 6% et ces entrées représentent plus de 10% des entrées de LPD. Ce constat peut être éventuellement relié au fait que LPD a une proportion relativement faible de noms comparativement aux autres tables; c'est-à-dire que les entrées «N» seraient aussi marquées «Q» ou «V» plus régulièrement que dans les autres dictionnaires. Cette hypothèse n'est absolument pas corroborée par la recherche sur éléments qui fera l'objet

---

[32] L'indication « tdmk » est p ut-être aussi imposée par des contraintes d'ordre légal.

d'analyses dans la dernière section de ce chapitre. Les différences d'étiquetage s'expliqueraient plutôt par le choix même des entrées.

4.3.2.4.2 Conclusion

L'étude de la distribution des différentes catégories syntaxiques dans chaque table permet de nous rendre compte de la forte proportion de noms, puis d'adjectifs (mais beaucoup moins) dans ces ensembles de composés; les autres catégories étant très minoritaires (de 4% à 14%, y compris les entrées plurifonctionnelles qui sont souvent et verbe et nom, par exemple).

En étudiant les différences dans la distribution des catégories par table, nous avons pris conscience une nouvelle fois du problème de la sélection des entrées. S'il est vrai qu'il y a beaucoup plus de composés appartenant à la catégorie «nom», cette proportion ne correspond pas à celle des entrées NPD (plus de 90%). Le peu d'adjectifs notés par ce dictionnaire est sûrement relié au choix des entrées, toutes écrites en un seul mot alors que la graphie des adjectifs est souvent avec un trait d'union. (62% des adjectifs de LPD-Cp s'écrivent ainsi). Il est vrai que le sens de beaucoup d'adjectifs peut se deviner à partir de leurs composants, mais qu'en est-il de leur schéma accentuel, puisque cet aspect est primordial pour un dictionnaire de prononciation des composés? Sous la rubrique suivante, nous allons étudier la corrélation entre la catégorie syntaxique et le schéma accentuel.

4.3.2.4.3 Corrélation entre le schéma accentuel et la catégorie syntaxique

Afin d'étudier la corrélation entre le schéma accentuel et la catégorie syntaxique, nous avons analysé les pourcentages d'entrées ayant une accentuation simple et une accentuation double pour chaque catégorie. Nous avons également

étudié ces chiffres par rapport au nombre de syllabes pour être en mesure de comparer ces résultats à ceux de l'ensemble des tables (voir 4.3.2.1.1.1, « Etude du schéma accentuel : Structure des tables », et Annexe B). Le tableau suivant ne reprend que les résultats pour l'ensemble de la table et pour l'ensemble de la catégorie ; un tableau récapitulatif de l'étude plus détaillée, syllabe par syllabe, se trouve en Annexe B.

Les chiffres entre parenthèses qui suivent les quatre catégories principales concernent le nombre d'entrées qui sont marquées comme ayant une tendance au déplacement de l'accent.

Figure 4.10 : Nombre d'entrées par catégorie syntaxique et par type de schéma, par table

Table	EPD-Cp		NPD-Cp		LPD-Cp	
% Entrées de la table avec schéma « s » / « d »	71,12% / 28,88%		94,69% / 5,31%		72,49% / 27,51%	
Catégorie & schéma	*Nb entrées*	% Entrées de la catégorie	*Nb entrées*	% Entrées de la catégorie	*Nb entrées*	% Entrées de la catégorie
« N » avec schéma « s » / « d »	*2737 / 756*	78,36% / 21,64%	*3633 / 102*	97,27% / 2,73%	*4739 / 734 (291)**	86,59% / 13,41%
« V » avec schéma « s » / « d »	*48 / 35*	57,83% / 42,17%	*181 / 32*	84,98% / 15,02%	*287 / 286 (76)**	50,09% / 49,91%
« Q » avec schéma « s » / « d »	*250 / 427*	36,93% / 63,07%	*238 / 41*	85,30% / 14,70%	*529 / 998 (938)**	34,64% / 65,36%
« A » avec schéma « s » / « d »	*1' / 68*	13,92% / 86,08%	*62 / 76*	44,93% / 55,07%	*69 / 157*	30,53% / 69,47%
« Pn » avec schéma « s » / « d »			*8 / 17*	32,00% / 68,00%	*19 / 5*	79,17% / 20,83%
« Ex » avec schéma « s » / « d »			*0 / 3*	0,00% / 100,00%	*8 / 34*	19,05% / 80,95%
« tdmk » avec schéma « s » / « d »					*135 / 17*	88,82% 11,18%

Si nous comparons les tendances accentuelles selon la catégorie entre les différentes tables, nous en concluons que les entrées appartenant à la catégorie

«N» (nom) sont les plus susceptibles de porter une accentuation simple. Ce résultat corrobore les remarques faites en 4.3.1 à propos des doublons qui appartiennent à des catégories différentes : nous avons constaté que, dans ces cas, les noms ont une accentuation simple. Si nous comparons ces pourcentages avec ceux pour l'ensemble de chaque table, nous constatons également que la proportion de «N» ayant une accentuation simple est systématiquement plus grande pour cette catégorie qu'elle ne l'est pour l'ensemble. Nous pouvons en conclure que cette catégorie tend à porter une accentuation simple.

La tendance à l'accentuation simple diminue successivement pour les catégories «V» (verbe), puis pour les «Q» (adjectifs), et enfin les «A» (adverbes), sauf dans la table NPD où la proportion des types d'accentuation est semblable pour les adjectifs et pour les verbes. Nous avons déjà fait remarquer le peu d'adjectifs que comporte cette table et nous avons émis l'hypothèse d'une sélection des composés très restrictive qui est reflétée par l'orthographe (en compact) ce qui expliquerait cette disparité. Si nous écartons ce résultat de NPD qui semble être marginal pour ces raisons, nous pouvons en conclure que la catégorie des adjectifs montre une forte tendance à l'accentuation double (*new-born*, *nightlong*) ; la catégorie des verbes aurait un schéma fréquemment à accentuation double (*underestimate*, *spotcheck*), mais les chiffres montrent qu'au moins la moitié (et beaucoup plus selon les tables) ont une accentuation simple (*blow-dry*, *shortlist*).

En ce qui concerne les adjectifs, il est nécessaire de tenir compte des indications sur le déplacement de l'accent que nous trouvons dans LPD[33]. L'entrée aurait un type d'accentuation (double) lors d'une prononciation hors contexte, et un autre (simple) en discours, plus précisément lorsque l'adjectif est en position d'épithète. Si nous regardons le tableau, le nombre d'entrées concerné par ce phénomène est très important. Si nous considérons que les entrées entre parenthèses sont susceptibles de porter une accentuation simple, le pourcentage des adjectifs avec ce type d'accentuation est de 97%, ce qui est encore plus important que pour les « N ». Le phénomène touche également d'autres catégories : les « N » justement (286 entrées en tout, y compris les plurifonctionnelles, qui sont souvent « Q ») ; les « V » (287 entrées en tout) ; enfin, les « A » qui ont une fonction adjectivale, et qui sont donc susceptibles de subir un déplacement de l'accent, représentent au total 157 entrées. Ce phénomène a donc un effet très important sur la perception de l'accent en discours. Nous l'étudierons de près dans l'analyse acoustique (voir 4.5.3.1, « Le déplacement de l'accent »), et nous en évaluerons sa portée.

Il est intéressant de noter que pour la catégorie « Pn » (pronom), la proportion des schémas de type « s » et de type « d » est inversée entre NPD et LPD. Ce résultat s'explique par le choix des entrées elles-mêmes. Nous trouvons

---

[33] Voir la requête **SHIFT-LPD**, Annexe A. Dans la version électronique, le symbole « < » indique le déplacement de l'accent, qu'il ne faut pas confondre avec « > < » qui correspond au point d'exclamation à l'intérieur d'un triangle de la version papier, indiquant une prononciation qui n'est pas recommandée. Il existe des entrées avec les deux symboles. Certaines entrées avec une accentuation de type « s » comporte ce symbole : il s'agit d'une indication portant sur la variante en « d » de l'entrée.

de nombreuses entrées en *-self* dans NPD, et une seule dans LPD (absente de NPD !), ce qui augmente considérablement la proportion des schémas à accentuation double. En ce qui concerne les « tdmk », nous remarquons qu'il y a une forte tendance à l'accentuation simple ; nous reviendrons à ces cas dans l'analyse des entrées noms propres (4.6.2.3, « Les entrées noms propres de la base de données »).

Enfin, la tendance que nous avions notée sous la première rubrique de cette section (4.3.2.1, « Etude du schéma accentuel »), que les entrées ayant le plus de syllabes ont tendance à avoir une accentuation double, se confirme dans l'étude des schémas par catégorie et par syllabe. Le tableau en Annexe B montre clairement que dans certains cas, pour les entrées de 2 et de 3 syllabes, les pourcentages se tiennent, mais à partir de 4 syllabes la proportion de « s » chute et la proportion de « d » augmente. C'est le cas de LPD pour les catégories « tdmk » et « A » ; de NPD pour les « N », les « V », les « A » et les « Pn » ; enfin, de EPD pour les « N » et les « V ».

4.3.2.4.3.1 Conclusion

Dans cette section, nous avons pu rendre compte de la corrélation entre la catégorie syntaxique et le schéma accentuel. En fait, nous avons surtout noté des tendances : les noms ont souvent une accentuation simple, les adjectifs ont une accentuation double hors contexte. Toutefois, ces tendances ne sont pas suffisamment étendues pour que l'on puisse conclure que telle catégorie est associée à tel type d'accentuation.

Nous avons à nouveau pris conscience qu'il est important de tenir compte de phénomènes de discours tels que le déplacement de l'accent. D'après les indications de LPD que nous avons étudiées, cela peut changer radicalement les tendances accentuelles de certaines catégories. Nous nous posons la question de l'opportunité de faire référence à ce genre de phénomène de discours dans les indications de l'accent lexical. Nous reviendrons à ces considérations dans la partie suivante.

Enfin, les résultats soulignent l'influence du choix des entrées sur ce genre d'étude statistique. La proportion de noms ou d'adjectifs incluse dans la table varie très considérablement d'une source à l'autre; elle ne varie pas dans le lexique que nous utilisons tous les jours. Il faut donc se fixer des critères de sélection cohérents. Dans la dernière section de cette partie, nous étudierons l'évolution du lexique; le premier point que nous aborderons porte sur son contenu.

4.3.3 Etude de l'évolution du lexique

Notre méthode consiste à comparer les tables provenant de la première édition du *Longman Pronouncing Dictionary*, LPD et N-COMPACT, et la table EPD afin d'étudier les similarités et les différences entre elles à plusieurs niveaux : la graphie, le schéma accentuel et le contenu. Nous avons analysé les entrées commençant par « a » et par « b », absentes de l'autre source, afin de comprendre les raisons de leur exclusion. Cela représente plus de 700 entrées de LPD et près de 200 de EPD. Nous avons comparé ces entrées avec le dictionnaire complet EPD12 ainsi qu'avec la dernière édition de EPD, EPD15, parue en 1997, pour

évaluer la portée des changements. Pour cette raison, nous précisons EPD12 ou EPD15 dans cette section pour qu'il n'y ait pas de confusion. Nous avons également fait une étude à partir de certains éléments qui se trouvent dans les extraits sonores afin d'être en mesure de tirer des conclusions générales sur ces exemples[34]. Cette analyse sera développée dans les parties 4 et 6 de ce chapitre, mais nous rendons compte ici des conclusions qui concernent l'évolution du lexique.

Nous commencerons par étudier les différences au niveau du contenu, et donc du choix des entrées de chaque table ; ensuite, nous aborderons les modifications de la graphie, puis du schéma accentuel afin d'étudier l'éventuelle corrélation entre les deux.

4.3.3.1 Le contenu

4.3.3.1.1 Les entrées manquantes dans LPD

Nous avons déjà pu constater l'importance des changements entre EPD12 et LPD à partir des chiffres comparatifs des tables (4.3.2.2.2. « La structure des tables »). LPD contient beaucoup d'entrées supplémentaires (plus de 10 000) mais il y a environ 1800 entrées de EPD12 qui ne sont pas présentes dans LPD. Certaines de ces entrées manquantes correspondent à des erreurs : les entrées ayant un trait d'union ne sont pas toutes des composés, comme nous l'avons montré dans l'évaluation de ce critère d'extraction (voir 2.2.2.3. « Extraction des composés selon des critères graphiques ») : des mots tels que *ack-ack*, par

[34] Voir aussi les indications sur l'accent lexical des extraits et de leurs composants en Annexe C.

exemple, ou des préfixés en *arch-* et *by-* que nous avons écartés (voir 2.1.2.1.1.1, « Les préfixes »). Certaines entrées de EPD12 sont des dérivés dont seul le radical figure dans LPD ou EPD15 : *air-cooled* dans EPD12 correspond à *air-cool* dans EPD15 et LPD.

Sur les 179 entrées composées de l'échantillon « a » et « b » de EPD12 qui manquent dans LPD, presque la moitié (83 entrées) manquent aussi dans EPD15. Cette table contient certainement des entrées obsolètes faisant référence à des techniques qui n'ont plus cours. L'étude des « a » et « b » ne nous a pas éclairée sur ce point ; en revanche, nous avons relevé plusieurs exemples de techniques ou d'appareils peu courants qui ne sont repris ni dans LPD ni dans EPD15 : *air-bath* (un mode thérapeutique) et *air-pump* (un appareil pour pomper l'air), par exemple. Ces dictionnaires récents ont également choisi d'omettre *altar-cloth*, pourtant tout autant en usage pour recouvrir l'autel qu'il ne l'était dans les années cinquante, à l'époque de la sortie de EPD12. L'étude des éléments entrant dans la composition des extraits sonores est plus révélatrice. Nous avons trouvé *carpet-broom* dans EPD12, qui a été définitivement remplacé par l'aspirateur, et *talking-film*, sans doute par opposition à *silent film*[35]. Aujourd'hui, il suffit de parler de *film*, le fait qu'il ne soit pas muet étant implicite.

Enfin, nous voudrions commenter quelques entrées qui figurent toujours dans EPD15 mais qui sont absentes de LPD. Le composé *booking office*, (la

---

[35] *Silent film* ne figure dans aucun des dictionnaires de prononciation, sans doute parce qu'il est considéré comme un simple adjectif suivi d'un nom, et son accentuation est donc parfaitement prévisible.

graphie est de EPD15) qui porte une accentuation simple est exclu de LPD. Quel est le critère de sélection ? Sûrement pas le type d'accentuation, car dans ce cas, il serait inclus puisque l'accentuation simple est caractéristique des composés par opposition aux syntagmes. Peut-être un critère sémantique s'applique-t-il, selon lequel les entrées dont le sens de l'ensemble se comprend aisément à partir de celui des composants seraient exclues ; mais si on devait appliquer ce critère, on retirerait un très grand nombre d'entrées. Nous percevons la difficulté importante qui consiste à sélectionner les composés.

Nous citons un autre exemple d'une entrée présente dans EPD12 et EPD15 mais qui manque dans LPD. L'entrée *apple tree* (la graphie est de nouveau de EPD15) est exclue tout comme *cherry tree* ou *pear tree*. Les deux derniers ne figurent pas non plus dans les deux éditions de EPD ce qui souligne le problème de la cohérence des choix d'entrées. Ici, LPD décide d'exclure tous ces composés (à accentuation simple) car il s'agit d'une famille dont les caractéristiques sont identiques. Nous précisons qu'il est logique que les entrées de EPD15 soient plus proches de la 12ème édition que ne l'est LPD car il s'agit du même dictionnaire : le choix se présente donc comme une reconduction ou une exclusion ; en revanche, le choix des entrées de la première version de LPD était libre.

4.3.3.1.2 Les entrées manquantes dans EPD12

Comme nous l'avons déjà signalé en 4.3.2.2.3.1. « Le contenu des tables. Critères de sélection des composés », il manque dans la table EPD-Cp tous les composés compacts du dictionnaire, puisque le critère d'extraction de cette table était la présence du trait d'union. L'étude des éléments entrant dans la

composition des extraits sonores est révélatrice : sont absents de la table des « composés » les mots *anyone, anything, armchair, broomstick, workshop* et, comme nous l'avons signalé, *greenhouse*. Tous figurent dans le dictionnaire mais non dans la table car ils s'écrivent sans séparateur.

Il manque aussi un grand nombre d'entrées formées à partir de particules telles que *-off, -up, -out*, comme dans les exemples suivants : *browned-off* (« Q »), *brush-up* (« NV »), *brownout* (« N »). EPD12 ne semblent pas inclure beaucoup d'entrées de ce type ; celles-ci manquent toutes dans la version papier sous quelle que graphie que ce soit.

Les entrées de LPD qui manquent dans EPD12 dépassent très largement le critère graphique de la sélection Nous avons vu sous la rubrique précédente « catégorie syntaxique » que LPD inclut des noms de marques. Dans l'analyse des « a » et « b », nous avons relevé plusieurs cas dont *BluTack* et *Band-Aid*. Ces deux exemples relèvent de deux autres facteurs qui expliquent l'absence de ces entrées dans EPD12 : d'abord les avances technologiques avec *BluTack*, une sorte de pâte collante, qui n'existait sans doute pas dans les années cinquante ; ensuite les phénomènes de société car le principe de *Band-Aid*, où les chanteurs font des concerts afin d'aider une bonne cause, a été lancé par Bob Geldorf il y a une vingtaine d'années. Un autre exemple de progrès technologique est le *babybouncer*, qui permet aux bébés de bien s'amuser en toute sécurité. Pour les phénomènes de société, nous tenons à citer *bra-burner*, type féminin apparu au moment du *Women's Lib* des années soixante-dix.

Enfin, certaines entrées rappellent des expressions toutes faites qui ne figurent pas dans EPD12: *bread-and-butter*, *bumper-to-bumper* (aussi un phénomène de société ?) et *by-your-leave*, par exemple. Comme ces unités sont très souvent utilisées, il est souhaitable effectivement de les inclure dans un dictionnaire de prononciation.

L'étude des éléments composant les extraits sonores apporte des éclairages supplémentaires, mais soulève également des questions. Les unités correspondant à des phénomènes de la société actuelle sont absentes de EPD12 : *greenhouse gas* (marquée par EPD15 seulement) et *greenhouse effect* (marquée dans les deux versions de LPD et EPD15). Toutefois, nous nous demandons si EPD12 aurait inclus ces entrées, proches de syntagmes par leur graphie : *bird of prey* ne figure pas non plus, alors qu'il ne s'agit pas d'une évolution de la société ; cette unité est présente dans les autres dictionnaires. La graphie n'explique pas l'absence de *kitchen-sink* ; mais peut-être cette unité a-t-elle été incluse dans les autres dictionnaires à partir des pièces de théâtre connues sous l'appellation *kitchen-sink drama* ?

Les mots *microchip* et *nonsmoker* ne figurent pas non plus dans EPD12 alors qu'ils sont présents dans les autres dictionnaires. Ce ne sont pas des composés, mais le choix de les écarter du dictionnaire doit être commenté car si l'on n'inclut pas les quasi-lexèmes et les préfixés de manière systématique, cela représente un gain de place considérable. Ce n'est pas le cas de EPD12 qui sélectionne certains mots formés avec *micro-* (*microfilm, micro-organism*) et *non-* (*non-resident*), mais pas ceux-ci. Ces cas peuvent s'expliquer : pour *microchip*, il

s'agit d'avance t[illegible]hnologique alors que pour *nonsmoker*, qui est une véritable catégorie de personnes dont la désignation, à l'oral comme à l'écrit, est devenue nécessaire.

L'absence de deux autres mots nous a surpris : *classroom, washroom*. Nous pensions qu'ils seraient présents dans le dictionnaire avec une graphie sans séparateur mais ce n'est pas le cas. Pour *washroom*, son absence peut s'expliquer par le fait que le mot est plutôt américain alors que les entrées de EPD12 sont vraiment des mots d'usage britannique. En revanche, il n'y a pas d'explication pour *classroom* si ce n'est que le dictionnaire a choisi d'inclure les mots sans contexte particulier et que ce mot se comprend uniquement dans un contexte d'école, mais il s'agit sans doute d'un oubli[36].

Sous cette rubrique, nous avons été amenée à discuter des critères de sélection des entrées d'un dictionnaire des composés. Nous reviendrons à ces considérations dans notre conclusion. Nous passons à la deuxième rubrique de cette section où nous allons étudier des facteurs très concrets : les changements de la graphie et du schéma accentuel.

4.3.3.2 La graphie et le schéma accentuel

4.3.3.2.1 La graphie

Nous rappelons brièvement les faits que nous avons montrés dans l'étude de la graphie dans la section précédente. La comparaison des tables de EPD12 et de

---

[36] Dans la 15ème édition d'EPD, ce mot est inclus, sans séparateur.

LPD et N-COMPACT révèle que sur les 2376 entrées communes, seulement 585 ne changent pas de graphie. Celle-ci se modifie dans les deux sens : 1098 entrées s'écrivent en compact (sans séparateur) et 693 s'écrivent en non compact (avec un espace). Les graphies des entrées communes ont la distribution suivante :

Figure 4.11 : Entrées communes EPD/LPD, par type de graphie

Type de graphie	Nb entrées au total	Graphie sans séparateur	Graphie avec trait d'union	Graphie avec espace
Nb entrées	2376	1098	585	693
% Entrées communes		46,21%	24,62%	29,17%

Seule une entrée sur quatre ne change pas. Il n'est pas possible de fournir des statistiques sur la graphie de EPD15 car nous ne disposons pas de fichier électronique. Toutefois, sous la rubrique où nous étudierons la corrélation entre la graphie et le schéma, nous y ferons systématiquement référence.

4.3.3.2.2 Le schéma accentuel

Nous nous sommes intéressée aux cas où le schéma accentuel est différent pour la même entrée dans EPD12 et LPD. Notre méthode est la même que pour l'analyse du contenu des tables, c'est-à-dire que nous avons d'abord comparé les tables de la base de données[37], puis nous avons procédé à une vérification de nos résultats en consultant la dernière édition de EPD, EPD15.

[37] Voir requête **comp-LPD-EPD**, Annexe A.

Figure 4.12 : "Evolution" du schéma accentuel EPD12/LPD

Nb et % d'entrées communes	Nb entrées	Nb entrées changées	Nb entrées **EPD12 « s » → LPD « d »**	Nb entrées **EPD12 « d » → LPD « s »**
Entrées communes EPD12 / LPD & N-COMPACT	2376	223	91	132
% Schéma accentuels changés		9,39%	3,83%	5,56%

Ces résultats, obtenus à partir de la comparaison entre ces tables, montrent que le phénomène semble assez étendu: au total plus de 9% des entrées communes sont différentes. Nous nous attendions à un changement dans le sens de la rétraction germanique, donc de l'accentuation double vers l'accentuation simple. Ce serait la tendance naturelle de l'anglais et conforterait l'hypothèse de l'accentuation simple comme étant le type d'accentuation normale des composés. Cependant, près de 4% des changements se font dans le sens inverse. Nous allons d'abord étudier les entrées qui se modifient dans le sens de l'accentuation simple vers l'accentuation double, puis celles qui subissent le changement inverse.

4.3.3.2.2.1 Accentuation simple vers accentuation double

Nous avons noté 91 entrées qui diffèrent entre EPD12 et LPD au niveau du schéma accentuel: le schéma de EPD12 est « s » alors que celui de LPD est « d ». Ces résultats nous ont surprise car ils vont dans le sens opposé de la tendance habituelle. En comparant ces entrées avec celles de EPD15, nous constatons que 76 d'entre elles sont également « d », comme LPD, et 4 sont absentes du dictionnaire. Seules 11 entrées de EPD15 auraient le même schéma accentuel que EPD12.

Nous étudierons d'abord la variation du schéma accentuel, puis les indications concernant le déplacement de l'accent, afin de savoir si nous pouvons parler d'une véritable évolution ou si ces différences s'expliquent d'une autre manière. Nous nous référerons aux chiffres donnés en Annexe B.

4.3.3.2.2.1.1 La variation

Lorsque les entrées sont marquées « d » dans LPD et EPD15, nous avons un taux de variation relativement important pour LPD (31 entrées sur 91, soit le tiers environ); EPD15 en a moins (10 sur 76) mais indique des prononciations différentes américaines pour 23 d'entre elles. Nous pouvons en conclure que ces entrées sont instables en ce qui concerne leur accentuation. Cette hypothèse est corroborée par le taux important de variation qui est signalé dans EPD12 (20 cas sur 91, soit 22% des entrées).

4.3.3.2.2.1.2 Le déplacement de l'accent

Si l'instabilité accentuelle est une hypothèse valable, la tendance au déplacement de l'accent fournit une explication encore plus convaincante. Ce phénomène reçoit un traitement beaucoup plus rigoureux dans les dictionnaires récents : donc les adjectifs signalés comme « s » dans EPD12 sont signalés « d » dans LPD et EPD15 de façon systématique, avec une indication de déplacement de l'accent (le symbole ◄ dans LPD, et *stress shift* dans EPD15). C'est le cas pour 68% des entrées de LPD (62 entrées) et 56% (51 entrées) pour EPD15. Ainsi, la majorité des différences du schéma accentuel dans le sens « s » pour EPD12 et « d » pour LPD s'explique par le fait que EPD12 donnait un schéma « s » à des adjectifs qui auront effectivement ce schéma en discours, mais qui sont marqués

« d » par LPD (et EPD15) dans la mesure où cela correspond à leur prononciation en tant que mot isolé ; le dictionnaire signale la tendance au déplacement de l'accent. L'étude des éléments permet de fournir un exemple : *green-eyed* a un schéma « s » dans EPD12 alors qu'il a un schéma « d » (plus ◀) dans LPD.

4.3.3.2.2.2 Accentuation double vers accentuation simple

Nous étudions l'autre cas de figure : EPD donne un schéma « d » et LPD donne un schéma « s », ce qui correspond à la tendance normale à la rétraction de l'accent. Sur les 132 entrées différentes, EPD15 donne un schéma « s » comme LPD pour 117 d'entre elles, 9 entrées sont absentes ; il ne subsiste que 6 entrées pour lesquelles EPD15 donne le même schéma que EPD12. Nous en concluons que les dictionnaires des années quatre-vingt-dix concordent très largement. Nous allons analyser ces cas de la même manière que précédemment en étudiant d'abord la variation puis le déplacement de l'accent.

4.3.3.2.2.2 1 La variation

Nous constatons que le nombre d'entrées indiqué avec une variante en « s » dans EPD12 est très important : il y en a 71, c'est-à-dire plus de la moitié des entrées. L'étude des éléments fournit l'exemple *motor-bike*, qui a un schéma « d » mais une variante en « s ». Il est intéressant de noter que sa graphie devient une graphie compacte dans LPD avec un schéma « s » sans aucune variante. En revanche, le taux de variation dans LPD et EPD15 est faible : ils signalent 21 et 11 cas respectivement (16% et 8%). Ces résultats nous permettent de conclure qu'il y a une évolution caractérisée du schéma accentuel, et non l'instabilité accentuelle que nous avons constatée précédemment.

4.3.3.2.2.2.2 Le déplacement de l'accent

Nous rappelons que le phénomène de déplacement accentuel ne concerne que les entrées à accentuation double[38]. Il n'y a donc pas de cas de déplacement signalés dans LPD et EPD15. En revanche, le schéma accentuel de EPD12 est « d » et le dictionnaire signale 8 entrées susceptibles de subir une rétraction de l'accent.

4.3.3.2.2.3 Conclusion

L'étude des entrées dont le schéma accentuel diffère entre EPD12 et LPD permet d'émettre deux hypothèses. Premièrement, l'instabilité accentuelle explique certains cas : le taux de variation des entrées marquées « s » dans EPD12 et « d » dans LPD confirme cette hypothèse. Par ailleurs, le nombre important de ces entrées s'explique par le traitement différent des adjectifs susceptibles de subir un déplacement de l'accent. EPD12 donnait directement la prononciation en discours (donc « s ») alors que les dictionnaires d'aujourd'hui donnent d'abord la prononciation à l'état isolé (« d ») avec une indication de déplacement de l'accent. Cette méthode plus rigoureuse nous semble la plus appropriée, et elle fait apparaître de nouveau la nécessité de tenir compte des phénomènes de discours dans les indications de l'accent lexical.

La seconde hypothèse peut être formulée à partir de l'analyse des entrées dont le schéma passe de « d » vers « s ». Il s'agit effectivement d'une évolution du

---

[38] C'est pourquoi, lorsque la prononciation GB est « d » et la prononciation US est « s », le dictionnaire indique que le déplacement accentuel concerne seulement la prononciation britannique : *British English only*.

schéma accentuel. Nous relevons plusieurs exemples grâce à l'analyse des éléments : *dark-room* qui est accentué « d » d'après EPD12 comme s'il s'agissait simplement d'une pièce sombre et qui devient « s » dans LPD (le schéma actuel souligne la fonction de la pièce plutôt que la caractéristique descriptive) ; la graphie se modifie également, et le mot s'écrit en compact. Les mêmes faits se remarquent dans *club-room* ou *hill-top* qui passent eux aussi de « d » à « s » avec une graphie modifiée. Sous la rubrique suivante, nous allons étudier la corrélation entre ces différences accentuelles des dictionnaires et de la graphie.

4.3.3.2.3 La corrélation entre le changement de schéma et la graphie

Afin d'étudier l'éventuelle corrélation entre le changement du schéma accentuel et de la graphie, nous avons adopté la même méthode, en tenant compte des entrées communes de LPD90, puis en comparant ces données avec celles de EPD15. Le tableau suivant permet de récapituler les résultats[39].

[39] Voir aussi le tableau en Annexe B qui récapitule la ventilation des graphies de EPD15 par rapport à LPD.

Figure 4.13: Graphie des entrées dont le schéma accentuel est différent EPD12 / LPD

Les entrées différentes : EPD12 «s» / LPD «d» 91 entrées	**EPD12**	**LPD**	**EPD15**
Graphie en compact		15	4
Graphie avec trait d'union	91	54	66
Graphie avec espace		22	16
Entrées absentes			5
Nb Entrées avec une graphie changée		37	20
% Entrées avec une graphie changée		**40,66%**	**21,98%**
Les entrées différentes : EPD12 «d» / LPD «s» 132 entrées	**EPD12**	**LPD**	**EPD15**
Graphie en compact		56	63
Graphie avec trait d'union	132	25	17
Graphie avec espace		51	43
Entrées absentes			9
Nb Entrées avec une graphie changée		107	106
% Entrées avec une graphie changée		**81,06%**	**80,30%**

Etudions d'abord les cas où le schéma de EPD12 est « s » et celui de LPD est « d ».

4.3.3.2.3.1 Accentuation simple vers accentuation double

Nous avons déjà constaté que les différences entre les dictionnaires des années 90 et EPD12 reflètent souvent un traitement plus rigoureux du déplacement de l'accent, ou bien une instabilité accentuelle. Il n'y a pas de véritable changement. De même, la graphie ne subit pas de changement notable dans EPD15: 20% seulement des graphies ont subi des modifications, très majoritairement vers une graphie avec un espace (16 entrées contre 4 écrites en compact). Les changements dans LPD sont plus nombreux : 40% des entrées, dont 22 vers une graphie avec un espace et 15 sans séparateur. Ces chiffres confirment la tendance que nous avons constatée à une corrélation entre l'accentuation double et la graphie non compacte, mais ce n'est pas très net. Nous notons que LPD

introduit davantage de modifications que EPD, ce qui est attendu dans la mesure où il s'agit d'un dictionnaire nouveau, et non d'une version nouvelle du même dictionnaire, comme c'est le cas pour EPD15.

4.3.3.2.3.2 Accentuation double vers accentuation simple

En revanche, les changements du schéma accentuel dans le sens d'une accentuation simple vers une accentuation double sont révélateurs d'une véritable évolution. Cette "évolution" est reflétée par des changements au niveau de la graphie dans plus de 80% des cas dans les deux dictionnaires, EPD15 et LPD90. Curieusement, la graphie change presque autant vers une graphie avec l'espace que vers une graphie en compact : 56 entrées écrites en compact dans LPD contre 51 avec un espace. La proportion est légèrement plus à l'avantage de la graphie en compact dans EPD15 : 63 entrées sont écrites sans séparateur contre 43 avec un espace.

4.3.3.2.4 Conclusions

Ces résultats ne nous permettent pas d'émettre d'hypothèses plus précises concernant la corrélation entre le type de graphie et le type de schéma. Cependant, nous remarquons la tendance vers une graphie autre que le trait d'union pour les entrées ayant une accentuation simple. Peut-être que la graphie avec un trait d'union est aujourd'hui associée à l'accentuation double ; la graphie en non compact pouvant être associée indifféremment aux deux schémas.

Dans cette section, nous avons soulevé à plusieurs reprises l'importance de la prise en compte du discours. Avant de procéder à l'analyse acoustique des

extraits sonores, nous allons étudier les indications de l'accent lexical par rapport aux schémas perçus pour ces extraits afin d'en évaluer la pertinence.

4.4 L'accent lexical et le corpus oral

Un des objectifs de nos travaux est d'évaluer la pertinence des indications de l'accent lexical par rapport au comportement des unités en discours. Dans cette section, nous nous intéresserons aux indications que donnent les dictionnaires[40] en ce qui concerne les extraits sonores qui composent le corpus oral et qui feront l'objet de l'analyse acoustique dans la section suivante. Nous tenterons de répondre aux deux questions suivantes : à quel point ces indications sont-elles fiables ? Et dans quelle mesure sont-elles complètes ?

4.4.1 Le taux de lexicalisation des extraits sonores

Les extraits sonores proviennent de deux types de sources différents : les sources documentaires et la source littéraire. Un récapitulatif des extraits (y compris les indications de l'accent lexical) est donné en annexe C de ce chapitre. Sur les 113 extraits sonores, il n'y en a que 41 pour lesquels nous avons trouvé des indications dans un dictionnaire de prononciation : 10 extraits sur 56 pour la source littéraire ; et 31 sur 57 pour les sources documentaires[41]. Parfois, ces indications ne sont valables que pour une partie de l'extrait, dans le cas des unités

---

[40] Nous utilisons les abréviations suivantes : EPD12 et EPD15 (12ème et 15ème éditions du *English Pronouncing Dictionary*) ; LPD90 et LPD00 (édition parue en 1990 et en 2000 du *Longman Pronouncing Dictionary*).

[41] Ces chiffres correspondent à chaque extrait sonore qui fait l'objet d'une analyse (et dont l'accent en discours peut varier). Ainsi, dans le cas de *armchair*, par exemple, qui apparaît quatre fois dans le corpus oral, cela représente trois des extraits lexicalisés de la source littéraire, et un des extraits de la source documentaire.

ayant une construction cumulative (des « surcomposés »). Dans ces cas, le tableau en annexe indique la prévision de l'accent entre parenthèses, car elle ne concerne qu'une partie de l'extrait.

Le faible taux de lexicalisation des extraits provenant de la source littéraire reflète les critères de sélection que nous avons adoptés pour cette source, nos objectifs étant l'analyse de néologismes comme nous l'avons expliqué dans le chapitre 2 (2.3.2.2, « Source littéraire »). Des unités comme *school broom, portrait hole* ou encore *flying-key room* sont intéressantes parce qu'elles ressemblent de par leur formation et leur sens à des composés lexicalisés ; nous pourrons donc étudier la prévision de l'accent sur ces unités en fonction des règles des composés existants. Cette étude sera développée dans la suite de ce chapitre.

En ce qui concerne les sources documentaires, plus de la moitié des extraits sont (du moins partiellement) lexicalisés. Néanmoins il subsiste 26 extraits qui ne figurent pas dans un des dictionnaires de prononciation que nous avons consultés. Ces absences peuvent-elles s'expliquer par des critères de sélection transparents ? Les extraits non lexicalisés se divisent en trois catégories : les unités ayant une construction cumulative, les unités correspondant à des sens nouveaux, à des inventions (donc des néologismes), puis celles qui ressemblent aux composés lexicalisés.

Dans le reportage sur l'invention du capuchon parlant, nous avons relevé plusieurs extraits ayant une construction cumulative : *talking bottle top, vodka bottle top, talking vodka bottle top*. Il est évident qu'il n'est pas envisageable pour un dictionnaire d'inclure ce type d'entrées, car nous pouvons fabriquer sur le

même schéma *talking+N* des centaines d'unités. Dans la mesure où leur sens est parfaitement prévisible, le dictionnaire fera l'économie de ce type d'entrée, sauf peut-être dans le cas où elle désignerait quelque chose de précis qui a une place centrale dans notre société. Dans EPD12, nous avons trouvé *talking-film* qui à l'époque devait tenir cette place; sa disparition des dictionnaires depuis montre qu'elle ne l'occupe plus. En revanche, nous étions surprise de ne pas trouver *bottle top*. Son absence des quatre dictionnaires ne s'explique que parce que son sens est prévisible; toutefois, son accentuation n'est pas du tout stable, comme nous le verrons dans la section suivante, l'accent primaire portant tantôt sur *bottle* tantôt sur *top*. Cette instabilité reflète peut-être le statut plutôt syntagmatique de cette unité et justifie son absence des dictionnaires.

Il y a plusieurs extraits qui ne sont pas lexicalisés parce qu'ils font référence à des idées ou à des inventions nouvelles : *vodka genie*, *climate change*, *mountain bike*, *genetically-modified crops*. Leur absence se justifie donc par leur nouveauté : si la nouveauté atteint un niveau d'institutionnalisation suffisant, peut-être l'unité figurera-t-elle dans une prochaine édition du dictionnaire Parmi les extraits que nous avons cités, il est peu probable que *vodka genie* atteindra ce statut; en revanche, les problèmes de l'environnement prennent une place de plus en plus grande dans notre société et il est tout à fait envisageable que *climate change* et *genetically-modified crops* soient lexicalisées; peut-être cette dernière le sera-t-elle sous une forme d'acronyme.

Nous avons déjà étudié l'évolution du lexique entre EPD12 paru en 1963 et les dictionnaires parus depuis 1990. Le processus de lexicalisation est aussi visible

si nous étudions les différences entre les deux éditions de LPD. L'édition 2000 inclut les unités suivantes, absentes de l'édition de 1990: *bad debt*, *underage drinking* cité seulement dans LPD00 ; *hard-earned, oilseed, razor blade*, qui sont cités par LPD00 et par EPD15. Il y a également des différences dans les choix d'entrées entre LPD et EPD: *dragon's blood* (qui est aussi dans EPD12), *greenhouse gas*, et *hard-earned cash* ne figurent que dans EPD15. Il est difficile d'expliquer tous ces choix: certains tiennent de l'institutionnalisation des unités qui est jugée plus ou moins grande selon le dictionnaire, ce qui justifie leur présence (ou absence): *bad debt* dans LPD00; *greenhouse gas* dans EPD15. D'autres semblent être données à titre d'exemple d'expressions courantes: *hard-earned cash* dans EPD15, *underage drinking* dans LPD00. Celles qui sont présentes dans EPD15 et LPD00 ont apparemment fait leur entrée dans le dictionnaire, mais selon quels critères? Il est possible que le critère de la fréquence soit utilisé mais il faut rappeler qu'en ce qui concerne les formations en Adv + *-en* comme *hard-earned*, ce sont des formations très productives et il ne serait pas possible de les inclure toutes.

Le même problème des critères de sélection est posé pour les extraits sonores qui ne correspondent pas à des entrées dans le dictionnaire alors que nous pouvions nous attendre à leur présence: *staff room* et *racing bike* dans la source littéraire, *distressed investing, publicity-shy* et *vulture fund* dans les sources documentaires. Le dernier extrait est sans doute du jargon, ce qui permet de justifier son exclusion. Sans doute est-ce la fréquence de *publicity-shy* qui nous fait penser que cette unité serait incluse dans le dictionnaire, mais l'éditeur doit

tenir compte non seulement de ce critère mais aussi du problème de productivité. Des expressions toutes faites qui sont souvent utilisées n'ont pas pour autant un statut lexicalisé[42]. L'unité *distressed investing* appartient à un domaine économique spécialisé (il ne figure pas dans *Chamber's* non plus) - nous rappelons que *bad debt* ne figure que dans LPD00 - donc les mots de spécialité peuvent être exclus[43]. Pour les deux derniers exemples, nous pourrions éventuellement évoquer la même explication. L'absence de *classroom* dans EPD12 a déjà été commentée; l'absence de *staff room* s'explique peut-être par le fait que l'unité appartient à une micro-société: celle de l'éducation. Nous ne sommes pas convaincue que son absence se justifie. L'unité *racing bike* soulève le même type de question. Il n'y a aucune entrée dont le premier élément serait *racing* ni dans LPD90 ni dans LPD00. Les entrées en *race(-)* ont toutes le sens de *race* désignant soit «course pour les chevaux» soit «race ethnique». EPD15 inclut l'entrée *racecar* (la graphie est en un seul mot) mais aucune entrée avec *racing*. Ces absences sont curieuses: si *racing* a une fonction adjectivale comme dans *racing pulse* («un pouls qui bat à pas de course»), son accentuation sera double, mais dans ces exemples, le mot n'a pas cette fonction et l'accentuation est simple. *Racing bike* et *racing car* (car ce terme existe) désignent tous les deux des engins destinés à faire la course (dans la nomenclature de Ginésy, il s'agit de leur «fonction»). Si nous considérons que ce sont des entrées institutionnalisées, il

---

[42] Voir Maniez, 2002, pour une discussion de la productivité de ce type de formation.

[43] Toutefois, nous notons avec intérêt le grand nombre de termes linguistiques qui figurent dans LPD, sans doute à l'intention des étudiants et des professeurs de la langue, c'est-à-dire des utilisateurs du dictionnaire. Il y a aussi beaucoup de termes relatifs au cricket.

faudrait les inclure dans le dictionnaire ; à l'inverse, si elles ne le sont pas, le fait que leur sens ne soit pas prévisible, et leur accentuation non plus, nous incite à les inclure de la même manière.

Sous cette rubrique, nous avons tenté d'expliquer les absences de certaines unités du dictionnaire mais nous en concluons que les critères de sélection des entrées ne sont pas toujours fiables. Dans la section suivante, nous allons étudier les prévisions de l'accent lexical des unités qui y figurent et évaluer les indications que donne le dictionnaire.

4.4.2 Les prévisions de l'accent et le comportement des unités en discours

Le tableau suivant nous permet de visualiser la corrélation entre les prévisions du schéma accentuel données par le dictionnaire et la réalisation effective de l'accent tel que nous le percevons. Nous distinguons les résultats pour les deux types de sources. Pour l'accent lexical, nous avons quatre indications possibles : « s » pour l'accentuation simple, « d » pour l'accentuation double, « s/d » s'il existe une variante, et enfin « d+SS » si l'entrée porte une accentuation double mais elle est susceptible de subir un déplacement de l'accent (*stress shift*) lorsqu'elle se trouve en position d'épithète. Pour la perception de l'accent, il n'y a que deux possibilités : « s » ou « d » selon le type d'accentuation perçue.

Figure 4.14: L'Accent prévu / perçu: extraits sonores figurant dans le dictionnaire

Type de source	Sources documentaires[44]		Source littéraire	
Type de schéma	Type de schéma prévu	Type de schéma perçu	Type de schéma prévu	Type de schéma perçu
«s»	16	23	6	8
«d»	11	11	1	2
«s/d»	1		3	
«d + SS»	6		0	

Nous analyserons d'abord les résultats en ce qui concerne la source littéraire. Nous constatons une très bonne corrélation entre les prévisions de l'accent et sa réalisation. Les trois entrées qui sont signalées avec une variation se réalisent avec un schéma «s» deux fois, et un schéma «d» une fois, ce qui correspond très exactement aux indications du dictionnaire. Nous notons que la variation sur l'extrait *armchair* est d'autant plus attendue que EPD, qui note aussi la variation, donne le schéma «d» comme prononciation principale et non «s» comme LPD.

Pour les sources documentaires, la corrélation entre les indications simples du dictionnaire («s» ou «d») est excellente: toutes les prévisions se réalisent. L'extrait *armchair* qui est sujet à variation se prononce avec le schéma «d». En revanche, l'analyse des résultats correspondant aux extraits qui sont marqués pour le déplacement de l'accent révèle deux anomalies. Il y a 6 entrées pouvant subir un déplacement accentuel, 5 d'entre elles se prononcent avec un schéma «s», donc avec déplacement, et 1 avec le schéma «d». En étudiant le contexte du

[44] En additionnant les chiffres correspondant à l'étude des extraits des sources documentaires, nous arrivons à un total de 34 pour 31 extraits en raison de la construction cumulative. En effet, pour trois extraits, il y a deux unités lexicalisées qui entrent dans la composition de l'ensemble.

discours de ces extraits, 5 d'entre eux sont dans une position d'épithète, 1 est isolé ; mais contrairement aux prévisions. c'est l'extrait isolé qui se prononce avec un schéma « s », donc l'accent se déplace sans que le contexte l'influence ; par ailleurs, un des extraits pour lesquels le contexte est propice au déplacement se prononce avec un schéma « d ». Toutefois, cette réalisation s'explique par le débit extrêmement délibéré du locuteur sur cet extrait.

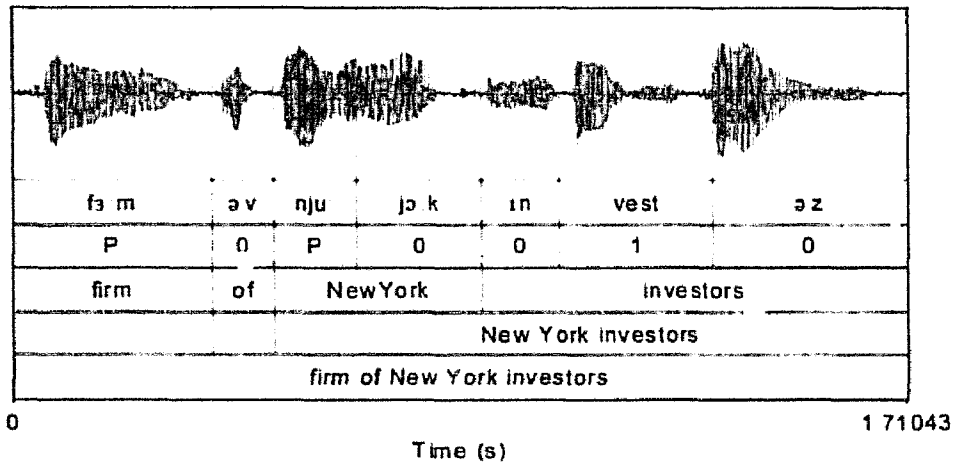
Ces résultats montrent à quel point le discours peut influencer la réalisation de l'accent. Les indications de variation et les indications du déplacement de l'accent se sont révélées efficaces, même si deux cas ne correspondent pas aux prévisions. S'il n'y avait pas eu d'indications, nous aurions eu 5 anomalies au lieu de 2 à cause du déplacement de l'accent, et 2 cas pour la variation. La prise en compte de ce genre de phénomène est donc souhaitable dans les indications de l'accent lexical.

Nous pouvons conclure que les indications de l'accent lexical sont très fiables. Dans la section suivante nous nous intéresserons à l'analyse acoustique de tous les extraits sonores : ceux que nous avons étudiés dans cette section et qui sont lexicalisés, puis tous les autres qui ne figurent pas dans le dictionnaire et pour lesquels nous ne disposons pas de prévisions de l'accent.

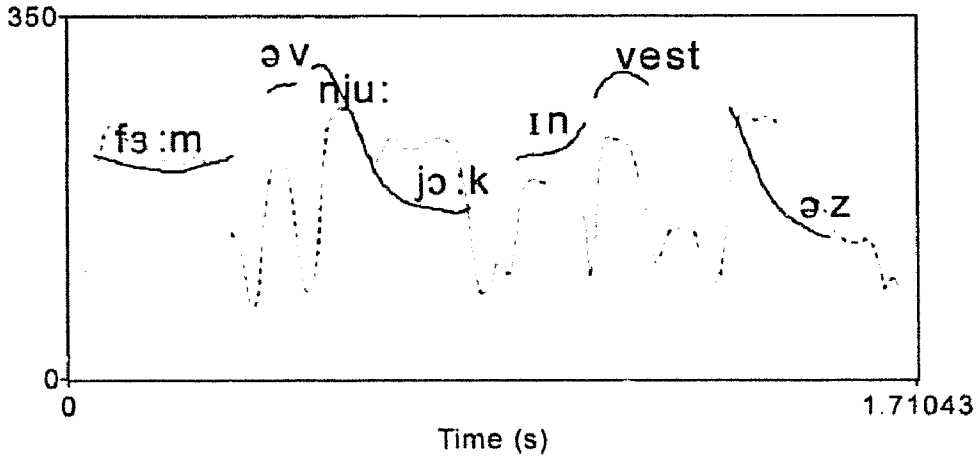
4.5 Analyse acoustique des extraits sonores de la base de données

Dans cette section, nous allons faire la synthèse des résultats de notre analyse des extraits sonores[45] qui s'est faite à partir des valeurs des corrélats acoustiques à l'aide des deux graphiques correspondant à chaque extrait: l'un montre les courbes du F_0 et de l'intensité, l'autre, l'oscillogramme avec la segmentation en syllabes et l'échelle temporelle, ce qui permet de rendre compte de la durée relative des syllabes et de l'amplitude du signal. A titre d'exemple, nous donnons les deux représentations graphiques qui correspondent à l'extrait « firm of New York investors ». La courbe du F_0 correspond à la ligne continue et celle de l'intensité à la ligne brisée.

Figure 4.15 : newyork\newyork1 : oscillogramme et courbes de F_0 et I



[45] Les extraits sonores se trouvent sur le CDROM dans le répertoire \corporal, dans le sous-répertoire désigné dans la légende de chaque figure. Un récapitulatif se trouve en Annexe E du chapitre 2 qui comprend le contexte de l'extrait (dans la colonne « Concordance »).



Nous avons adopté les conventions de la transcription phonétique large qui correspondent aux notations du dictionnaire LPD (1990), ainsi que la coupe syllabique proposée dans cet ouvrage. Le graphique des courbes de F_0 et I comportent cette notation qui se place au dessus de la courbe de F_0, syllabe par syllabe.

En ce qui concerne les mesures précises des valeurs du F_0, de l'intensité et de la durée, nous les avons prises pour tous les extraits provenant du reportage de *Reporters* sur les capuchons parlants (le répertoire est appelé « vodka »), ainsi que ceux de la discussion des essais des OGM, qui fait partie de la collection de reportages *Documentaires divers* (le répertoire est appelé « oilseed »). Certaines mesures ont également été prises pour d'autres extraits où les seuls graphiques ne

fournissent pas des indications suffisamment précises. Les tableaux récapitulatifs des mesures se trouvent en annexe C.[46]

Nous nous intéresserons d'abord aux corrélats acoustiques de l'accent, dans des contextes neutres et non neutres, et en fonction des différents types de contours intonatifs. Ensuite, nous rendrons compte de la qualité vocalique des éléments non accentués qui est citée, comme nous l'avons vu précédemment[47], comme un critère de la composition. Puis, nous étudierons l'accent par rapport à différentes formations: en les comparant à des non-composés d'abord, puis en fonction des types différents de composés. Enfin, nous analyserons l'influence de la position syntaxique du composé sur la réalisation et la perception de l'accent, c'est-à-dire le comportement des composés en discours, et nous tenterons d'en tirer des conclusions.

4.5.1 Le schéma accentuel et les corrélats acoustiques

4.5.1.1 Le rythme

Nous rappelons d'abord les caractéristiques rythmiques de l'anglais, avec ce qu'elles impliquent en ce qui concerne les unités rythmiques. Les expériences d'Allen (1972) visent à en étudier le fonctionnement et il arrive à la conclusion que l'anglais est bien une langue dont le rythme dépend de l'accent. Le type de battement rythmique auquel il fait référence est bien visible dans l'exemple que

---

[46] Des mesures précises ont également été relevées pour les extraits en « dark » dont nous avons rendu compte dans le chapitre 3, et qui se trouvent dans l'Annexe A de ce chapitre-là.

[47] Voir 1.1.1.1.1, « Qualité vocalique » et 2.1.1.1.4, « Le critère phonétique et phonologique ».

nous avons donné, au début de cette partie (Figure 4.15) « firm of New York investors ». Bien entendu, dans cet exemple, nous avons une proposition nominale qui comporte plusieurs mots dont une unité que nous pouvons considérer comme un composé (« New York », voire « New York investors »). Nous devons, cependant, situer nos analyses par rapport à ces caractéristiques générales, et en tenir compte dans nos analyses.

Dans le chapitre 3 (3.1.2), nous avons fait référence à deux types d'unités rythmiques : le pied et la syllabe. Dans l'exemple ci-dessus, il apparaît que la segmentation en pieds va aboutir à trois pieds de longueur semblable, le dernier étant plus long, ce qui est logique dans la mesure où il reçoit l'accent nucléaire :

firm of – New York in – vestors

Si *New York* forme un ensemble, cette unité n'est pas identifiée du fait de l'inclusion du *in* de *investors*, à l'intérieur du même pied. En revanche, ce découpage laisserait supposer que l'unité est bien l'ensemble : *New York investors*. Toutefois, si nous incluons l'article indéfini qui précède cette proposition, ***a firm of New York investors***, il ne fera pas partie du pied « firm of » : s'il est le début de la phrase phonologique, il sera extramétrique, sinon, il fera partie du pied précédent. Ce manque de cohérence sémantique peut gêner l'analyse du composé.

Si nous appliquons la syllabe comme unité rythmique, la segmentation est plus précise, et l'inégalité des syllabes est apparente. Il est vrai que cette segmentation ne permet pas de montrer des ensembles rythmiques, mais la

localisation des pics du F_0 et de l'intensité sont intéressants à analyser par rapport à la syllabe accentuée. Dans cet exemple, les deux pics du F_0 se trouvent respectivement sur *New* et sur *vest*, ce qui correspond aux accents forts dans le schéma accentuel de l'ensemble.

4.5.1.2 Les corrélats acoustiques de la perception de l'accent

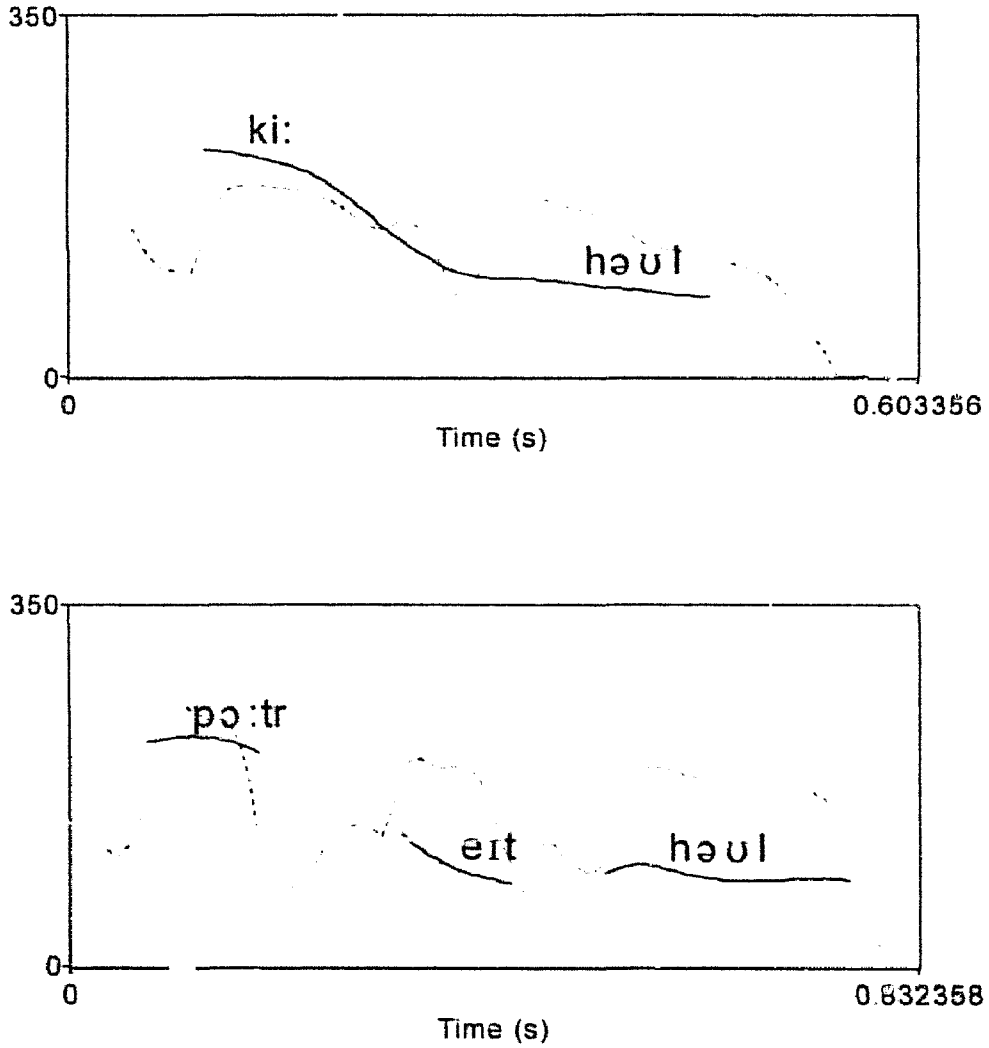
Les corrélats acoustiques de l'accent tels qu'ils ont été définis dans le chapitre précédent (3.1.1) sont: le F_0, l'intensité et la durée. Nous avons effectivement constaté que ces trois corrélats de la syllabe accentuée sont très souvent présents pour signaler son placement. Nous allons d'abord rendre compte des corrélations que nous avons trouvées dans des contextes intonatifs neutres. Puis, sous la rubrique suivante, nous étudierons plus en détail l'interprétation des corrélats dans des contextes non neutres, où le contour intonatif présente des caractéristiques particulières.

4.5.1.2.1 L'accentuation simple

D'après nos pré-analyses, à partir de l'échantillonnage des unités en « dark », nous nous attendions à trouver un niveau du F_0 plus élevé sur la syllabe accentuée, qui se conjugue très souvent avec un niveau élevé de l'intensité, et une durée plus longue par rapport aux syllabes inaccentuées. Dans nos analyses, nous avons effectivement constaté ces corrélations, qui sont plus marquées en ce qui concerne le F_0 que l'intensité. Nous avons choisi deux exemples pour illustrer cette corrélation: *keyhole*, qui est lexicalisé et qui désigne une ouverture (trou) qui est à l'intention d'une clé, et *portrait hole*, qui est un composé inventé dans le cadre du livre, *Harry Potter and the Philosopher's Stone*, et qui désigne une ouverture par

laquelle on peut passer, la présence de l'ouverture étant signalée par le portrait de la Grosse Dame qui en cache l'entrée. Nous notons au passage la différence des relations entre les deux éléments, mais le schéma accentuel est identique[48], comme le montre le graphique :

Figure 4.16 : hole\keyhole1 et hole\portrh1 : Courbes de F_0 et I



[48] Les deux extraits se trouvent en position nucléaire (voir la concordance en Annexe E du chapitre 2). La signification de la position syntaxique du composé est étudiée en 4.5.3.

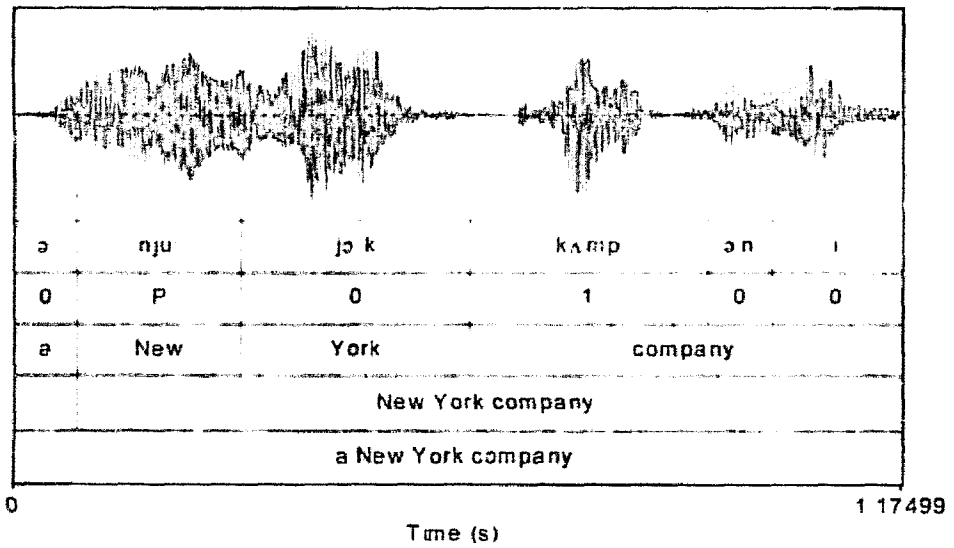
Nous constatons la bonne corrélation entre le F_0 élevé et la syllabe accentuée : le contour intonatif descend sur celle-ci. Nous avons rencontré ce type de courbe dans la moitié des cas d'accentuation simple de notre corpus. L'intensité est également plus élevée sur cette syllabe. Toutefois, la durée relative des syllabes indique plutôt une longueur supérieure sur la syllabe non accentuée *hole*, que sur la syllabe accentuée *key* ou *portr*. La corrélation entre une durée plus courte et une syllabe inaccentuée dont nous avons rendu compte dans le chapitre précédent concerne surtout des syllabes inaccentuées et réduites. Dans la mesure où le deuxième élément du composé est plutôt désaccentué qu'inaccentué, ce qui est indiqué par la non-réduction de sa voyelle, ce paramètre n'est pas très révélateur, du moins pour l'accentuation simple. Par contre, le paramètre a de l'importance pour l'accentuation double.

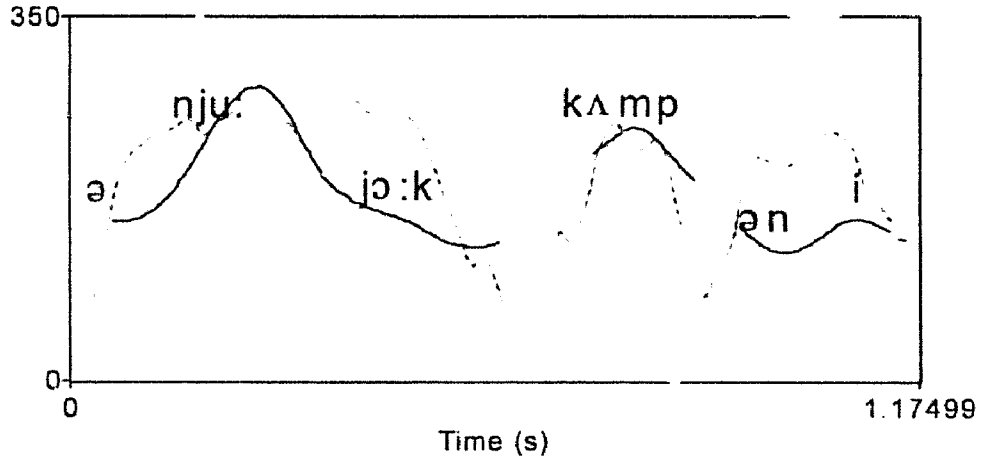
Nous rappelons que nos mesures tiennent compte de l'ensemble de la syllabe et pas seulement de la voyelle ce qui peut fausser les calculs au niveau de la durée relative. La syllabe inclut toutes les consonnes qui font partie de l'attaque, mais aussi de la coda. Comme nous avons adopté la coupe syllabique de LPD qui ne tient pas compte d'une éventuelle ambisyllabicité, les consonnes sont raccrochées à la syllabe accentuée précédente. Par ailleurs, certaines coupes ne sont pas complètement satisfaisantes : dans les cas déjà cités, il nous semble curieux que *portr-ait* soit segmenté ainsi, peut-être parce que le mot nous rappelle *portray* qui serait segmenté *por-tray* en raison de son schéma accentuel, mais aussi parce que «r» ne se prononce pas à la fin d'une syllabe en anglais britannique standard : or, ici il se prononce.

4.5.1.2.2 L'accentuation double

Les mêmes corrélats se trouvent réunis dans les cas d'accentuation double, mais comme il s'agit de l'accentuation du second élément du composé, il est souvent nécessaire de tenir compte du phénomène de la déclinaison du F_0 (voir 3.1.1.1.2.3. « Déclinaison de F_0 »). Dans l'exemple suivant, *a New York company*, nous constatons que l'accent nucléaire sur *company* ne correspond pas à un niveau du F_0 plus élevé, ni à une plus grande intensité. Nous donnons les deux graphiques :

Figure 4.17 : newyork\newyork3 : oscillogramme et courbes de F_0 et I





D'après le graphique de l'oscillogramme qui montre la coupe syllabique, nous devinons une durée légèrement plus longue sur *comp* que sur *York*, et bien plus que sur *New*. Est-ce suffisant pour que l'on perçoive l'accent nucléaire ? Sans doute que non. Le graphique des courbes du F_0 et de l'intensité montre que cet extrait comporte deux contours très similaires, dont le premier correspond à *New York* et le second à *company*, mais celui-ci est légèrement plus bas. Cette reprise de courbe est typique de l'accentuation double, comme nous verrons plus loin, car l'unité *New York company* comporte bien ce type d'accentuation. Sur le plan acoustique, le fait que cette reprise soit plus basse s'explique par la déclinaison de F_0. La baisse de l'intensité soulève à nouveau l'hypothèse que nous avons formulée dans le chapitre 3 (3.1.3.2.2.2, « L'intensité ») selon laquelle il existe également une tendance à une déclinaison de l'intensité.

Dans cet extrait, nous avons un composé *New York*, qui entre dans la construction d'une unité lexicale *New York company*. Le composé *New York* comporte ici une accentuation simple alors qu'isolément il comporte une accentuation double. Nous reviendrons à ces considérations dans la discussion du

déplacement de l'accent. En ce qui concerne les corrélats de l'accentuation double, nous notons l'importance de la durée: *York* dans cet exemple ne représente que 58% de l'unité *New York* alors qu'il en représente 76% dans l'unité à accentuation double.

Le cas de *kitchen sink*, qui est aujourd'hui un composé lexicalisé,[49] montre à nouveau la reprise de la courbe descendante pour l'accentuation double, mais cette fois-ci il y a une corrélation entre le F_0 et l'intensité aussi (voir Annexe D: vodka\kitchen1). La chute est surtout apparente sur le deuxième élément, peut-être du fait que le locuteur emploie un style de discours plutôt exclamatif et qu'il y a par conséquent une mise en relief nette de la syllabe accentuée *sink*. Nous reviendrons à ces considérations plus loin.

Enfin, nous signalons un dernier corrélat que nous n'avons pas relevé dans l'analyse des composés à accentuation simple de notre corpus: la pause. Dans un style de discours assez délibéré, nous remarquons la présence d'une pause entre les éléments d'un composé à accentuation double, qui renforce la perception de l'accent.[50] C'est le cas de l'exemple, *school owls*, dont la formation rappelle les terribles *school dinners*, ceux que nous mangions à l'école (voir Annexe D, broom\schowl1).

---

[49] Il figure dans LPD, éditions 1990 et 2000, avec le signalement de déplacement de l'accent, ainsi que dans EPD15, mais non dans EPD12. Voir Annexe C.

[50] Dujardin (2001) a constaté l'importance de la pause par rapport à l'emphase.

Dans les exemples précédents, nous avons illustré les occurrences d'accentuation simple et d'accentuation double dans des contextes neutres, où le contour intonatif est descendant (*fall* de Nicaise & Gray, 1998). Lorsque le contexte est non neutre, ou lorsque l'intonation présente un autre type de contour, il est nécessaire de tenir compte de ces caractéristiques différentes afin d'interpréter la courbe du F_0. C'est le sujet de la rubrique suivante: nous récapitulons d'abord les différents contours intonatifs que nous avons rencontrés, puis nous nous intéresserons aux contextes non neutres.

4.5.1.2.3 L'interprétation du contour intonatif

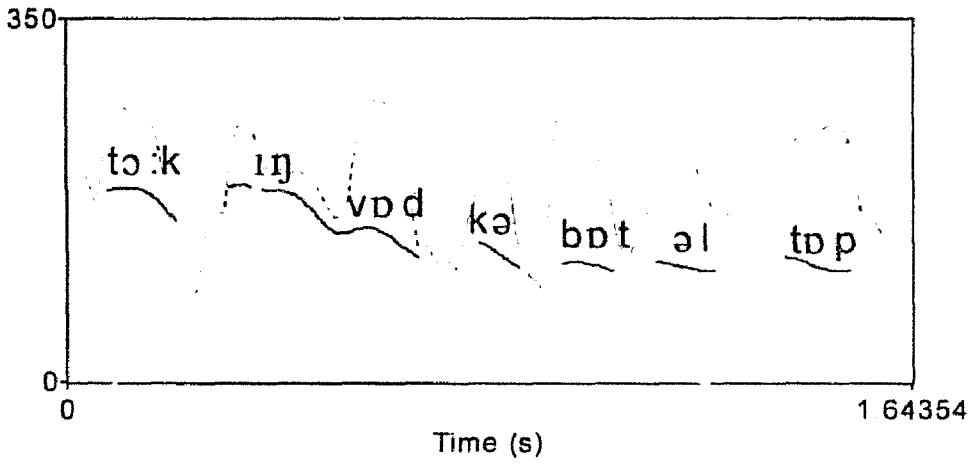
4.5.1.2.3.1 Les différents contours intonatifs

4.5.1.2.3.1.1 Intonation descendante: cas particuliers

L'intonation descendante peut présenter des caractéristiques particulières. Au lieu d'avoir une chute sur la syllabe nucléaire, avec des pics sur des syllabes accentuées avant celle-ci, nous constatons que la phrase phonologique tout entière descend par paliers. Dans ces cas, chaque pic est plus bas que le précédent, et la syllabe nucléaire s'identifie en étant sur le dernier. Ce contour s'appelle *downstep* (van den Berg, Gussenhoven & Rietveld in Docherty & Ladd 1992) et ne doit pas se confondre avec la déclinaison de F_0 qui est un phénomène perceptif qui implique que le dernier pic est perçu comme ayant la même hauteur mélodique que le précédent, alors que sa valeur absolue est plus basse. Dans le cas de *downstep*, l'interlocuteur perçoit la baisse graduelle de la valeur des pics. L'interprétation d'une courbe qui descend par paliers se fait donc en tenant

compte de la dernière chute qui correspond à la syllabe nucléaire. Dans la figure suivante, l'extrait *talking vodka bottle top* montre bien ce phénomène :

Figure 4.18 : vodka\talkbot1 : courbes de F_0 et I



En effet, il y a une légère baisse de F_0 tout au long de l'unité. En regardant les deux derniers éléments, *bottle top*, nous constatons que la hauteur de F_0 est à peu près égale (124 Hz et 123 Hz respectivement) et il y a une reprise de la courbe sur *top* qui reflète celle de *bottle*, ce qui est caractéristique de l'accentuation double. Ce type de contour a également des implications au niveau de l'interprétation des composés par rapport à leur position syntaxique : nous y reviendrons.

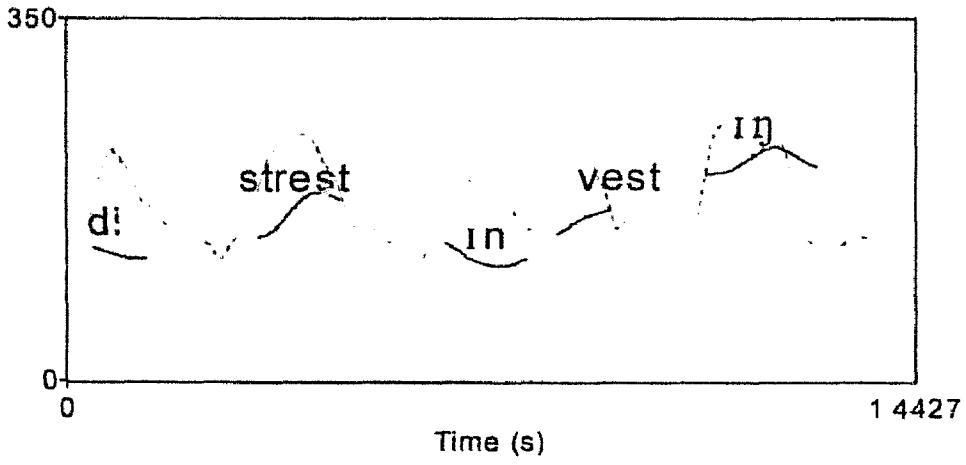
Parfois une intonation descendante s'accompagne d'une petite montée à la fin. Ce n'est pas forcément ce que Nicaise & Gray (1998) appellent *fall-rise* qui correspond à un contour où la chute est très nette et la montée est bien perçue comme faisant partie de l'ensemble. Dans une intonation de ce genre le message d'incertitude est souvent apparent. Dans les exemples que nous avons trouvés, la petite montée, qui se trouve sur la dernière syllabe inaccentuée (ou désaccentuée),

indique plutôt qu'il y a une suite : la phrase n'est pas finie. Dans le cas de *dragon liver*, par exemple, nous percevons une accentuation simple car la chute se fait sur *drag* et la petite montée sur la dernière syllabe de *liver* (voir Annexe D, dragon\dragliver1). Comme le montre la concordance (voir Annexe E du chapitre 2), ce composé est suivi dans le texte du signe de ponctuation deux points, ce qui montre bien la continuation de la phrase. Si ce composé avait été à accentuation double, il y aurait eu une chute sur la première syllabe accentuée du deuxième élément, *liver*, toujours suivie d'une petite montée sur la dernière syllabe.

4.5.1.2.3.1.2 Intonation montante

Ces caractéristiques de l'accentuation double sont illustrées par l'exemple, *distressed investing*, dont le contour est montant (*rise* de Nicaise & Gray, 1998).

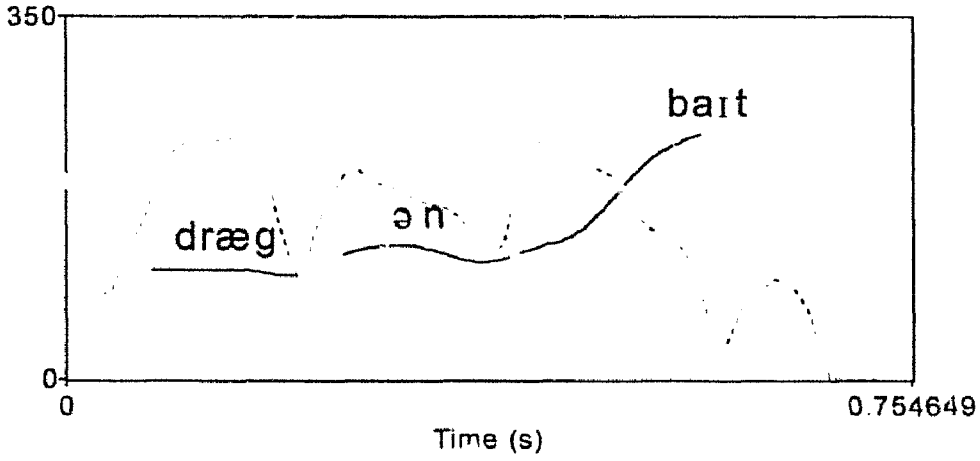
Figure 4.19 : debt\distinv1 : Courbes de F_0 et I



Dans ce cas, il y a une première montée sur *distressed* suivie d'une deuxième sur *investing* qui remonte encore plus haut. Nous avons déjà signalé dans la section précédente que la reprise de la courbe est une des caractéristiques de

l'accentuation double. C'est encore le cas ici, sur un contour montant. Nous voyons la différence entre les contours de ces deux types d'accentuation si nous faisons la comparaison avec *dragon bite*, qui est à accentuation simple, et qui est la dernière unité dans une question fermée (voir la concordance en Annexe E du chapitre 2).

Figure 4.20 : dragon\dragbite1 : Courbes de F_0 et I



Pour *distressed investing*, nous avons deux contours montants, alors que dans *dragon bite*, il n'y en a qu'un ; la montée commence dès le premier élément et se poursuit en continuité jusqu'à la fin.

Dans l'extrait *dragon dealer*, nous retrouvons les caractéristiques d'une accentuation simple dont le contour intonatif est montant, car ce composé fait partie d'un énoncé avec un « question tag » ; son contour intonatif est donc le reflet de celui du « tag ».

4.5.1.2.3.1.3 Intonation montante-descendante

Enfin, nous avons un cas d'un contour de type *rise-fall* (Nicaise & Gray, 1998) dans l'extrait *anyone anything*, où le second composé est mis en avant. Le mouvement mélodique commence sur la première syllabe de *anything* et se poursuit jusqu'à la fin du composé. Nous notons que le pic du F_0 est sur la syllabe inaccentuée et la chute est sur le deuxième élément. Nous insistons sur le fait que l'accentuation double aurait entraîné le même mouvement, mais à partir du deuxième élément *thing* seulement. Dans le cas d'un contour complexe comme celui-ci, où il y a une montée et une chute, il n'y aurait pas de double courbe comme celles que nous avons vues précédemment. Cet exemple est illustré plus loin sous la rubrique « insistance » (Figure 4.22).

4.5.1.2.3.2 Résumé des caractéristiques de l'accentuation simple et de l'accentuation double en fonction du contour intonatif

Pour conclure cette sous-section, nous résumons ce que nous avons expliqué en ce qui concerne les deux types d'accentuation en fonction des différents contours intonatifs. Il y a un mouvement mélodique (chute ou montée) sur l'élément accentué /1/ du composé, donc sur le premier élément dans le cas de l'accentuation simple et sur le deuxième dans le cas de l'accentuation double. Pour l'accentuation double, le graphique montre souvent que le même mouvement se fait sur les deux éléments (sauf dans des contours complexes) et il y a donc une reprise de la courbe. Cette caractéristique correspond au fait que les deux éléments sont accentués, mais le dernier est mis en relief.

Nous ferons référence à ces considérations sous la rubrique suivante où nous étudierons l'effet des contextes non neutres et leur influence sur la perception de l'accent et l'interprétation des courbes.

4.5.1.2.4 Les contextes non neutres

Des contours intonatifs différents entraînent un comportement des unités lexicales différent selon le contexte de discours; ce constat est tout aussi significatif lorsque le contexte est non neutre. Nous étudierons d'abord la mise en relief d'un élément, ce qui correspond au *focus*, concept fondamental en ce qui concerne l'intonation. Puis, nous étudierons des cas spéciaux de mise en relief: le contraste et l'insistance. Enfin, nous terminerons cette rubrique en montrant les effets acoustiques dans des contextes influencés par des facteurs psychologiques, et nous tirerons des conclusions sur les implications pédagogiques de nos travaux.

4.5.1.2.4.1 Focus

Comme nous le précisent les chercheurs qui s'intéressent à l'intonation, (Bolinger, 1972, Collier, 1992, Selkirk, 1995, Nicaise & Gray, 1998, Gray, 2001, Huart, 2002), la mise en relief d'un élément correspond à l'accent nucléaire; la non-mise en relief entraîne une désaccentuation[51]. Nous avons plusieurs cas dans notre corpus oral où les corrélats acoustiques subissent l'influence d'une mise en relief qui ne serait pas présente si l'unité était hors contexte.

---

[51] Voir la discussion de la désaccentuation et sa signification pour l'accentuation des composés dans ce chapitre, 4.2.2.2.1.

Cette mise en relief peut même entraîner un changement du schéma accentuel attendu. Ainsi, dans l'exemple, *emissions trading*, nous nous attendions à un schéma à accentuation simple en raison du caractère général du deuxième élément[52]; or, nous constatons que le schéma est à accentuation double (voir Annexe D, emiss\emiss1). Ceci s'explique par la mise en relief de *trading* par rapport à *emissions* qui a déjà été mentionné dans le reportage. Au niveau des corrélats acoustiques, cela se traduit par une double courbe descendante, une sur *emissions*, l'autre sur *trading*, et la seconde montre une chute plus nette, partant d'un pic du F_0 qui est la valeur maximale de cet extrait.

Nous retrouvons un phénomène semblable dans *total greenhouse gas emissions* (voir Annexe D, emiss\emiss2) où la courbe du F_0 indique la mise en relief de *total*; *greenhouse gas emissions* a déjà été mentionné. Cependant, dans cet exemple, si *total* présente une hauteur mélodique bien plus élevée que par la suite, il est clair que l'accent nucléaire porte sur l'une des syllabes du reste de la phrase; sinon, cet ensemble serait interprété comme étant un composé à accentuation simple. La complexité lexicale (quatre éléments) ainsi que la longueur (huit syllabes) rendent cette hypothèse invraisemblable. Nous y reviendrons dans la conclusion de cette partie.

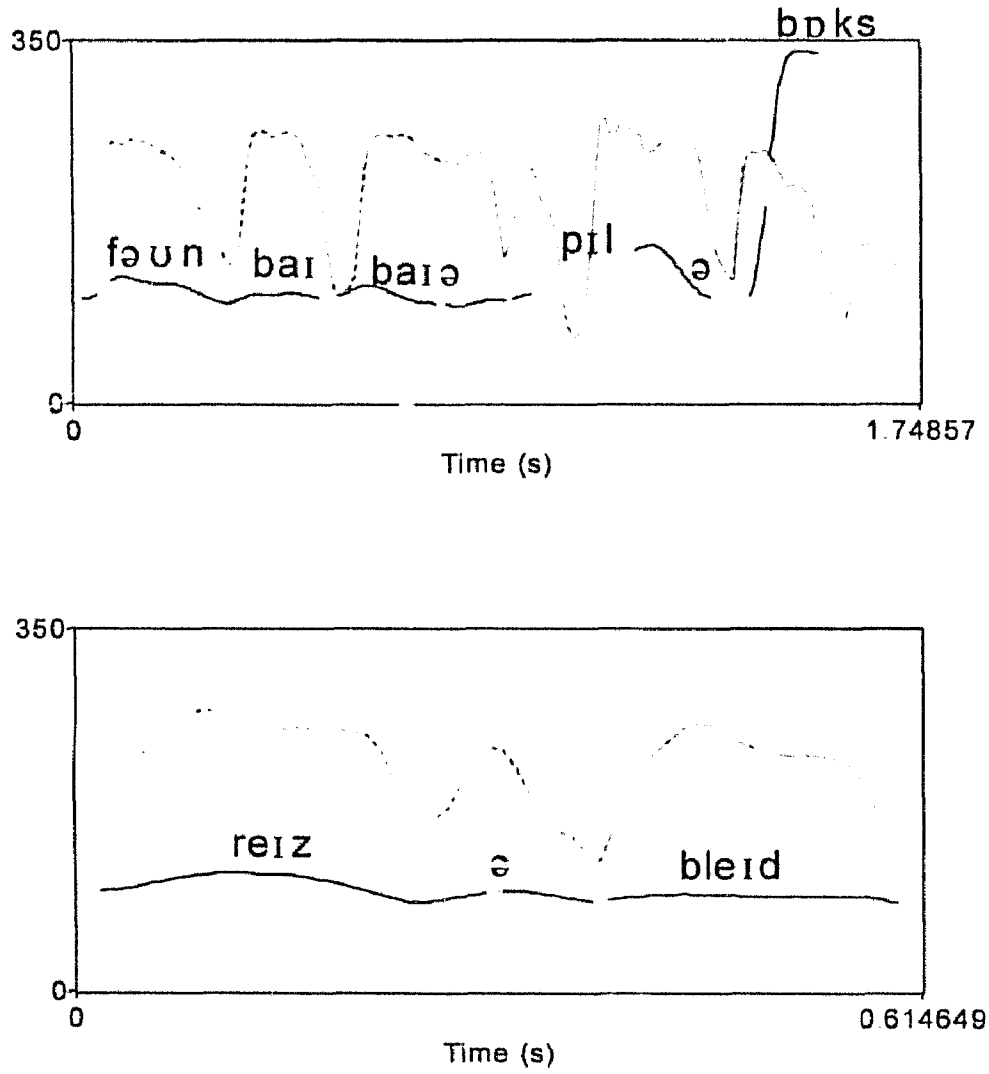
Enfin, nous citons un dernier cas de mise en relief qui est en fait une reprise. Dans le reportage qui traite du problème des publicités pour les cigarettes, le locuteur se trompe puis se reprend: [...] *phone by by a pillar box*. Le résultat de

---

[52] Voir les remarques de Boisson (1980) cité en 4.2.2.2.2.3

cette reprise est clairement démontré si l'on compare cet extrait à *razor blade*, dit par le même locuteur.

Figure 4.21 : test\pillarb1 et test\razorbl1 : Courbes de F_0 et I



Dans les deux cas, nous avons des composés à accentuation simple, mais la hauteur du F_0 sur *pil* est bien plus élevée qu'elle ne l'est sur *raz*. La courbe de *razor blade* est relativement plate mais montre quand même clairement une chute lente à partir du premier élément ; il y a également une baisse d'intensité. Pour *phone by by a pillar box*, de la même façon, la courbe commence avec une hauteur

mélodique assez basse, mais comme le locuteur se reprend, la hauteur sur *pil* est nettement plus élevée. Nous signalons au passage que la hauteur sur *box* est erronée[53].

4.5.1.2.4.2 Le contraste

L'accent de contraste est une des réalisations de ce qui a été appelé « accent de phrase » dans le premier chapitre (1.3, « Les différents types d'accent »). Il s'agit d'un phénomène de discours. Le contraste est une forme particulière de mise en relief qui consiste à faire ressortir une des unités lexicales qui s'oppose donc à une autre. Dans l'extrait, *spring-sown genetically-modified oilseed rape*, le locuteur insiste sur le mot *spring* qui s'oppose à *autumn-sown oilseed rape* et *genetically-modified autumn-sown oilseed rape*. Les deux derniers extraits ont une courbe avec des variations normales (le F_0 est d'environ 140 Hz au début de l'extrait, et 175 Hz sur la syllabe nucléaire *rape*) alors que la valeur du F_0 sur *spring* est très élevée (229 Hz sur *spring* et seulement 167 Hz sur la syllabe nucléaire *rape*). Il y a également un écart des valeurs de l'intensité considérable : 83 dB sur *spring*, alors que sur le début des deux extraits sans contraste, l'intensité est de 78 dB. Elle est de 81 dB pour tous sur la syllabe nucléaire. Nous constatons une nouvelle fois que le contraste a tendance à faire jouer les deux facteurs de l'intensité et du F_0.

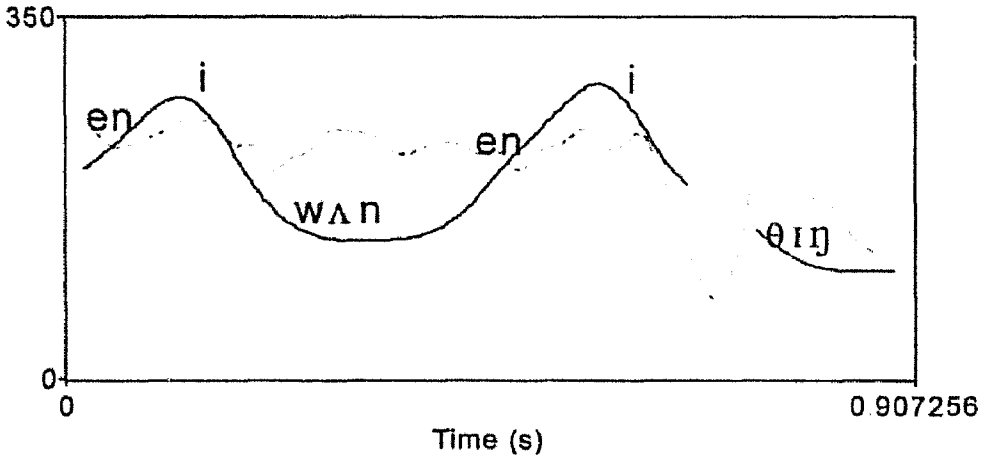
---

[53] Voir chapitre 3 et les problèmes de mesures: 3.1.1.1.2.2, « Comment interpréter la fréquence fondamentale : Paramètres contextuels ».. Ici, c'est la courbe brisée et la hauteur totalement en discontinuité qui signalent qu'il y a une erreur. En réalité, le locuteur parlait très bas.

4.5.1.2.4.3 L'insistance

L'accent d'insistance est une autre réalisation de l'accent de phrase, mais la mise en relief ne l'est pas par opposition à d'autres unités lexicales de la phrase ; elle signale l'importance capitale de l'unité en question. Nous reprendrons l'exemple cité plus haut à titre d'exemple pour le contour intonatif *rise-fall* : *anyone anything*. Dans cet extrait, le locuteur insiste sur le fait que les agissements de certaines compagnies d'investissement (*vulture funds*) ont empêché le Pérou de rembourser quoi que ce soit (à quiconque). Il insiste sur *anything* :

Figure 4.22 : test\anyOne1 : Courbe de F_0 et I



Les deux composes lexicalisés, *anyone* et *anything*, ont le même type d'accentuation (simple) et le même premier élément. Le graphique montre deux courbes parallèles, avec une légère pause entre les deux, ce qui contribue à l'effet d'insistance. La seconde courbe monte légèrement plus haut que la première, ce qui fait que la hauteur mélodique est perçue comme étant nettement plus haute, compte tenu de l'effet de la déclinaison normale de F_0. Nous constatons également

une chute plus importante sur *thing* que sur *one* qui peut s'expliquer comme étant un renforcement de l'effet d'insistance, ou bien parce qu'elle signale la fin de la phrase.

4.5.1.2.4.4 Les facteurs psychologiques

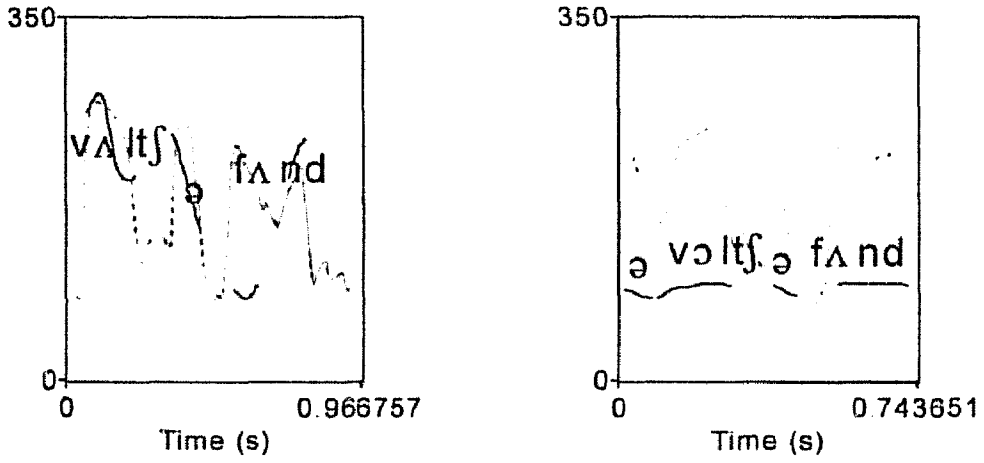
Nous savons que notre état psychologique a une influence sur notre façon de parler et ces différents styles de discours correspondent à des différences au niveau des corrélats acoustiques. L'émotion, par exemple, aura comme effet d'amplifier le F_0 et l'intensité. Nous prenons à titre d'exemple *portrait hole*, déjà cité sous la rubrique précédente pour illustrer l'accentuation simple. Nous avons donc un exemple de variation, due au contexte. Comme le montre les graphiques en annexe (voir Annexe E, hole\portrh3 et hole\portrh4), cet extrait comporte des valeurs du F_0 et de l'intensité bien plus élevées que dans un contexte neutre.

Pour nos analyses, ce genre de variation n'est pas gênante, mais nous avons aussi relevé à plusieurs reprises des extraits qui correspondent à des courbes extrêmement plates, au point qu'il n'est pas toujours possible de les interpréter. Lorsqu'il s'agit de locuteurs étrangers, nous avons écarté l'extrait en considérant qu'il n'est pas représentatif, mais lorsqu'il s'agit d'un locuteur natif, nous devons en tenir compte.

Dans le reportage sur les problèmes économiques du Pérou, l'unité lexicale *vulture fund* est énoncée à deux reprises par un locuteur britannique et à deux

reprises par un locuteur américain[54]. Il s'agit de journalistes ou d'analystes dans les deux cas. Nous avons constaté que la courbe du F_0 correspondant au locuteur américain est très basse. Les graphiques suivants permettent de comparer les deux réalisations. Le graphique de gauche est la première occurrence relevée de cet extrait (numéroté « 1 ») et correspond au locuteur britannique, celui de droite correspond au locuteur américain (et à la troisième occurrence) :

Figure 4.23 : debt\vultfun1 (locuteur GB) et debt\vultfun3 (locuteur US) : Courbes de F_0 et I



Nous faisons le même constat dans un autre reportage, mais nous ne pouvons pas tirer de conclusions précises à partir de si peu de cas. Toutefois, la tendance est à noter.

Enfin, nous trouvons des courbes semblables dans le reportage sur le satellite envoyé pour étudier le climat de Mars, mais cette fois-ci, il s'agit d'un locuteur britannique, professeur de physique atmosphérique. Nous émettons

[54] Le numéro à la fin du nom de l'extrait indique l'ordre de l'occurrence.

l'hypothèse que, lors de cet entretien, ce professeur ne fait aucun effort particulier pour transmettre son message, contrairement aux journalistes, ce qui aboutit à des énoncés avec ce style de contours intonatifs plats.

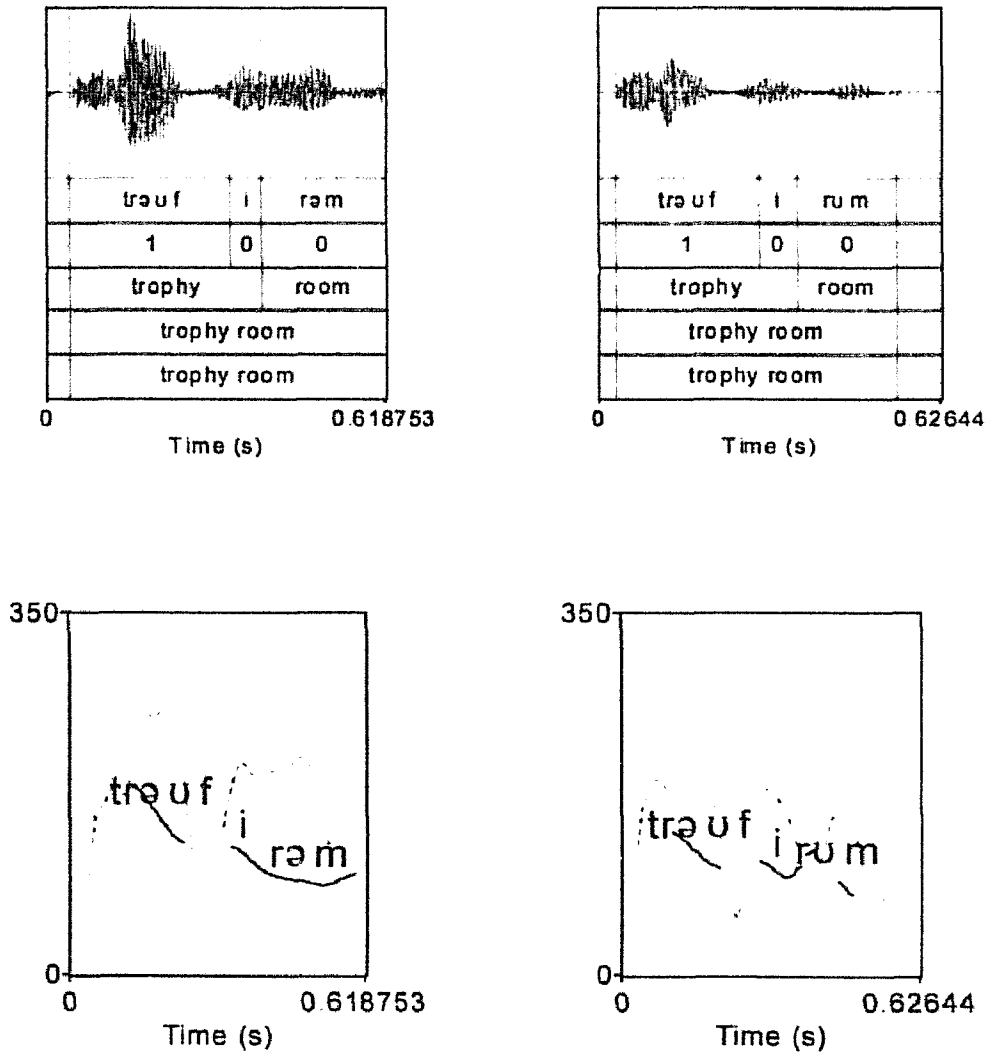
Ces constats nous font réfléchir au niveau des implications pédagogiques de nos travaux. Nous ne pouvons pas écarter les résultats qui contiennent des corrélats acoustiques peu probants, car il s'agit de discours de locuteurs natifs provenant d'enregistrements dont la qualité n'est pas mise en cause. Toutefois, nous faisons la distinction entre le style de discours que nous entendons effectivement dans une conversation, et celui que nous adopterons pour faire un discours en public, par exemple. Sans doute est-ce ce dernier qu'il faudrait enseigner.

4.5.2 Le schéma accentuel et la phonie

Dans le corpus oral, nous avons pu vérifier la règle de la non-réduction des éléments désaccentués dans les composés. En effet, nous n'avons que quelques cas qui méritent d'être étudiés. Le premier ne met absolument pas en cause la règle puisqu'il s'agit de la syllabe inaccentuée de l'élément accentué, *portrait*, dans *portrait hole*. Lors de la première occurrence de ce composé, la voyelle inaccentuée se prononce [eɪ] alors que dans les autres occurrences, cette voyelle se prononce avec un schwa. La longueur de la voyelle est sensiblement modifiée : 832 ms pour la voyelle pleine contre un temps compris entre 662 et 768 ms pour la voyelle réduite.

Le second cas, et le seul où on peut véritablement parler de réduction, concerne les mots en *room*. Le mot isolé peut se prononcer /ru:m/ ou /rum/ donc nous ne pouvons pas considérer que cette prononciation (qui est généralement utilisée dans les composés de notre corpus) correspond à une voyelle réduite. Cependant, dans au moins un cas, nous trouvons la voyelle schwa.

Figure 4.24 : room\trophy1 & room\trophy2 : Courbes de F_0 et I



Il est intéressant de noter que la courbe de l'intensité correspondant à la voyelle réduite est ininterrompue, et que l'oscillogramme ne laisse deviner que deux segments au lieu de trois.

Enfin, nous rencontrons un phénomène similaire dans les mots avec *broom*, où il peut y avoir une alternance /u:/ / /ʊ/. Nous remarquons avec intérêt que, dans les composés à accentuation simple, c'est non seulement sur l'élément désaccentué que nous trouvons *broom* avec la voyelle brève [ʊ], mais aussi en position accentuée, alors que nous pouvions attendre la voyelle longue [u:]. En fait, nous ne trouvons la voyelle longue que dans le composé à accentuation double : *school brooms*. Peut-être que la présence de la voyelle longue est aussi l'effet d'un écho de la voyelle de *school*, sans laquelle la proéminence relative de *broom* ne pourrait ressortir.

L'analyse acoustique des composés ne nous a pas permis d'apporter des informations complémentaires en ce qui concerne la signification de la phonie des segments. Nous nous sommes volontairement limitée à l'étude de la voyelle de l'élément désaccentué. Il est vrai que l'étude de la co-articulation entre les éléments pourrait également nous intéresser, mais cela dépasse le cadre strict d'une étude de l'accentuation, et nous souhaitons nous restreindre à ce propos, déjà vaste. Sous la rubrique suivante, nous passons à l'étude des composés par rapport à leur position syntaxique dans la phrase. Ainsi, nous abordons la question du comportement des composés en discours. Cette analyse devrait nous apporter des éclairages nouveaux.

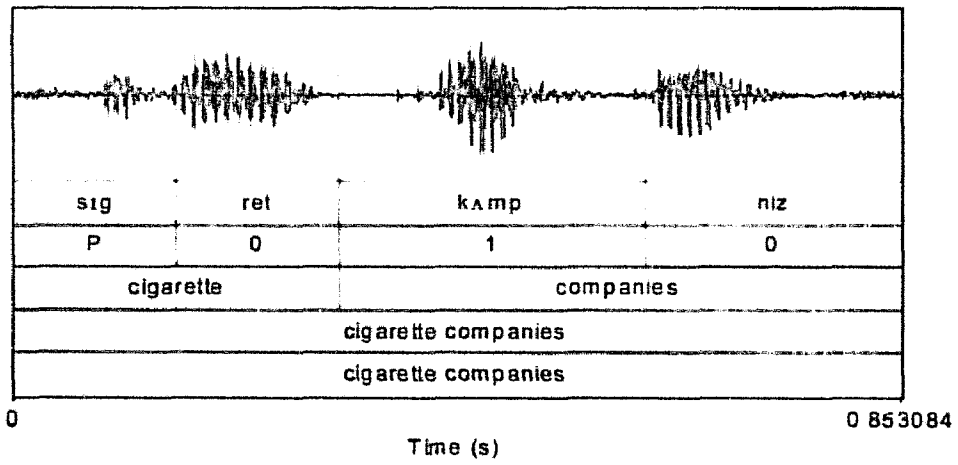
4.5.3 Le schéma accentuel et la position syntaxique des composés dans le discours

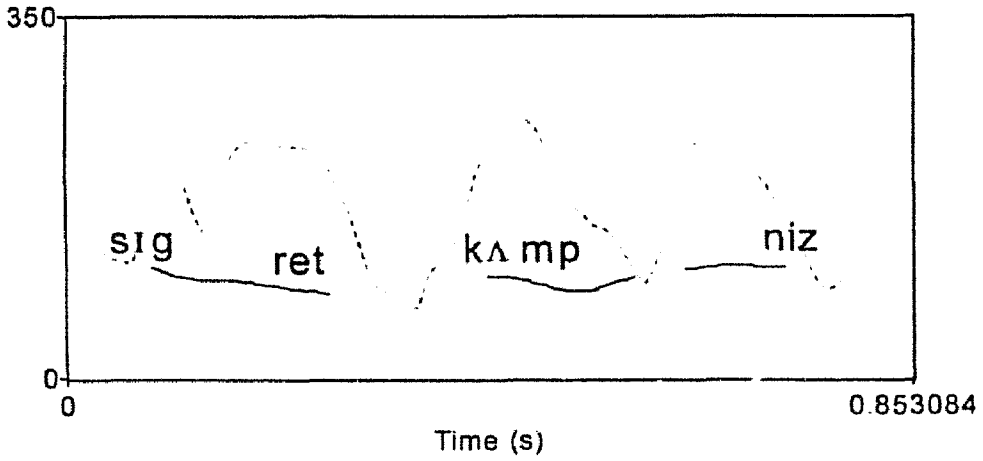
Nous allons étudier la portée de la position syntaxique du composé par rapport à son accentuation. Nous commençons cette étude par le phénomène très connu du déplacement de l'accent dont nous avons déjà discuté (en 1.6 et 4.2.3).

4.5.3.1 Le déplacement de l'accent

Ce phénomène du discours est suffisamment répandu pour qu'il soit répertorié de manière spécifique par les dictionnaires de prononciation: Wells indique un triangle ◄ après l'entrée, alors que EPD15 indique «*stress shift*». Le déplacement se fait sur les mots comportant un accent secondaire et un accent primaire, et sur les composés à accentuation double, surtout lorsqu'ils sont suivis d'un mot accentué. Une des explications qui est donnée est la raison rythmique: le locuteur évite deux accents adjacents. L'exemple suivant illustre bien ce phénomène.

Figure 4.25: shift\cigcomp1 : Oscillogramme et courbe de F_0 et I





En effet, le schéma accentuel indiqué par le dictionnaire est /201/ pour *cigarette*. Or, lorsqu'il apparaît dans l'unité lexicale *cigarette companies*, qui est une unité lexicale à accentuation double, nous constatons que le F_0 est plus élevé sur *cig* que sur *ette* et la chute de la courbe se fait à partir de cette syllabe, même si l'intensité est plus élevée sur *ette*. Nous remarquons également la compression aussi bien de la syllabe intermédiaire de *cigarette* que celle de *companies*. Le déplacement de l'accent de *ette* sur *cig* est évident.

En revanche, nous avons noté plusieurs cas où le déplacement de l'accent ne s'est pas fait. Dans *thirty-sixth floor*, par exemple, l'accent primaire reste sur *sixth*, et ne recule pas sur *thirt* (voir Annexe E, shift\thirty1). L'explication est simple: comme l'élément *floor* n'est pas accentué, il n'y a pas de raison rythmique de reculer l'accent principal de *ˌthirty-ˈsixth*.

Dans l'unité lexicale *hard-earned cash*, cité comme un exemple du déplacement de l'accent dans EPD15, les éléments restent tous bien séparés (voir Annexe E, shift\hardear1). L'explication repose sur le style de discours où le

commentateur pèse chaque mot afin de les faire ressortir comme des éléments individuels. Nous remarquons la présence d'une petite pause entre *hard* et *earned* ainsi qu'une intensité plus forte qui correspond à l'insistance dans cet extrait. La hauteur du F_0 est similaire pour les trois mots, ce qui donne cette impression d'éléments séparés.

Nous concluons que le déplacement effectif de l'accent est une possibilité mais non une obligation et qu'il dépend notamment du débit du discours. Le facteur rythmique n'est pas le seul qui joue un rôle. Sous la rubrique suivante, nous proposons une autre explication du phénomène que nous avons déjà mentionnée dans ce chapitre : *Phrase Edge Prominence* [55].

4.5.3.2 *Phrase Edge Prominence* et le déplacement de l'accent

Notre premier exemple est typique d'un contexte de déplacement de l'accent. L'adjectif composé non lexicalisé, *publicity-shy*, se trouve en position d'épithète devant le nom propre *Elliots*. En effet, nous constatons que l'accent est porté par la deuxième syllabe de *publicity* et non par *shy* comme il le serait sans doute en position d'adjectif attribut (voir Annexe E, shift\publici). La courbe montre bien les battements rythmiques de la langue. Cependant, la hauteur de ce premier élément est telle que la courbe donne l'impression qu'il est mis en relief ; or, nous ne le percevons pas ainsi. Une autre explication pour l'accentuation de cet élément est « *Phrase Edge Prominence* », qui est une hypothèse qui tient compte de ce que Selkirk appelle le « phrasing factor » (Goldsmith, 1995) selon

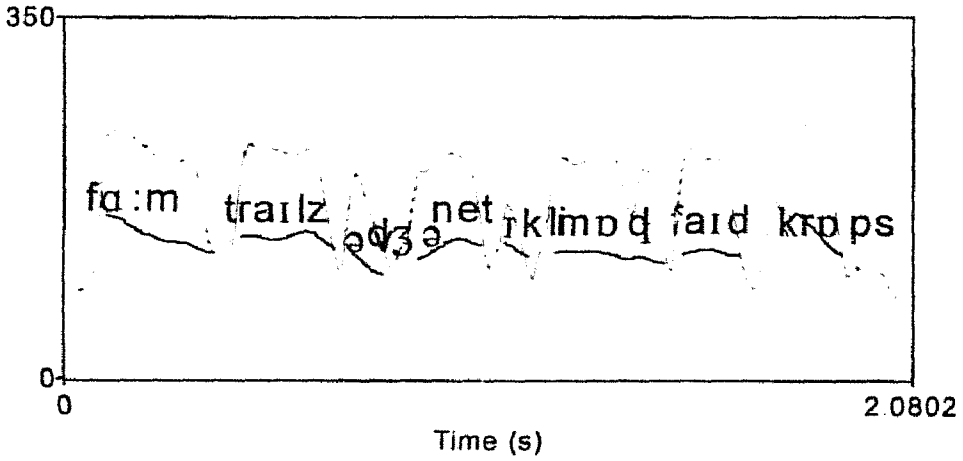
---

[55] Voir surtout les travaux de Selkirk (Goldsmith 1995).

lequel les éléments qui se trouvent aux extrémités de l'unité tonale ont tendance à être accentués. Ce point de vue peut rejoindre les propos de Boisson (1980) qui explique le fonctionnement mixte de la langue anglaise où l'accent lexical est à gauche, mais l'accent nucléaire de la phrase est à droite.

Selon cette hypothèse, la tendance à rendre le début de la phrase proéminente dépasse les contraintes rythmiques : il n'y a pas besoin d'un conflit accentuel pour que l'accent se déplace. Nous le voyons clairement dans des extraits plus longs, tels que *farm trials of genetically modified crops*, où la hauteur mélodique a effectivement deux pics majeurs : l'un sur *farm* et l'autre sur *crops*.

Figure 4.26 : oilseed\farmtri2 : Courbe de F_0 et I



Ce genre de courbe souligne le fait que les degrés d'accent n'ont rien d'absolu et qu'il faut étudier leur valeur dans le contexte du discours. Toutefois, l'importance relative des syllabes accentuées est claire et nous avons à nouveau des battements rythmiques bien distincts. Nous considérons que l'hypothèse de Selkirk peut être conciliée avec les contraintes rythmiques de la langue dans la

mesure où le fait d'aller au-delà des juxtapositions, comme elle le fait, ne remet pas en cause l'analyse locale des éléments. Cela permet de les relativiser en les remettant dans un contexte plus vaste.

Cependant, la tendance à accentuer les extrémités ne se réalise pas toujours. L'unité lexicale *bad debt* est relevée à deux reprises, avec un type d'accentuation différente à chaque fois. Le contexte nous aide à expliquer cette variation. Dans le syntagme *bad debt of a distressed country*, l'unité comporte une accentuation simple : les pics du F_0 sont sur *bad* et *countr*, c'est-à-dire sur les extrémités, ce qui soutient l'hypothèse de «*Phrase Edge Prominence*» (voir Annexe E, debt\baddebt1). En revanche, dans le syntagme *dead companies and bad debts*, qui pourrait également suivre cette tendance, les pics sont sur les noms *companies* et *debts* (voir Annexe E, debt\deadcom1). Sans doute que, dans ce cas, il y a une mise en relief des noms, ce qui explique l'accentuation double.

Sous la rubrique suivante, nous nous intéresserons à des unités qui apparaissent isolément et aussi dans une construction cumulative afin d'en étudier le comportement accentuel.

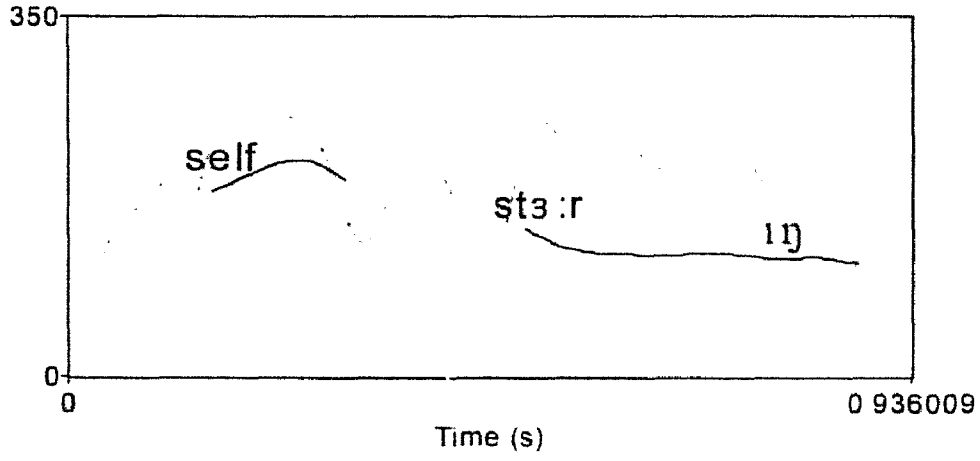
4.5.3.3 Les composés isolés et les composés dans une construction cumulative

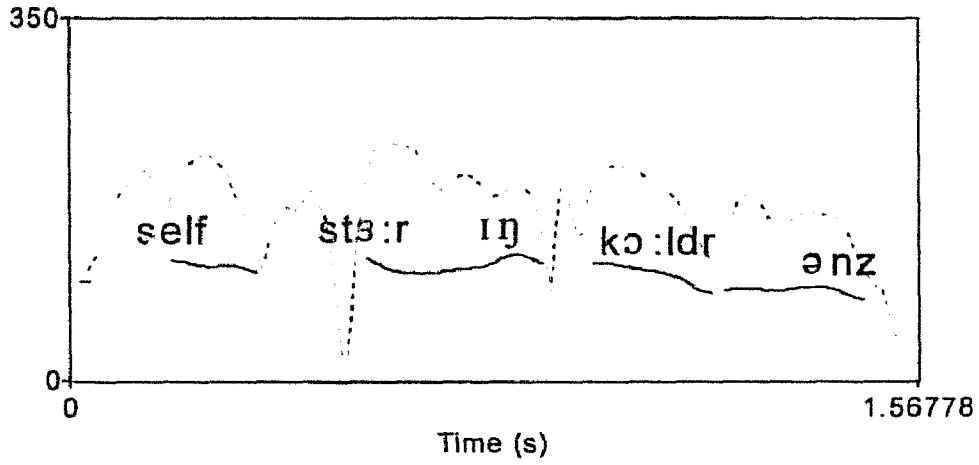
Dans les trois occurrences de *New York*, l'importance de la position syntaxique du composé apparaît clairement. Ce composé se trouve isolément, puis dans des constructions d'unités lexicales : *New York company* et *New York investors*, que nous avons déjà citées. Lorsque nous étudions le contexte de discours, nous nous apercevons que *New York* isolément fonctionne comme un complément de nom ce qui explique la courbe plate du F_0, car il n'y a aucune mise

en relief (voir Annexe, E, newyork\newyork2). Nous arrivons quand même à l'interpréter: l'unité comporte une accentuation double. Dans les deux autres cas, *New York* subit très nettement un déplacement de l'accent, comme le montre les courbes (en 4.5 et 4.5.1.2.2), ce qui est caractéristique de ce type de formation.

Les deux autres exemples de notre corpus sont: *under-age*, qui entre aussi dans la construction *under-age smoker*, *under-age smoking*; puis, *self-stirring*, qui sert aussi à former *self-stirring cauldrons*. Le bel exemple de *New York* n'est pas suivi. L'accent s'est déjà déplacé dans *under-age* isolément, qui comporte donc une accentuation simple, de même que dans *self-stirring*, isolément. De plus, l'accentuation simple de *self-stirring* apparaît plus nettement seul que dans la construction *self-stirring cauldrons*, dont la courbe du F_0 rappelle plutôt l'accentuation double, comme le montre les graphiques suivants.

Figure 4.27: shift\selfstir1 & shift\self2 : Courbes de F_0 et I





A partir de ces constats, nous ferons une première remarque. Nous voulons insister sur l'importance d'une prise en compte du contexte de discours et de la fonction du composé dans ce contexte. Dans le cas de *self-stirring*, l'accentuation simple peut s'expliquer par la mise en relief de l'élément *self*. Il s'agit d'une pancarte destinée à vendre des produits : le fait de pouvoir remuer le contenu de la bassine n'est pas surprenant ; ce qui est nouveau, c'est le fait que la bassine remue le contenu toute seule.

Dans la section suivante, nous revenons au phénomène de la variation en dehors du contexte syntaxique.

4.5.4 Le schéma accentuel et la variation

Sous la rubrique précédente, nous avons étudié des cas de variation d'une même unité dans des types de constructions différentes, ce qui nous a permis de tirer des conclusions sur l'importance du contexte de discours et de la position syntaxique du composé dans la construction. Nous nous intéresserons dans la

discussion qui suit à des extraits qui présentent les caractéristiques de variation accentuelle, sans que cette variation soit liée à une construction particulière.

Les trois extraits, *vodka genie*, tirés du reportage sur les capuchons parlants (c'est lui qui parle) comportent une accentuation simple pour deux des occurrences et une accentuation double pour l'autre. Il s'agit d'une unité non lexicalisée qui a été créée pour nommer la nouvelle invention technologique russe. Dans le second extrait (vodka\vodgeni2), l'accentuation simple pourrait s'expliquer par le déplacement de l'accent car il entre dans la construction *vodka genie's speech* (ou le *'s* correspond à un génitif déterminatif), ou pour des raisons rythmiques (l'accent nucléaire portant sur *speech*) ou en fonction de « *Phrase Edge Prominence* ». Toutefois, les deux autres occurrences correspondent à des contextes semblables, l'une étant en fin d'unité tonale *The voice belongs to the vodka genie (who...)*, l'autre se situant en fin de phrase. C'est cette dernière occurrence qui comporte une accentuation double ; de plus, elle est à la fin de l'enregistrement, ce qui ne nous permet pas d'adopter l'hypothèse selon laquelle l'accent se rétracte en fonction du degré de familiarité des unités. Nous ne pouvons que conclure à une variation libre, dictée par le choix du locuteur, peut-être aussi par son style de débit qui est relativement lent.

Notre second exemple, repris quatre fois au cours du reportage, est *vulture fund*. Nous en avons déjà analysé deux occurrences (voir Figure 4.23) afin de démontrer la variation entre locuteurs et la difficulté d'interpréter certaines courbes de F_0. Sur les quatre occurrences, les trois premières comportent une accentuation simple, la dernière une accentuation double. Ce quatrième extrait se

trouve en position de sujet, comme le troisième. Nous pouvons comparer le premier et le quatrième extrait qui proviennent du même locuteur et dont la qualité de la représentation graphique est bonne (voir Annexe E, debt\vultfun1 et debt\vultfun4). En ce qui concerne les corrélats acoustiques, la reprise de la courbe du F_0 sur *funds* ainsi que la hauteur plus élevée sur le second élément ne laisse aucun doute sur son schéma accentuel. Même le niveau de l'intensité est presque aussi élevé sur *funds* que sur *vult*. Il n'y a pas d'explication syntaxique, ni sémantique ; nous nous retrouvons face au choix du locuteur.

Enfin, nous terminons cette rubrique avec un exemple bien connu de composé lexicalisé qui est sujet à variation : *armchair*. Nous en avons quatre occurrences dont trois chez le même locuteur ; dans deux cas, la proéminence porte sur *arm*, dans les deux autres, sur *chair*. Les premiers graphiques montrent les extraits à accentuation simple : ils sont faciles à lire car l'intonation descend à chaque fois. Pour les extraits à accentuation double, qui figurent à la suite, il y a un cas où l'intonation monte, l'autre où il descend.

Figure 4.28 : vodka\armchai1 & vodka\HParmcha3 (accentuation double) : courbes de F_0 et I

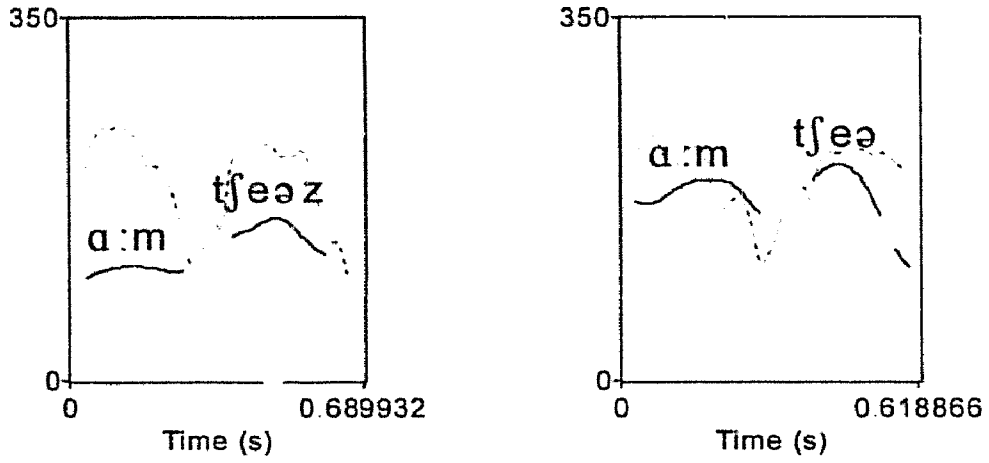
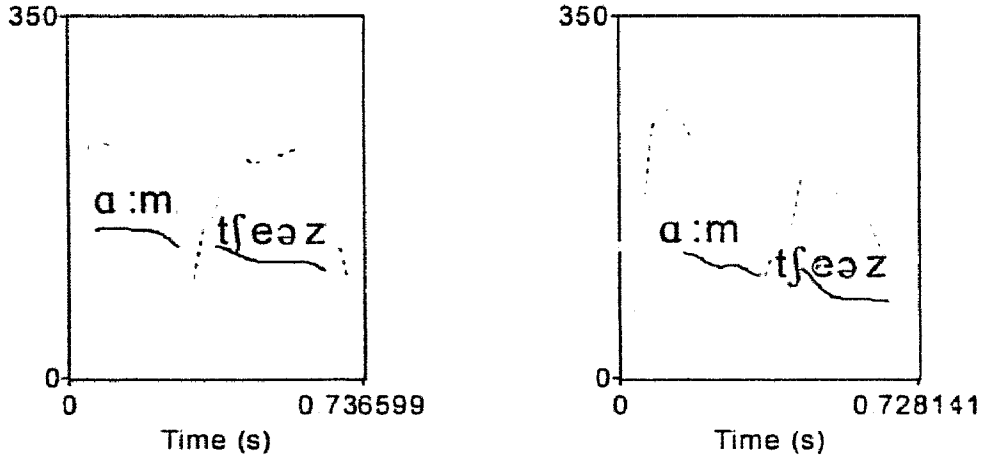


Figure 4.29 : vodka\HParmcha1 & vodka\HParmcha2 (accentuation simple) : Courbes de F_0 et I



En étudiant les mesures précises qui ont été relevées pour ces extraits (voir Annexe C : vodka), nous constatons que la valeur maximale du F_0 correspond bien à la syllabe accentuée /1/ dans tous les cas. La pente est également une indication intéressante (F_0MAS) si l'on tient compte également de la moyenne du F_0 (F_0Mean) et de l'écart type (F_0SD). Les trois mesures correspondent à des valeurs

plus grandes sur la syllabe accentuée (sauf dans un cas d'accentuation simple où ces calculs sont influencés par le niveau très bas de la hauteur mélodique qui atteint le seuil de 75 Hz). Il n'y a pas de reprise de courbe pour l'extrait avec une intonation montante, mais l'accent sur *chair* correspond à la courbe du F_0 très plate sur *arm*, suivie de l'importance de la montée sur *chair*. Ainsi, nous constatons que chez un même locuteur, nous pouvons trouver des variations accentuelles sans raison apparente; la variation ne semble donc pas toujours prévisible.

Toutefois, nous avons vu que le contexte du discours, sur le plan syntaxique et sémantique, peut expliquer un grand nombre de cas de variation. Pour conclure cette partie sur l'analyse acoustique des extraits sonores du corpus, nous émettons l'hypothèse que pour un certain nombre d'unités lexicales, nous avons le choix du type d'accentuation. Ce choix dépend des facteurs que nous avons déjà cités : le contexte sémantique, et notamment la mise en relief d'un élément; le contexte syntaxique et la position du composé dans la construction des unités et des phrases; enfin, le style de débit qui favorise l'accentuation double lorsqu'il est lent et délibéré, et l'accentuation simple lorsqu'il est plus rapide et fluide.

Il est évident que tous les composés ne sont pas sujets à cette variation, même s'il faut en tenir compte dans les conclusions que nous pourrons émettre sur les règles d'accentuation. Nous allons aborder la dernière partie de ce chapitre où nous tenterons de tirer des conclusions de l'analyse de l'accent lexical et de l'analyse acoustique des extraits sonores, par rapport aux règles d'accentuation des composés dont nous avons discuté au début de ce chapitre (section 4.5.2).

4.6 L'analyse acoustique, l'analyse de l'accent lexical et les règles d'accentuation des composés

L'accent lexical est indiqué par les dictionnaires en tant que prévision du schéma accentuel ; l'analyse acoustique a permis d'établir le type de schéma pour chacun des extraits sonores du corpus oral. Dans cette dernière partie, nous ferons la synthèse de notre analyse de l'accent lexical et de l'analyse acoustique par rapport aux règles de l'accentuation des composés afin de pouvoir faire des prévisions de l'accent, notamment en ce qui concerne les néologismes.

Nous commencerons par une étude des cas où il existe deux types d'accentuation correspondant à des interprétations sémantiques différentes. Ces entrées ont été relevées à partir de la vérification de N-COMPACT et des entrées doubles de LPD-Cp et N-COMPACT. De plus, nous avions marqué au fur et à mesure de la saisie de N-COMPACT certaines entrées qui auraient pu être associées à des types de schémas différents.[56] Cette étude nous permettra de situer les principes généraux de notre analyse des règles.

Ensuite, dans la deuxième section, nous ferons une étude d'une partie de la base de données : les entrées de N-COMPACT dont le premier élément commence par une majuscule. Cette étude a deux objectifs : premièrement, voir le nombre d'entrées de cette table dont il ne faut pas tenir compte en appliquant les critères de sélection que nous avons utilisés pour LPD-Cp. Il y a un certain nombre de

---

[56] Nous les avons extraites de la table à l'aide de la requête **RECH-COMM-NC** (voir Annexe A).

noms propres, par exemple. Nous pourrons ainsi évaluer la portée des statistiques que nous avons données en 4.3, et faire les ajustements appropriés. Deuxièmement, pour les nombreuses entrées correspondant à nos critères de sélection, nous nous servirons de cette mini-base de données afin de tester les règles proposées pour l'accentuation simple et l'accentuation double, en faisant des rapprochements avec le corpus oral où cela est possible. Nous étudierons dans cette section les composés qui désignent des noms de marque, car cette mini-base en comporte forcément, mais il y en a aussi un nombre considérable dans LPD-Cp, indiqué par le code «tdmk» dans la catégorie syntaxique, comme nous l'avons fait remarquer en 4.3.2.4.

Enfin, dans la troisième section, nous étudierons les composés des dictionnaires de la base de données qui comportent un des éléments entrant dans la composition des extraits sonores. Cela nous permet d'établir une relation entre les sources écrites et les sources orales, et de comparer les prévisions de l'accent lexical avec le comportement des extraits sonores. Nous pourrons ainsi tirer des conclusions sur la prévision de l'accent des extraits sonores non lexicalisés. A partir de cet ensemble, nous étudierons les composés selon leur catégorie en nous servant de celles que nous avons proposées dans la partie 4.2 de ce chapitre : les composés endocentriques (radicaux), les co-ordonnés (*dvandvas*), les exocentriques (*bahuvrihis*), les synthétiques (verbaux), puis les composés réduplicatifs et ceux ayant une apostrophe. Enfin, nous ferons une étude des composés qui comportent plus de deux éléments ainsi que des unités ayant une construction cumulative.

4.6.1 Etude des entrées ayant deux types d'accentuation

Dans cette section, nous nous intéressons aux cas où les différences accentuelles correspondent à une forme de désambiguïsation afin de déterminer quelles sont les règles d'accentuation qui permettent d'expliquer les différences. Notre étude porte sur les 9 cas de doublons dans N-COMPACT, les entrées doubles de LPD-Cp et N-COMPACT, et les entrées de N-COMPACT pouvant être associées à des types de schémas différents (relevées à la saisie). Nous étudierons d'abord les doublons de N-COMPACT.

L'étude de ces cas démontre l'efficacité des règles, mais aussi la nécessité de prendre en compte plusieurs facteurs. Sur ces 9 cas, 3 s'expliquent par l'opposition que fait Ginésy[57] entre « nature » pour l'accentuation double et « fonction » pour l'accentuation simple : *funny man, toy factory, paper bag* ont un schéma « d » lorsque le premier élément fait référence à la nature du second (l'homme est drôle, l'usine est un jouet, le sac est fait de papier) ; ils ont un schéma « s » lorsqu'il s'agit de la fonction du second (le métier de l'homme est de faire rire, l usine fabrique des jouets, le sac est pour les journaux).

D'autres s'expliquent par le contraste implicite dont parle Boisson[58] : *family name, old boy, old school, party line* illustrent cette idée. Si l'entrée porte une

---

[57] Ginésy, 1995. Voir ce chapitre : 4.2.2.2.2.4 : « L'accentuation : reflet d'une relation de fonction / nature ».

[58] Voir ce chapitre 4.2.2.2.3 : « L'accentuation : reflet d'une relation sémantique entre les éléments ».

accentuation double, alors il s'agit d'une relation simple (voir aussi Huart[59]) où le premier élément décrit le second (on peut encore parler de sa nature); pour l'accentuation simple, le premier élément reflète un contraste implicite ce qui implique une mise en relief (*focus*) de l'élément accentué. Cette accentuation est souvent associée à une forme de sous-catégorisation. Pour *family name*, le schéma « d » correspond au sens de « réputation de la famille » (« name of the family »). Nous constatons à nouveau que la relation « of » désignant l'appartenance est associée à ce type de schéma. L'accentuation simple correspond au sens de « nom utilisé afin de désigner la famille », par opposition à *first name* ou *nickname*. Pour les deux entrées en *old*, l'accentuation double correspond au schéma du syntagme où *old* est un simple adjectif qualificatif alors que l'accentuation simple sous-entend d'autres relations plus complexes : l'école n'est pas vieille, c'est celle que l'on fréquentait autrefois. Il y a une opposition implicite avec *new boy* (élève qui vient d'arriver dans l'établissement) par exemple. Enfin, pour *party line*, le schéma « d » indique le sens « line of the party » donc dans un contexte politique, alors qu'avec le schéma « s », le composé appartient au contexte des télécommunications et s'oppose à d'autres types de lignes téléphoniques (*land line*, par exemple, qui n'est pas encore lexicalisé, et qui désigne les lignes fixes par opposition aux mobiles).

---

[59] Huart (1989, 2002). Voir ce chapitre 4.2.2.2.5 « L'accentuation : reflet d'une relation préconstruite (ou non) entre les éléments ».

Le cas de *lightning strike* s'apparente aux deux explications déjà données : le schéma « d » correspond à une relation où le premier élément décrit la nature du second et il s'agit d'une grève qui s'est déclarée très vite, une grève eclair. Le schéma « s » correspond plutôt à la deuxième explication selon laquelle il y a une mise en relief du premier élément et une sous-catégorisation : *lightning strike* s'apparente à *thunderbolt*, tous les deux ayant une accentuation simple.

Enfin, pour expliquer *hard core*, il faut s'inspirer d'Huart (1989 & 2002) qui démontre que si le premier élément est la caractéristique principale que nous voulons mettre en valeur, l'ensemble aura une accentuation double ; la relation entre les éléments est simple : dans ce cas, cette unité lexicale désignera un noyau. En revanche, si le premier élément correspond à une des caractéristiques (celle qui a été choisie pour représenter le tout, comme pour *blackbird* qui désigne un type d'oiseau dont la couleur de la femelle est marron d'ailleurs), elle porte une accentuation simple et *hard core* désigne alors de la pierre concassée.

Si nous étudions les 76 entrées communes de LPD-Cp et de N-COMPACT, nous constatons que certaines ont des schémas différents (10 cas nouveaux et 1 cas, *hard-core*, qui a été expliqué). Il y en a 7 qui correspondent à des différences de catégorie syntaxique, les 3 autres à des différences de sens. Nous citons un cas de catégorie différente qui nous intéresse par rapport à *bottle top* qui fait partie du corpus oral : *screw-top/screw top*. Dans LPD-Cp, il s'agit d'un adjectif à accentuation double susceptible de subir un déplacement de l'accent ; dans N-COMPACT, c'est certainement un nom. La différence d'accentuation correspond à ces fonctions syntaxiques et le sens est : un capuchon que l'on visse par

opposition à des capuchons avec d'autres types de fermeture. Les relations entre les éléments relèvent bien du schéma «s» correspondant à la «fonction» du premier élément par rapport au second. Pour *bottle top*, nous nous attendions à ce même schéma car il s'agit de la fonction du capuchon qui est de fermer les bouteilles. Nous étudierons ce cas à la fin de cette section avec les unités à construction cumulative.

Passons aux cas où il y a une différence de sens: *free hand, second hand, white water*. L'opposition «nature/fonction» explique *second hand* où le schéma «d» correspond au sens «d'occasion», «de deuxième main». Pour le schéma «s», *second hand* désigne l'aiguille d'une montre dont la fonction est de montrer les secondes. Notre deuxième explication, la mise en relief du premier élément (et / ou le contraste implicite) est celle qui convient pour comprendre le schéma «s» de *whitewater* et *freehand*. Avec un schéma «s» le premier désigne l'eau du bord de mer où les vagues se brisent, ou encore l'eau des rapides; le second est un adjectif qui s'applique aux réalisations non guidées. Quand ces unités ont un schéma «d» nous retrouvons alors l'interprétation linéaire de l'adjectif suivi du nom. Dans le cas de *white water*, il s'agit sans doute d'un terme de chimie; pour *free hand*, il s'agit de liberté d'action.

L'étude des entrées de N-COMPACT que nous avons relevées lors de la saisie nous permet de confirmer les explications que nous avons proposées. Parmi les 156 entrées à commenter, certaines ont déjà été traitées dans les paragraphes précédents et il existe aussi quelques noms propres que nous écartons de notre

analyse. Il subsiste 111 entrées[60] qui méritent d'être étudiées. Il y a 58 entrées ayant un schéma « s » et 53 ayant un schéma « d ».

Les explications pour les entrées à accentuation double s'appuient surtout sur l'opposition « fonction/nature » : 36 d'entre elles correspondent à des cas où le premier élément indique la nature du second. Cependant, ces indications ne sont pas toutes du même type : trois catégories se distinguent. La première catégorie correspond aux cas où le premier élément est en *-ing*, comme *helping hand* : l'ensemble désigne la main qui aide. Une deuxième catégorie comporte les entrées où le premier élément désigne la matière qui constitue l'ensemble, comme *pith helmet* (un casque de paille). Ces exemples sont intéressants car il s'agit d'une forme de sous-catégorisation ; or, nous l'avons vu précédemment, nous pouvons souvent rapprocher la sous-catégorisation du contraste implicite et donc de l'accentuation simple. Laquelle des deux règles doit s'appliquer ? Nous reviendrons à cette question dans la section suivante. La troisième catégorie correspond aux simples adjectifs qualificatifs suivis de noms, comme *sharp practice*. Si le premier élément ne fait que décrire le second (la relation entre eux est simple, l'adjectif est la caractéristique unique destinée à décrire le second élément, comme pour les syntagmes), nous aurons un schéma « d ».

---

[60] Voir la liste de ces entrées en Annexe F.

En ce qui concerne les autres entrées en « d », la plupart s'expliquent par la localisation[61]. Si le premier élément sert à localiser le second, comme pour *desert rat*, alors le schéma est « d ». Nous retrouvons le même problème pour la sous-catégorisation que pour les entrées où le premier élément désigne la matière. Il s'agit ici d'une sorte de rat : or, *water rat* aura une accentuation simple car le premier élément comporte un contraste implicite, alors que *desert rat* aura une accentuation double car la relation que nous établissons est celle de localisation.

Il subsiste quelques cas qui sont plus difficiles à expliquer : *liege lord, harvest home, horse chestnut, cotton candy, seller's market, stormy petrel, vegetable marrow*. Deux de ces entrées peuvent être considérées comme des composés co-ordonnés où la relation entre les éléments n'a pas d'ordre hiérarchique ce qui se reflète dans l'accentuation : *liege lord* et *vegetable marrow* peuvent s'expliquer ainsi. Le second composé montre la complexité des règles car nous pourrions très bien l'opposer à *bone marrow* qui s'accentue avec un schéma « s » car il y a un contraste implicite. Toutefois, il n'y a pas de sous-catégorisation dans ce cas car l'élément *marrow* désigne des référents très différents, comme nous nous en apercevons si nous les traduisons : *vegetable marrow* désigne « la courge », *bone marrow* « la moelle osseuse ». En ce qui concerne *liege lord*, nous avons aussi rencontré *liege man* avec un schéma « s », mais ce n'est pas un

---

[61] Voir Huart (1989 & 2002) et ce chapitre : 4.2.2.2.5 « L'accentuation : reflet d'une relation préconstruite (ou non) entre les éléments »

composé co-ordonné et ce type d'accentuation est celui que nous attendons pour les composés en *-man* auxquels nous avons déjà fait référence [62].

La localisation pourrait expliquer *harvest home*, car il s'agit d'une célébration de la rentrée de la moisson. La localisation est temporelle et non spatiale dans ce cas. Pour *cotton candy*, le premier élément désigne non la matière du second mais la ressemblance au coton. Pour *seller's market*, la relation entre les éléments est simple ; nous retrouvons la relation avec « of » dont nous avons parlé. Enfin, pour les deux derniers cas, *horse chestnut* et *stormy petrel*, le premier élément donne une précision par rapport au second, mais il n'y a pas de sous-catégorisation. Si l'on peut utiliser *petrel* ou *chestnut* seul, leur appellation pleine comprend aussi ce premier élément. Peut-être que la mise en relief reste sur le second élément pour cette raison.

Pour les composés à accentuation simple, les entrées correspondent à trois explications différentes, certaines pouvant relever de deux d'entre elles. Une quinzaine de composés comporte l'idée de « fonction » par opposition à « nature ». Ainsi, un composé ayant un premier élément en *-ing* comme *'aughing gas*, désigne un gaz qui fait rire (et non « un gaz qui rit », ce qui serait l'interprétation s'il avait une accentuation double, comme *helping hand*). De même, *front man*, s'explique par le fait que le rôle de l'homme est d'être devant (ce n'est pas sa localisation). Cet exemple peut aussi s'expliquer d'une autre manière : la généralité du second élément. Cette deuxième explication s'applique à une dizaine

---

[62] Voir 4.3.2.3 « Etude de la phonie ».

de composés tels que *solar system* ou *women's movement*. Nous remarquons qu'il y a toujours une mise en relief de l'élément spécifique par l'accentuation simple qui peut aussi fournir une explication pour ce type de schéma. Enfin, la troisième explication repose sur la mise en relief du premier élément souvent dans un contexte de contraste implicite et donc de sous-catégorisation : elle s'applique à une quinzaine de cas. Les composés en *high* (*high jump*, *high road*) fournissent des exemples clairs. Le composé *high jump* est une sorte de saut (comme *long jump* ou *triple jump*) et non pas la description d'un saut particulier qui était haut (qui serait un syntagme avec un schéma « d ») ; *high road* est une sorte de route avec une fonction particulière ; il ne s'agit pas de sa description (sa nature). Nous revenons donc à notre première explication qui est l'opposition « nature/fonction ».

Il subsiste une dizaine d'entrées pour lesquelles nous avons de la difficulté à catégoriser les explications. Certaines correspondent aux cas où le premier élément désigne une caractéristique choisie pour représenter l'ensemble : *hot rod*, *sausage dog*, *green belt*. Il s'agit de ressemblances dans les trois cas : le premier composé est une automobile qui a été convertie afin d'obtenir des performances maximales, ce qui correspond à l'idée de *hot* en langage familier ; en enlevant tout ce qui est superflu, d'où l'élément *rod*, sans doute. Le second est un chien qui ressemble de par sa forme à une saucisse ; le troisième est la ceinture, dite verte, car c'est une zone autour d'une agglomération où il est interdit de construire. Cela ressemble beaucoup aux explications que nous pourrions donner pour *greenhouse*. Deux autres composés avec *long* comme premier élément utilisent l'adjectif dans

un sens non littéral : *long johns*, *long shot*. Pour le premier, il y a un contraste implicite, mais qui n'est pas direct, avec des sous-vêtements moins chauds parce qu'ils sont courts; pour le second, ce qui est sous-entendu par *long* est l'improbabilité de la réalisation effective et non pas littéralement une référence à un coup qui serait long (comme au golf, par exemple). En effet, l'accentuation simple peut souvent être utilisée afin de montrer que les relations entre les éléments ne sont pas simples et directes[63]. Les composés suivants rentrent dans le cadre de cette explication : *giant killer, little people, real estate, set theory*. Dans tous ces cas, le premier élément ne joue pas un rôle de description; c'est l'accentuation simple qui nous permet de le comprendre.

Dans les sections suivantes, nous ferons référence aux explications qui nous ont permis de désambiguïser les entrées pouvant s'accentuer des deux manières. Le tableau suivant les récapitule.

---

[63] Voir les propos d'Huart, 4.2.2.2.5 « L'accentuation : reflet d'une relation préconstruite (ou non) entre les éléments ».

Figure 4.30 : Règles d'accentuation des composés

Accentuation simple		Accentuation double	
Règle	Commentaire	Règle	Commentaire
Opposition Fonction/Nature[64]	Le premier élément indique la fonction du second	Opposition Fonction/Nature	Le premier élément indique la nature du second : Il est en *-ing*. Il désigne la matière dont le second est constitué. Il décrit le second.
Relation complexe entre les éléments[65] Mise en relief de la caractéristique représentative	Relation préconstruite ou lien à rétablir entre les éléments (à partir du contexte) Caractéristique pouvant désigner une ressemblance	Relation simple entre les éléments Localisation du premier élément par rapport au second	Le premier élément décrit le second Localisation spatiale ou temporelle
Contraste implicite[66] Deuxième élément général	→ sous-catégorisation		

Dans la suite de nos analyses, nous nous servirons de ces règles et nous formulerons également quelques autres hypothèses. A présent, passons à l'étude des composés de la table N-COMPACT qui commencent par une majuscule.

4.6.2 Etude des entrées ayant une majuscule

Nous avons utilisé la requête **Majuscules**[67] afin d'être en mesure d'étudier toutes les entrées de N-COMPACT qui commencent par une majuscule. Ce sont ces entrées que nous étudierons dans les deux premières rubriques ; sous la troisième, l'analyse s'élargit pour inclure des entrées de LPD-Cp étiquetées « tdmk » et quelques autres noms propres qui ont été inclus dans la table.

---

[64] Ginésy (1995).

[65] Les règles proposées dans cet encadré sont d'Huart (1984, 1989, 2002).

[66] Les règles proposées dans cet encadré sont de Boisson (1980).

[67] Voir Annexe A.

4.6.2.1 Vérification des statistiques de N-COMPACT

Notre premier objectif dans l'étude de ces entrées était de vérifier la portée des statististiques concernant la proportion d'entrées ayant un certain type d'accentuation par rapport à leur graphie[68]. En effet, il y a environ 1100 entrées qui commencent avec une majuscule; 790 doivent être exclues de l'analyse selon les critères de sélection définis dans le chapitre 2. En tenant compte des entrées restantes seulement, la proportion de N-COMPACT des entrées en «s» (accentuation simple) est de 53% et les entrées en «d» (accentuation double) représentent 47%, au lieu de 49% et 51% respectivement lorsque nous tenions compte de toutes les entrées. Cela fait une différence non négligeable car la proportion s'inverse, mais comparativement aux autres tables où la proportion d'entrées en «s» est de plus de 70%, ces changements ne remettent pas en cause les conclusions que nous avons pu formuler. Si nous ne tenions aucun compte de ces entrées, cela ne modifierait que de 1% les résultats (54% de «s» contre «46% de «d»).

4.6.2.2 Etude des entrées

En ce qui concerne les quelques 310 entrées qui correspondent aux critères de sélection établis pour LPD-Cp[69], 190 ont une accentuation double, contre 120 avec une accentuation simple. Le premier constat que nous faisons est qu'un très grand nombre de celles qui ont un schéma «d» correspondent à des cas où la

---

[68] Voir la partie 4.3.2.2 « Etude de la graphie ».

[69] Voir 2.1.3 « Les critères de sélection des composés pour le corpus d'analyse ».

relation entre les éléments est celle de la localisation, parce que l'élément se réfère à un nom de lieu, ou bien parce qu'il est un adjectif de nationalité : *Norfolk terrier, Cheddar cheese, Worcester sauce* sont des exemples de nom de lieu ; *Welsh rarebit, Scotch egg, Dutch courage, Siamese cat* ont des premiers éléments qui sont des adjectifs de nationalité. Cela représente environ 120 entrées, soit près des deux-tiers des entrées à accentuation double. Il y a une seule exception : *French doors*, qui a une variante en « d ». Nous reviendrons à la question de la variation à la fin de cette rubrique. Il y a un peu plus d'exceptions parmi les noms de lieux mais la moitié de celles-ci désigne des noms de gâteaux : *Madeira cake*, par exemple. En effet, Wells (Wells, 1990 & 2000) donne les mots en *cake* comme formant une sous-catégorie de composés à accentuation simple. Les autres ont des deuxièmes éléments très généraux, ce qui explique que la mise en relief porte sur le premier élément : c'est le cas pour *Oxford movement*. Dans le corpus oral, nous en avons des exemples avec le nom propre *New York* : ainsi, *New York company*[70], a une accentuation double, ce qui est reflété par la courbe du F_0 où il y a une reprise du contour (ce qui est typique de ce schéma, comme nous l'avons vu dans la partie 5). Cette accentuation s'explique donc par la localisation de *company* par rapport à *New York*.

L'étude de ces entrées soulève le problème de la sous-catégorisation en tant qu'explication du type de schéma. Peut-être y a-t-il une opposition entre les composés où nous nous servons du premier élément pour lui donner une

[70] Voir le graphique en 4.5.1.2.2.

appellation, comme les composés avec un schéma «d» que nous venons de citer, et les composés avec un schéma «s» où le premier élément fonctionne comme une sorte de sous-catégorisation, avec l'idée de contraste implicite et une mise en relief de cet élément. Au moins le tiers des composés à accentuation simple s'explique ainsi, et beaucoup d'autres peuvent se prévaloir de deux explications, dont celle-ci : c'est le cas de tous les composés en *cake* ainsi que d'autres : *Joshua tree, St John's wort, O'level, Easter egg*. Les trois premiers exemples sont des cas typiques de sous-catégorisation. En revanche, nous aurions pu attendre une accentuation double pour le dernier en fonction de la localisation temporelle. Toutefois, comme pour *Sunday school* qui a une fonction spécifique, l'œuf de Pâques, même si son existence est reliée à Pâques, désigne un type d'œuf très particulier qui a une fonction différente de l'œuf de poule[71].

L'explication selon laquelle le deuxième élément a une fonction spéciale qui est désignée par le premier peut expliquer environ un quart des entrées à accentuation simple : *Stanley knife, Bren gun, Geiger counter, Wendy house, Davy lamp* relèvent toutes de cette explication. Nous la proposons avec une certaine prudence, car il est beaucoup plus difficile d'associer une fonction à un nom propre qu'à un nom commun parce que le nom propre n'est pas *a priori* porteur de sens ; mais comme ces entrées le montrent, il peut le devenir.

Néanmoins, la majorité des entrées à accentuation simple s'expliquent tout simplement par la mise en relief du premier élément car le second est très général.

---

[71] Voir aussi les commentaires d'Huart (1989 : 75)

Ainsi, tous les composés en *disease* et *law* ont un schéma « s », ainsi que d'autres dont le dernier élément est *test* ou *syndrome* : *Paget's disease, Stanford-Binet test, Down's syndrome, Verner's law*. Nous pouvons aussi relier ces cas au contraste implicite et à la sous-catégorisation.

Il y a une dernière catégorie dans cet ensemble de composés qu'il est intéressant d'étudier: ceux où le premier élément est un acronyme ou une abréviation. Ces composés semblent avoir un schéma « s » : c'est souvent le cas aussi parce que le second élément est général : *ENT specialist*, par exemple, mais également dans d'autres cas : *X-chromosome, T-cell*. Nous avons relevé deux exceptions qui ont des constructions particulières car le second élément est un adjectif : *HIV positive* et *RH negative* (ou *positive*, bien sûr).

Il reste à expliquer les 80 entrées à accentuation double qui ne relèvent pas de la localisation. Nous avons formulé l'hypothèse qu'il existe une opposition entre les appellations qui portent un schéma « d » et des sous-catégorisations qui portent un schéma « s ». Les entrées avec le schéma « d » correspondent souvent à cette hypothèse, c'est-à-dire qu'elles ne désignent pas des sous-catégories mais donnent des noms particuliers à des référents auxquels ils sont associés : *Fosbury flop* est ainsi nommé, non pour entrer dans une série de *flop* mais pour rendre hommage à son inventeur et donner un nom à son saut. Peut-être est-ce le même phénomène que nous avons remarqué pour *horse chestnut* et *stormy petrel*. Nous avons également remarqué un nombre important de composés en *'s* avec un schéma « d » qui sous-entendraient l'interprétation d'appartenance : *Hobson's choice, Job's comforter, Achilles' heel, Occam's razor*, et qui désignent par

extension le même référent chez un autre individu. Nous reviendrons à ces considérations dans la prochaine section.

Enfin, nous voudrions terminer cette rubrique en comparant quelques autres cas. Nous avons retenu la localisation comme étant une explication convaincante de l'accentuation double des composés dont le premier élément désigne un lieu; *Vichy water* a donc un schéma exceptionnel car il est à accentuation simple. Une explication possible repose sur l'idée de sous-catégorisation et de contraste implicite: ce composé entre dans une série d'eaux, avec et sans nom propre, qui désigne des choix qui figureront peut-être sur un menu. C'est ce que son schéma accentuel nous indique. Si nous comparons *Beaufort scale* et *Bunsen burner*, nous devons expliquer pourquoi le premier a une accentuation simple alors que le second a une accentuation double. Ici, nous pouvons proposer deux explications: premièrement, en faisant référence à la sous-catégorisation et au contraste implicite, car il existe plusieurs types d'échelles, ce qui justifie le schéma «s» de *Beaufort scale*, alors que le *Bunsen burner* est unique, donc il a un schéma «d». Deuxièmement, *Bunsen*, comme *Fosbury*, semble jouer le rôle d'appellation plutôt que celui d'une fonction, ce qui implique l'accentuation double.

Le mot *number* se trouve à plusieurs reprises en deuxième élément, mais il est associé à des schémas différents: nous en citons trois exemples: *PIN number*, et *Fibonacci numbers* qui ont un schéma «s», et *Avogadro('s) number* qui a un schéma «d». Le premier exemple, *PIN number*, est le plus facile à expliquer car il correspond au type d'accentuation («s») que nous trouvons dans les composés dont le premier élément comporte un acronyme ou une abréviation. En revanche,

dans le cas de *Avogadro('s) number*, qui est à accentuation double, peut-être pouvons-nous considérer qu'il existe une relation d'appartenance entre *Avogadro* et le nombre, qui serait indiqué par l'apostrophe, comme pour *Achilles' heel*, mais l'apostrophe est elle-même une variante graphique. Cette explication n'est pas convaincante, et bien qu'il y ait une série de nombres, dont *Fibonacci numbers*, l'entrée *Avogadro number* ne rentre pas dans cette logique de sous-catégorisation.

Enfin, nous terminons avec la question de la variation qui a été soulevée précédemment. Nous émettons l'hypothèse que l'existence de variantes sous-entend qu'il y a deux manières d'interpréter la même réalité pour ces entrées: *Gallup poll* et *Gladstone bag*, par exemple, ont un schéma « d » et une variante en « s ». La prononciation principale de *Gallup poll* reflète simplement son appellation: ce type de sondage a été nommé d'après son inventeur; en revanche, la variante sous-entend sa fonction d'indicateur de l'opinion, surtout en période électorale. Il est probable que cette deuxième prononciation remplacera la première dans les années à venir; c'est déjà la prononciation américaine principale. La variante de *Gladstone bag* est en fait la prononciation américaine: il s'agit d'un type particulier de sac, ce qui nous aurait plutôt fait penser à un schéma « s », mais il a été ainsi appelé pour faire honneur à Gladstone: c'est ce qui explique sa prononciation britannique. C'est aussi le cas de *Worcester sauce* qui a une accentuation double en anglais britannique qui s'explique par la localisation, mais en américain, elle a une accentuation simple: c'est donc la sous-catégorisation qui l'emporte. Un dernier exemple de variation est *French doors* qui est indiqué avec un schéma « s » contrairement à toutes les autres entrées ayant

un adjectif de nationalité en premier élément. Peut-être l'accentuation signale-t-elle la perte de l'identité de localisation du premier élément, mais elle n'a pas complètement disparu car il existe toujours la variante en « d ». Il est probable que cette variante sera souvent utilisée à cause du parallèle qui existe entre *French doors* et *French windows*, qui est indiqué comme ayant toujours le schéma « d ».

Dans cette section, nous avons voulu étudier en détail une partie de notre base de données. Nous avons d'abord vérifié sa portée par rapport à l'ensemble de la table de N-COMPACT, puis analysé les entrées qui rentrent dans notre problématique. Les explications de la première section ont été reprises et quelques hypothèses complémentaires ont été émises: premièrement, qu'il y aurait une opposition entre le rôle de l'appellation du premier élément (schéma « d ») par rapport à son rôle de sous-catégorisation (schéma « s ») qui résulte parfois dans des conflits accentuels; deuxièmement, le type d'accentuation des composés ayant des premiers éléments qui sont des acronymes ou des abréviations serait « s ». Avant de passer à la section suivante, nous allons étudier en plus des entrées dont le premier élément est un nom propre, celles qui sont composées d'éléments qui sont tous les deux des noms propres.

4.6.2.3 Etude des entrées noms propres de la base de données

Dans le chapitre 2, nous avons expliqué que nous voulions inclure dans notre base de données un certain type de noms propres qui désignent des objets, des associations, des faits de la société[72]. Lors de l'extraction des données, nous

[72] Voir 2.1.4.3 « Noms propres ».

avons marqué ces entrees de LPD-Cp par un code que nous avons pu utiliser pour les isoler grâce à la requête **RECH-COMM-LPD**[73]. La requête **RECH-catgr-LPD**[74] nous a également permis d'isoler toutes les entrées considérées comme étant des noms de marque qui sont étiquetés «tdmk». Ces deux requêtes retournent des entrées souvent communes. La mini-base que nous venons d'étudier en comporte aussi un certain nombre. Il y a environ 200 entrées à étudier.

Le premier constat qui s'impose est le très faible taux d'entrées à accentuation double dans la table LPD-Cp: seulement 17 entrées sur 152 étiquetées «tdmk» et une vingtaine dans la liste des entrées marquées «OPC» sur 180. En revanche, N-COMPACT en comporte une vingtaine pour seulement quelques entrées à accentuation simple.

Nous allons d'abord expliquer le schéma des entrées à accentuation double des noms de marque, qui sont donc minoritaires par rapport à cet ensemble. Parfois, elles comportent des éléments qui se suivent sans ordre hiérarchique: ils se comportent comme des composés co-ordonnés, comme *Ryvita* qui est à la fois la céréale *rye* et la vitalité que l'on devine dans *vita*. L'accentuation est une manière de les associer et de souligner le double effet du produit. Toutefois, le statut de composé de ce nom de marque est discutable car le second élément n'est pas connu, mais deviné. C'est souvent le principe des noms de marques qui

---

[73] Voir Annexe A. Le code recherché est {OPC} dans le champ « Type de composé ».

[74] Voir Annexe A.

cherchent à accrocher l'attention et se livrent à des troncations et à des jeux sur l'orthographe.

Dans *Pepsi-Cola*, nous retrouvons le même principe d'égalité, mais cette marque s'inspire de *Coca-Cola* dont le nom rappelle les composés réduplicatifs. Le rime des deux premières syllabes des éléments est accrocheur et se mémorise bien: l'objectif du publicitaire est atteint. Deux autres catégories de noms de marques avec un schéma «d» peuvent être retenues: celle des noms qui sont des phrases: *Sleepeezee*, qui jouent à partir de l'orthographe *sleep easy*; et les noms qui comportent plusieurs éléments: *Punch and Judy*, où nous retrouvons le principe d'équilibre des composés co-ordonnés, qui se reflète dans le schéma accentuel.

Pour les noms de marque à accentuation simple, qui sont très majoritaires, ils correspondent parfois aux règles que nous avons étudiées: *Play-Doh* est une pâte pour jouer (nous constatons le même jeu sur l'orthographe). Toutefois, l'accentuation simple sur ces mots semble jouer un rôle plus subtil: *Kwiksave* et *Freefone* sont des exemples de jeux d'orthographe, mais aussi d'une sorte de concentration des caractéristiques du produit qui se reflète dans l'orthographe et par le schéma accentuel. Le choix de ces noms de marque n'a rien d'arbitraire: l'idée d'économie rapide est établie par l'accentuation pour *Kwiksave*, de même que pour *Freefone*, qui est un service qu'offre British Telecom et qui permet à celui qui appelle de ne pas payer la communication (elle est payée par le destinataire). Nous en comprenons immédiatement l'intérêt. L'accentuation simple du nom de la marque *Flymo* qui fabrique des tondeuses sur coussin d'air

permet de comprendre le côté léger de l'appareil qui est le grand argument de vente, le second élément désignant *mow*, en tronquant à nouveau l'orthographe.

Ces exemples sont anecdotiques, mais montrent que la composition peut être utilisée à des fins publicitaires avec des jeux de langage et d'écriture, et que l'accentuation joue aussi un rôle dans ce type de manipulation de la langue. Nous notons au passage la préférence pour des noms ayant un petit nombre de syllabes, surtout les entrées ayant un schéma «s»: 129 des 135 entrées à accentuation simple ont deux ou trois syllabes; les 6 autres en ont 4, alors que sur les 17 entrées à accentuation double, 11 ont deux ou trois syllabes et 6 comportent 4 ou 5 syllabes. C'est la préférence marquée pour les mots courts qui est intéressante: elle est sûrement motivée par le souci d'accrocher le public, de proposer des noms faciles à mémoriser et qui correspondent à un rythme naturel, d'où le choix de l'accentuation simple.

Les autres noms propres de la base de données ont des caractéristiques plus traditionnelles. Ces entrées sont aussi généralement à accentuation simple: *Roundhead, Greenpeace, Goldilocks, Superbowl* ont toutes des adjectifs en premier élément mais ces adjectifs ne décrivent pas l'ensemble. Ce sont des caractéristiques qui ont été choisies pour les représenter. Les soldats de Cromwell avaient des casques ronds; l'association *Greenpeace* milite pour l'environnement, d'où le mot *green*; chacun sait qu'une jeune fille aux grands cheveux dorés (Boucle d'Or) a mangé la soupe (dans la version anglaise, il s'agit de *porridge*) du plus petit des trois oursons; enfin, la compétition sportive du *Superbowl* est la plus grande des manifestations de ce type. Nous citons un autre exemple qui

figure dans le corpus oral: *Dark Ages* qui est à accentuation simple. Nous pouvons proposer la même explication car *dark* est utilisé au sens figuratif; c'est la caractéristique qui reflète le mieux cette époque.

Les quelques entrées ayant une accentuation double sont souvent des adjectifs suivis de noms, où l'adjectif sert à déterminer la nature du deuxième élément, comme dans *Iron Curtain, Black Mass, Moral Majority, Yellow Pages*. Nous rencontrons aussi, dans d'autres types de formation, une localisation du deuxième élément par rapport au premier, comme dans *Highway Code*, et peut-être *Hell's Angel*.

Cette étude nous a permis de confirmer le bien-fondé des règles pour l'accentuation des composés. Dans la dernière section de cette partie, nous allons étudier les éléments qui figurent dans le dictionnaire et dans le corpus oral. Ainsi, nous pourrons poursuivre l'analyse du comportement des composés lexicalisés dans un contexte de discours que nous avons commencée dans la partie précédente; nous pourrons aussi établir des prévisions pour les unités non lexicalisées et les évaluer.

4.6.3 Etude des entrées comportant un des éléments des extraits du corpus oral

Afin d'être en mesure d'étudier le lien entre les règles d'accentuation des composés lexicalisés et leur pertinence pour les formations nouvelles, nous avons

créé une autre mini-base[75] à partir des éléments qui entrent dans la composition des extraits sonores du corpus oral. A l'aide de la requête **RECH-EL**[76], nous avons listé les entrées des 4 tables de la base de données correspondant à une recherche sur la graphie. Cette recherche s'est effectuée avec certains éléments en position initiale et finale, mais nous ne ferons pas de synthèse en ce qui concerne ces différences, car l'étude a montré que la position n'est pas pertinente : c'est la relation entre les éléments qui est importante[77].

L'analyse des résultats se fera selon les différents types de composés que nous avons cités dans la deuxième partie de ce chapitre[78]. Nous étudierons d'abord les formations de base : les composés endocentriques, qui sont très majoritaires dans notre corpus ; les composés co-ordonnés ; les composés exocentriques ; les composés synthétiques ; puis, nous nous intéresserons à des formations particulières : les composés réduplicatifs ; les entrées avec une apostrophe ; enfin, les unités ayant une construction cumulative et les entrées ayant plusieurs éléments. La plupart de ces catégories sont représentées par des extraits sonores du corpus oral et nous essayerons de tester les prévisions de l'accent par rapport à ces exemples. Sous cette rubrique, nous ferons la synthèse des résultats, sans passer toutes les entrées en revue.

---

[75] Voir Annexe F. Nous avons inclus un échantillon de cette base comprenant les éléments qui se sont révélés les plus intéressants à l'analyse : *break, fly, green, work*. Cet échantillon représente environ 250 entrées.

[76] Voir Annexe A.

[77] Voir aussi Huart (1989).

[78] Voir la conclusion de 4.2.1.2

4.6.3.1 Les différents types de composés

4.6.3.1.1 Composés endocentriques

Si nous étudions d'abord les entrées non lexicalisées du corpus oral, nous en avons un nombre important qui correspond à cette catégorie: *portrait hole*, par exemple. Chacune des occurrences de ce composé inventé porte une accentuation simple, tout comme le composé lexicalisé *keyhole*[79]. Ce schéma s'explique aisément par le fait que le trou a une certaine fonction[80]. La même explication peut être proposée pour *portrait hole*, mais la relation sémantique entre les deux éléments est différente. Le composé lexicalisé désigne une ouverture (trou) qui est à l'intention d'une clé, alors que *portrait hole*, qui est un composé inventé dans le cadre du livre, *Harry Potter and the Philosopher's Stone*, désigne une ouverture par laquelle on peut passer, la présence de l'ouverture étant signalée par le portrait de la Grosse Dame qui en cache l'entrée.

Cette même relation de fonction fournit l'explication du schéma d'autres composés tels que *racing broom, racing bike, mountain bike*. La réalisation de l'accent dans tous ces cas est sur le premier élément. Si nous comparons *mountain bike* à *mountain path*, qui aura un schéma «d», la différence est claire: pour le premier, il s'agit d'un vélo ayant une fonction particulière; nous pourrions également parler de sous-catégorisation, par opposition à *racing bike*. En revanche, *mountain path* est une simple localisation du chemin par rapport à la

---

[79] Voir les graphiques en 4.5.1.2.1.

[80] Voir Ginésy (1995).

montagne.[81] Comme nous l'avons vu dans la partie 4, nous n'avons pas trouvé *racing* dans la composition d'entrées composées lexicalisées, mais nous avons d'autres exemples de vélo: *exercise bike, push bike*, tous les deux ayant un schéma « s ». L'idée de la fonction du vélo est présente, surtout pour le premier ; le second relève plutôt de la règle du contraste implicite.

Le corpus oral présente d'autres exemples qui permettent de comparer les néologismes et les entrées du dictionnaire[82] ; nous en avons relevé avec l'élément *broom* : *broomstick, broomshed.* Ces deux extraits sonores se prononcent avec une accentuation simple, et les courbes du F_0 et de l'intensité montrent la proéminence sur *broom.* Pour ces deux composés, nous retrouvons les explications que nous avons déjà proposées : pour *broomstick*, accentué /10 d'après les indications du dictionnaire, il s'agit de contraste implicite, de sous-catégorisation, qui entraînent la mise en relief du premier élément ; alors que pour *broomshed*, le composé désigne un lieu dont la fonction est de stocker les balais (nous sommes à Poudlard, école de sorciers). Les entrées en *broom* de notre mini-base ne sont pas nombreuses : *carpet-broom* (EPD), *broomstick, broomrape* (LPD et NPD), *new broom* (N-COMPACT). Seule la dernière entrée s'accentue avec un schéma « d ». Nous avons déjà expliqué le schéma à accentuation simple de *broomstick* ; *broomrape* suit la même logique, alors que pour *carpet-broom* le premier élément

---

[81] Voir Huart (1989) qui donne l'exemple *mountain road.*

[82] Nous n'avons pas pu commenter toutes les entrées du corpus oral qui correspondent à ce schéma et à cette explication : *dragon liver, dragon egg, dragon heartstring* et bien d'autres. Voir Annexe C.

désigne la fonction du second. Pour *new broom*, cette entrée fait partie de l'expression « new brooms sweep clean » et se réfère aux personnes nouvellement nommées dans une entreprise et à leurs actions. Le schéma à accentuation double correspond à la relation directe entre les éléments : *new* se réfère à la nature de *broom*[83].

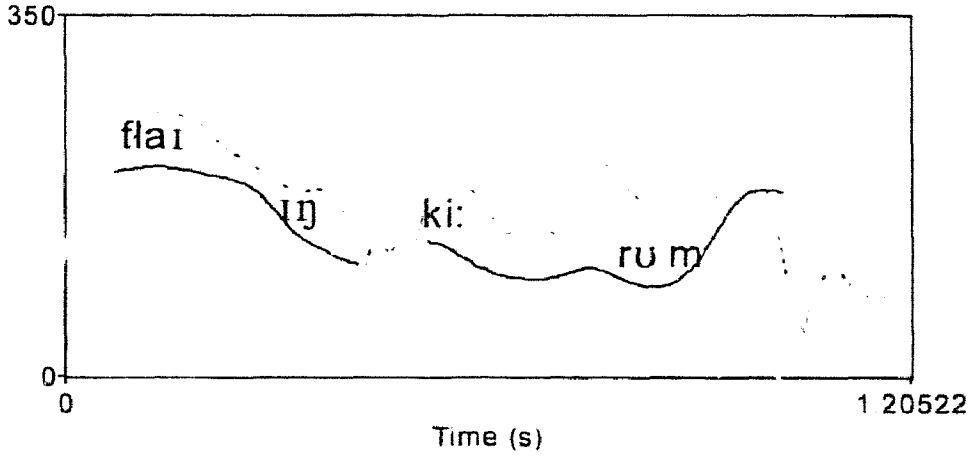
Nous voudrions donner un dernier exemple de ce type de néologisme à accentuation simple : *flying-key room*. Toutes les entrées en *room* (près de 70 entrées dans LPD-Cp et N-COMPACT, une trentaine dans NPD) sont à accentuation simple, sauf *front room* dont le schéma à accentuation double peut s'expliquer par la localisation du deuxième élément par rapport au premier[84]. L'unité inventée *flying-key room* suit cette tendance et entre dans la série des mots en *room* avec une accentuation simple. Nous pouvons parler de contraste implicite ou de sous-catégorisation. Cet exemple mérite d'être commenté parce que son premier élément est en fait déjà composé, et nous avions émis l'hypothèse dans le chapitre 3 (voir 3.1.2.2 « Le pied ») que l'accent serait sur *key* et non sur *fly*. Or, l'analyse du graphique démontre le contraire.

---

[83] Toutes les entrées en *new* ont un schéma à accentuation double car l'adjectif désigne la nature du second élément. Il n'y a qu'une exception sur une centaine d'entrées : *new town*, qui a d'ailleurs une variante en « d ».

[84] Dans le corpus oral, nous avons un exemple d'un composé endocentrique dont l'accentuation est double et qui s'explique de cette manière : *kitchen sink*, qui se réfère à un évier de cuisine, c'est-à-dire que le premier élément localise le second. Ce n'est pas sa fonction, mais sa localisation.

Figure 4.31 : room\flykeyrm : Courbe de F_0 et I



La proéminence est clairement sur *fly*. Deux explications sont possibles: La proéminence sur le début de l'unité pourrait être l'effet d'un déplacement de l'accent. Toutefois, cette explication n'est pas convaincante car il n'y a pas de raison rythmique de le déplacer puisque *room* est désaccentué. La seconde explication repose sur l'analogie entre cette unité et les autres de la série qui s'accentuent sur le tout premier élément. Le rythme de l'ensemble ressemble donc à celui de *trophy room* (voir les graphiques en 4.5.2). Nous retiendrons cette hypothèse. La prévision de l'accent doit donc tenir compte de tous ces facteurs.

Les entrées de la mini-base avec *fly* sont très intéressantes. La recherche n'a retourné aucune entrée en *-ing* dans NPD, car ce dictionnaire écarte tous les dérivés; il n'y en a pas dans LPD non plus et très peu dans EPD; en revanche, la recherche N-COMPACT est fructueuse (voir Annexe F). Toutes les entrées à accentuation double correspondent au sens où le second élément vole: *flying fish, flying saucer*; y compris lorsque cette activité n'est pas la raison d'être du

référent : *flying doctor*. Les entrées à accentuation simple représentent des cas où la fonction du second élément est de voler : *flying squad, flying boat*. Nous retrouvons donc l'opposition « fonction/nature »[85].

Cette catégorie de composés regroupe la majorité des entrées de notre corpus oral et un grand nombre de celles de la mini-base. L'analyse nous permet de conclure que les règles d'accentuation permettent effectivement de prévoir le schéma des nouvelles unités dans la plupart des cas. Nous avons été surprise par *flying-key room*, que nous avons commenté plus haut ; sinon, les autres entrées correspondent bien à nos attentes. Nous allons nous intéresser à une autre catégorie, beaucoup plus restreinte : celle des composés co-ordonnés (ou *dvandvas*).

4.6.3.1.2 Composés co-ordonnées

Nous n'avons trouvé que quelques exemples de ce type de formation qui n'est pas très courant. La recherche des éléments se terminant par *-ing* nous donne *drink-driving* ; l'élément *work* nous donne *worker-priest*, et *top* nous permet de relever deux entrées : *tip-top* et *top-hole*. Ces deux dernières ne sont pas complètement convaincantes, surtout *top-hole* qui est ressentie comme étant une formation d'Adj + N, qui ne peut être considérée au même niveau hiérarchique du point de vue de la syntaxe. Cependant, nous l'avons relevée afin de la comparer à *tip-top* qui est une séquence de deux noms ayant le sens de l'adjectif « génial ». En

---

[85] La même opposition existe pour les entrées avec *working* : *working class* a une accentuation double : la classe travaille ; alors que *working party* a une accentuation simple : c'est un groupe dont la fonction est de travailler.

fait, cette formation rappelle celle des composés réduplicatifs car nous avons deux fois le même sens répété afin de renforcer le sens du tout, associé à une phonie semblable (seule la voyelle change).

Dans tous ces cas, le schéma accentuel est «d» ce qui reflète le statut d'égalité des deux éléments. Pour *drink-driving*, deux activités sont associées: boire et conduire; les deux activités réunies constituent un délit. Pour *worker-priest*, le prêtre est à la fois homme d'église et un travailleur au sens social du mot. L'accentuation double souligne cette double compétence; il ne s'agit pas d'une sous-catégorisation: dans ce cas, l'accentuation serait simple.

Nous n'avons pas d'extrait sonore qui correspond à cette catégorie, donc nous ne pouvons pas tester les prévisions du schéma accentuel. Pour ce genre d'entrée, nous nous attendons à une accentuation double pour les raisons que nous avons expliquées.

4.6.3.1.3 Composés exocentriques

La catégorie des composés exocentriques, aussi appelés *bahuvrihis*, est plus étoffée, mais il y a un seul exemple possible dans le corpus oral: *vulture fund*. Il s'agit d'une firme d'investisseurs qui se comporte comme le font les oiseaux de proie, en s'intéressant aux compagnies moribondes: ce n'est pas un fonds d'argent en soi. Nous nous attendons à un schéma «s» pour ce type d'unité car la métaphore ne permet pas d'interpréter la relation entre les éléments simplement. Néanmoins, nous l'avons vu dans la partie 5, une des quatre occurrences de *vulture fund* a une proéminence sur *fund*. Nous en avons conclu qu'il s'agissait d'un cas de variation libre selon le choix du locuteur car il n'y avait pas de raison

syntaxique ni sémantique. La prévision de l'accent est donc troublée par cette variation.

La mini-base nous permet de relever quelques autres exemples: *greenhorn* et *greensha ık*, dans la liste des entrées en *green*; *fly-weight* et *fly-by-night*, dans la liste des entrées en *fly*[86] et *dark horse* à partir de l'élément *dark*. Nous remarquons avec intérêt que chacun de ces exemples utilise un sens différent de l'élément: *green* dans *greenshank* emprunte le sens littéral de l'adjectif; l'oiseau désigné a de grandes pattes de couleur verte, alors que c'est l'inexpérience qui est sous-entendue dans *greenhorn* qui est un jeune homme avec la caractéristique indiquée, et non pas une corne. De même, pour *flyweight*, il s'agit d'une personne (un boxeur) dont le poids est léger; le sens de *fly* correspond à celui de la mouche. Pour *fly-by-night*, ce composé désigne une personne irresponsable, qui papillonne la nuit, à qui on ne peut pas faire confiance; le mot *fly* s'emploie dans son sens verbal. Toutes ces entrées ont le même schéma «s» ce qui reflète la relation complexe entre les deux éléments. Selon Huart (2002), l'accentuation et la désaccentuation indiquent qu'il y a un lien qui doit être reconstruit par l'interlocuteur pour comprendre le sens du tout[87]. Le dernier exemple est donc exceptionnel car *dark horse* porte une accentuation double alors qu'il s agit non d'un cheval mais d'une personne qui ne montre pas ce qu'il a derrière la tête, d'où l'adjectif *dark*. Toutefois, nous pourrions considérer que cet adjectif désigne la

---

[86] Voir Annexe F.

[87] Voir Huart (2002) et 4.2.2.

nature du nom *horse* ; ce qui surprend est le lien entre le cheval et l'homme. S'il y avait une accentuation simple, peut-être que l'on pourrait interpréter l'ensemble comme un espèce de cheval. Sans doute est-ce la raison de cette accentuation inhabituelle. Néanmoins, le schéma accentuel attendu est «s» pour ce type de composés.

4.6.3.1.4 Composés synthétiques

Avant d'étudier les composés verbaux, nous devons rappeler ce que nous entendons par ce terme. Pour Roeper & Siegel (1978), il s'agit de toutes les entrées dont le deuxième élément est verbal, et comporte une terminaison en *-er*, *-ing* ou *-ed*. Nous nous sommes intéressée aux cas des entrées en *-ing* et quelques cas en *-er* et *-ed* en élément final ; nous avons aussi relevé des entrées qui commence par un élément en *-ing*.

Nous avons déjà commenté des entrées en *flying* et *working* dans l'étude des composés endocentriques (4.6.3.1.1). Nous retrouvons la même opposition «fonction/nature» pour les entrées à accentuation simple et à accentuation double qui commencent par un élément en *-ing* : *walking stick* en est un exemple. Le composé désigne une canne dont la fonction est de nous aider à marcher ; il a une accentuation simple. Dans le cas d'une accentuation double, ce sera interprété comme une unité désignant un bâton qui marche (ce serait la nature du bâton). Dans le corpus oral, nous n'avons que des extraits où l'élément en *-ing* détermine la nature du second élément : c'est le cas des unités à construction cumulative qui commencent avec *talking* : *talking bottle top*, *talking vodka bottle top*. Le

capuchon parle : ce n'est pas sa fonction mais sa nature, ce qui se reflète par son accentuation double[88].

Pour les entrées en *-er*, nous nous intéresserons à celles avec *breaker* que nous avons trouvées à partir de la recherche sur *break*[89]. Lorsque le second élément est agent, donc en *-er*, le premier élément est toujours accentué[90]. Cette accentuation indique que, dans *horsebreaker*, il s'agit d'une personne qui dompte les chevaux ; *icebreaker* est quelque chose qui permet de briser la glace. Ces cas ressemblent à ceux des composés exocentriques où l'on reconstitue la relation entre les éléments, mais pour les exocentriques, l'agent est absent alors qu'il est présent dans le suffixe pour les composés verbaux en *-er*.

Dans l'ensemble de la table LPD-Cp, les trois-quarts des entrées se terminant par *-ing* ont une accentuation simple. Pour ces entrées, si le premier élément est l'objet du second, alors il y a une accentuation simple : *backbreaking*, par exemple, qui peut se gloser « qui brise le dos ». Ce cas s'apparente aux composés en *-er*. Pour *groundbreaking* qui fait partie de notre corpus oral, son accentuation est parfaitement conforme aux indications, et le schéma « s » est clairement perçu. Les entrées à accentuation double correspondent à des cas où le premier élément indique la nature du second : *long-playing*.

---

[88] Voir Figure 4.18.

[89] Cette requête nous a aussi retourné *breakfast* dans NPD que nous avions écarté de l'analyse car il s'agit d'un composé historique.

[90] Voir Huart (1989), Ginésy (1995).

Les entrées se terminant par *-ed* sont majoritairement à accentuation double. Sur 226 entrées de LPD-Cp, 22 ont un schéma « s » et 13 d'entre elles ne sont pas des composés en *-ed* mais des composés dont le second élément est *bed*, *weed*, *seed*, *feed* et *shed*, ce qui donne un aperçu du type de problème que pose la recherche de séquences courtes. De plus, certains participes passés ne se terminent pas par *-ed* Nous n'avons pas poursuivi l'analyse, mais nous pouvons nous fier à l'ordre de grandeur que ces résultats nous donnent. Toutefois, ces adjectifs qui sont majoritairement à accentuation double, sont souvent sujets au déplacement d'accent d'après les indications du dictionnaire.

4.6.3.2 Les composés réduplicatifs

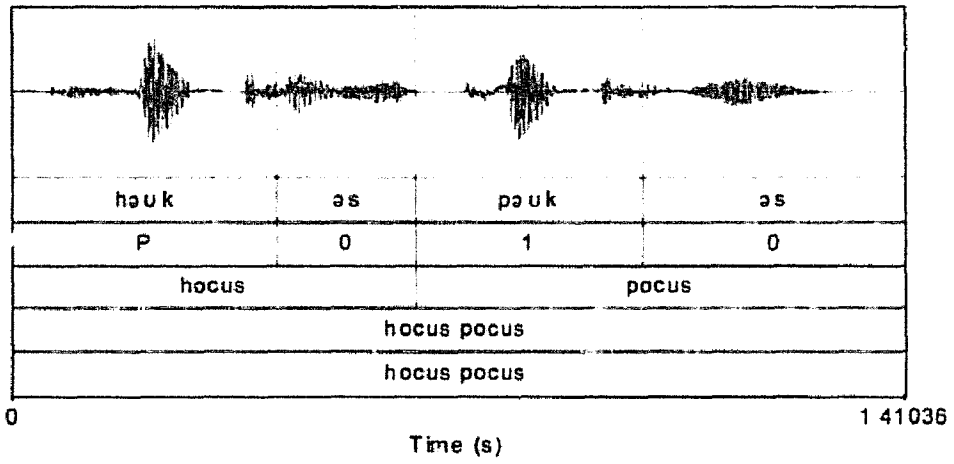
Les composés réduplicatifs que nous avons relevés dans LPD-Cp[91] sont au nombre de 118 avec un peu plus d'entrées ayant un schéma « d » (65) que « s » (53). Nous remarquons que les entrées à accentuation simple ont tendance à être plus courtes puisque 31 d'entre elles ne comportent que 2 syllabes alors que la moitié des entrées à accentuation double en ont 4. Toutefois, il ne s'agit que de tendances Un autre phénomène intéressant est le taux de variation. Cette tendance a été notée par Thun (1963), qui a fait une étude de ces mots. Le taux indiqué dans LPD-Cp est de près d'une entrée sur cinq ce qui montre l'instabilité accentuelle de ce type de formation. La variation concerne plus particulièrement les entrées à accentuation simple (11 sur 53).

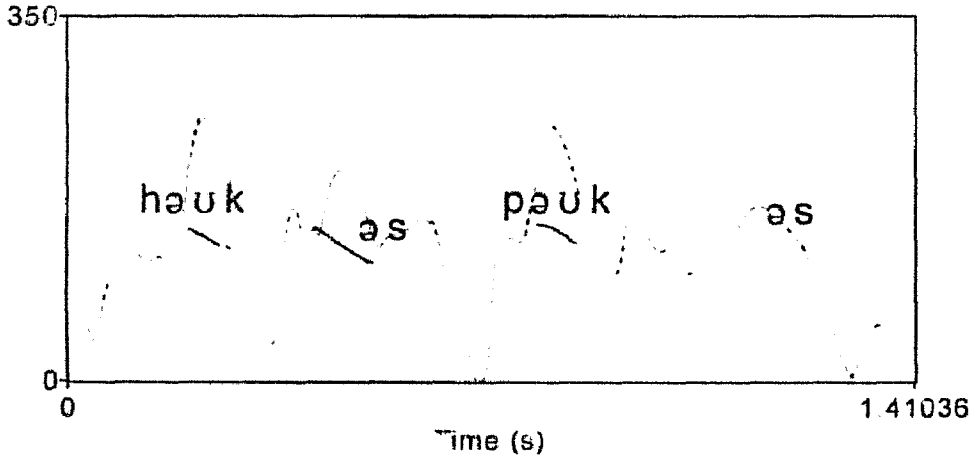
---

[91] Voir Annexe A, requête **RECH-COMM-LPD**.

Nous avons un exemple dans notre corpus oral : *hocus pocus*, qui a une accentuation double. Le graphique suivant montre bien la reprise de courbe que nous associons à ce type de schéma, mais aussi la déclinaison de F_0 car la seconde courbe est plus basse que la première, au point où le F_0 sur la deuxième syllabe n'apparaît pas :

Figure 4.32 : test\hocuspo1 : Oscillogramme et courbe de F_0 et I





Nous avons fait référence sous la rubrique des composés co-ordonnés aux caractéristiques de cette formation. Les deux éléments se ressemblent phonétiquement. Souvent, un segment change : si c'est une consonne, alors les

éléments riment, comme dans *walkie-talkie* ou *hurly-burly* (le premier a un schéma « d », le second « s » avec une variante en « d »). Dans d'autres cas, c'est la voyelle qui change : *ticktack*, *topsy-turvy* (le premier est « s », le second « d »). Parfois, il y a des simples reprises de l'élément : *choo-choo*, *clever-clever* (le premier est « s », le second « d »). Il ne semble pas qu'il y ait une corrélation entre le type de changement et le schéma accentuel. Il est évident que ces entrées relèvent souvent d'un niveau de langue qui est ou enfantin ou familier, parfois humoristique ; ce sont des créations qui s'apparentent à des jeux de langue comme nous en avons vu pour les noms de marque. Ainsi, nous ne sommes pas en mesure de proposer des règles pour cette formation, car, en général, il n'y a pas de relation particulière entre les éléments. Néanmoins, il y a quelques cas où nous retrouvons des relations dont nous avons discuté : *pooper-scooper* (outil indispensable pour promener le chien) et *gender-bender* (personne qui affiche une image sexuellement ambigüe) s'apparentent aux composés synthétiques en *-er* : leur schéma est « s ».

Les composés formés de cette manière peuvent être facilement mémorisés ; ils imitent souvent des bruits (comme *ticktack* qui correspond au bruit des aiguilles d'une montre et *choo-choo* qui est celui du train à vapeur) ; d'autres n'utilisent que partiellement des mots existants mais ces pseudo-reprises permettent de renforcer le sens de l'ensemble, comme *argy-bargy*. Peut-être que les cas de renforcement auraient tendance à avoir une accentuation double alors que les entrées qui doivent être aisément retenues seraient plutôt à accentuation simple, mais *fuddy-duddy*, par exemple, a une accentuation simple alors qu'elle

relève autant de l'interprétation de renforcement du sens que de celle de la mémorisation facile. La question reste ouverte et mérite un travail plus approfondi, mais ce type de formation est en marge de nos travaux et nous ne pouvons pas développer l'analyse sans entrer dans des considérations de formation des mots qui dépassent notre cadre.

4.6.3.3 Les composés en 's

Nous avons fait remarquer sous la rubrique de l'étude des entrées commençant par une majuscule qu'il se trouve un nombre considérable d'entrées avec une apostrophe où il existe une relation d'appartenance entre les éléments : *Achilles' heel* est le talon appartenant à Achille, ce qui sous-entend un endroit fragile car c'est de cette manière qu'Achille a été touché : par extension, cette unité désigne un endroit fragile pour toute autre personne. Ces entrées ont une accentuation double.

Dans le reste de la base de données, nous trouvons beaucoup d'autres exemples de ce type. Bien que la relation soit directe, le sens de l'ensemble est une sorte de généralisation du cas individuel auquel le premier élément fait référence : *shepherd's pie, mother's ruin, beginner's luck, athlete's foot*. Ainsi, le hachis est le plat des bergers, le gin est la ruine des mères (cette entrée provient du *Cockney rhyming slang*), la réussite d'un non-spécialiste est la chance du débutant, et les mycoses du pied sont la maladie des athlètes. Cette relation directe d'appartenance entre les éléments s'applique aussi à des entrées ayant un sens tout à fait littéral, comme *potter's wheel, life's work, state's evidence*.

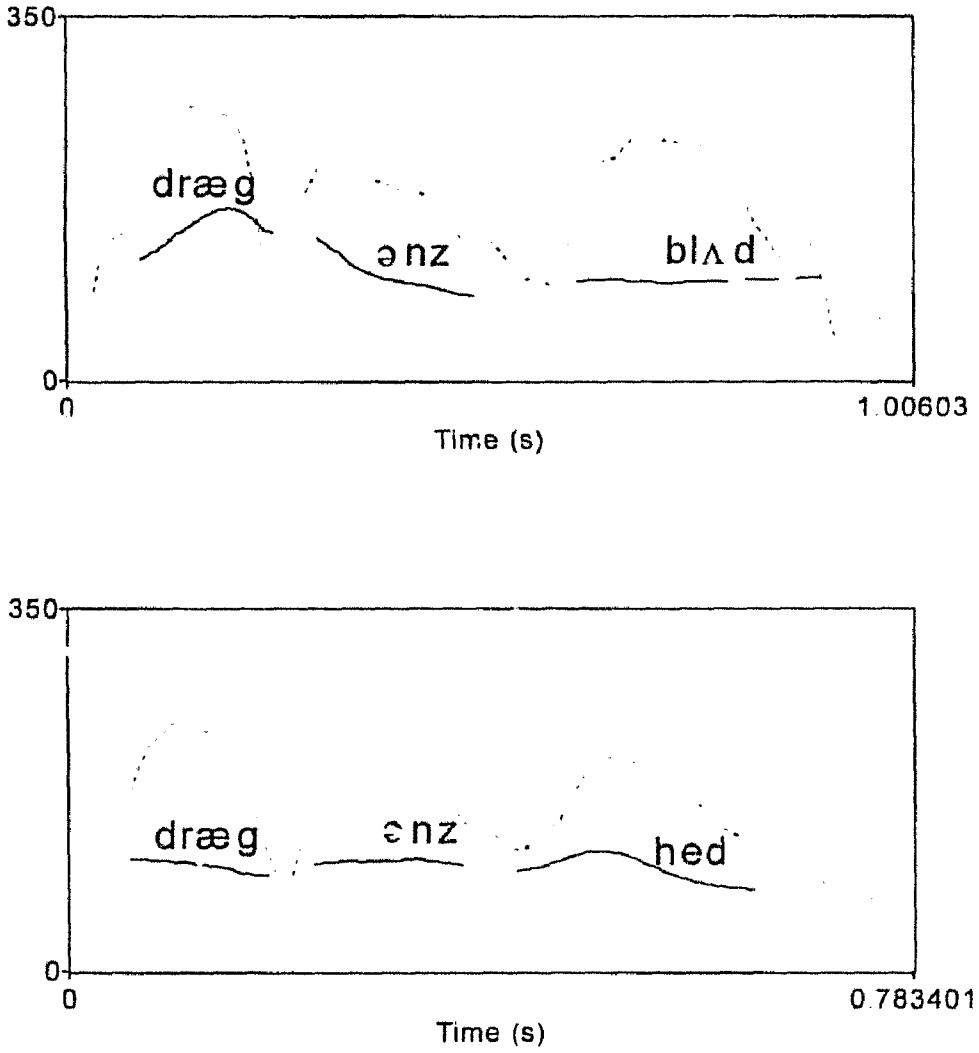
Néanmoins, il existe une autre catégorie d'entrées avec apostrophe où ce séparateur joue le rôle de mise en relation (voir Marchand, 1969 : 27). Ce sont souvent les génitifs dits génériques[92]. Dans ces cas, les entrées sont à accentuation simple, et le sens de ces ensembles est souvent figuratif : *bull's eye* ne se réfère pas à l'œil du taureau, mais au centre d'une cible. Nous retrouvons les explications que nous avons proposées pour les composés exocentriques. La relation entre les éléments est complexe. C'est aussi le cas pour les exemples suivants : *donkey's years, child's play, doll's house*. Le premier désigne une très longue durée ; le second implique la facilité de la démarche, comme si c'était un jeu d'enfant ; enfin, le troisième se réfère à une maison en miniature : elle n'appartient pas à la poupée, elle est à son intention. Nous retrouvons donc aussi l'idée de fonction.

D'autres exemples à accentuation simple peuvent s'expliquer par un contraste implicite : *driver's licence, master's degree*. C'est aussi le cas d'un des extraits sonores : *dragon's blood* qui a une accentuation simple, comme *dragon liver* : il s'agit d'un contraste implicite, d'une sous-catégorisation. Cette accentuation s'oppose à celle de *dragon's head* qui est un syntagme où l'apostrophe correspond au sens du génitif déterminatif et il y a une relation d'appartenance entre *head* et *dragon*. Nous proposons les deux courbes qui montrent clairement ces différences accentuelles.

---

[92] Voir Huart (1986).

Figure 4.33 : dragon\dragblood1 & dragon\draghead1 : Courbes de F_0 et I



L'intensité n'est pas révélatrice mais nous avons la reprise de la courbe du F_0 sur *head* qui est caractéristique de l'accentuation double, alors que le F_0 reste bas sur *blood*. De plus, la durée de *head* est supérieure à celle de *blood*.

Pour terminer cette rubrique, nous voulons étudier deux paires qui ont des schémas différents : *sod's law* / *Verner's Law* et *New Year's Day* / *Mother's Day*. Pour le premier exemple de chaque paire, il y a un schéma « d », et un schéma « s » pour le second. Y a-t-il une différence dans la relation entre les éléments pour

expliquer ces contradictions apparentes? Nous avons proposé en 4.6.2.2 que l'opposition entre le rôle de l'appellation du premier élément et celui de la sous-catégorisation peut expliquer certains schémas accentuels.

En effet, pour la série des mots en *law* dont nous avons discuté dans l'étude des entrées commençant par une majuscule, ces unités ont habituellement une accentuation simple car leur premier élément implique un contraste implicite, et elles entrent dans une série. Toutefois, *sod's law* n'est pas une loi au même titre que les autres, mais une loi qui s'applique aux malchanceux ; elle leur appartient. Nous retrouvons la même logique que pour *Achilles's heel* et *beginner's luck.*

Pour les entrées en *day*, nous pouvons proposer la même explication : *New Year's Day,* accentué avec un schéma « d » est une appellation, non une sous-catégorisation. Il relève certainement aussi d'une localisation temporelle. En revanche, *Mother's Day*, qui n'est pas un jour fixe du reste, comporte un contraste implicite avec *Father's Day* : ces deux journées ont été créées à l'intention des mères et des pères ; elles ont donc une fonction spéciale. Toutefois, *Mothering Sunday* a une accentuation double; nous retrouvons la logique de l'appellation.

L'étude des entrées avec *'s* permet de confirmer certaines des règles d'accentuation ainsi que notre hypothèse que l'opposition entre appellation et sous-catégorisation est pertinente pour expliquer les différents schémas accentuels. Pour clore cette étude, nous allons analyser les unités plus longues : celles qui ont aux moins trois éléments, et celles qui entrent dans une construction cumulative.

4.6.3.4 Les constructions cumulatives et les composés à plusieurs éléments

Dans la partie 5 de ce chapitre, nous avons étudié le comportement des extraits sonores en discours et nous avons soulevé le problème de l'accentuation des unités longues qui est influencée par leur placement dans l'ensemble. Deux phénomènes de discours ont été cités : le déplacement de l'accent et *Phrase Edge Prominence*. Ces deux principes permettent d'expliquer le fait que dans des unités longues, il y a un accent secondaire sur l'élément en début de l'unité, ce qui inverse parfois la hiérarchie accentuelle de cet élément. Cela ne remet pas en cause le placement de l'accent nucléaire qui est à la fin.

Sous cette rubrique, nous voulons comparer ce type d'unité qui est souvent formé pour l'occasion et des unités lexicalisées de notre base de données. Si nous prenons les extraits sonores du reportage sur le capuchon parlant, nous avons des constructions cumulatives: *talking bottle top, vodka bottle top, talking vodka bottle top*. En étudiant les mesures et les graphiques, nous nous apercevons que la proéminence principale est ou sur *bottle* ou sur *top* (plus souvent), mais jamais avant, sauf dans le cas où l'ensemble de l'unité subit une baisse graduelle[93]. Nous remarquons également que le début de l'unité a tendance à être plus proéminente que le milieu. Le tout début a souvent une valeur maximale du F_0. Ces corrélats acoustiques correspondent à l'accentuation double que nous retrouvons dans un grand nombre d'unités à éléments multiples.

---

[93] Voir 4.5.2.2. « L'interprétation du contour intonatif ».

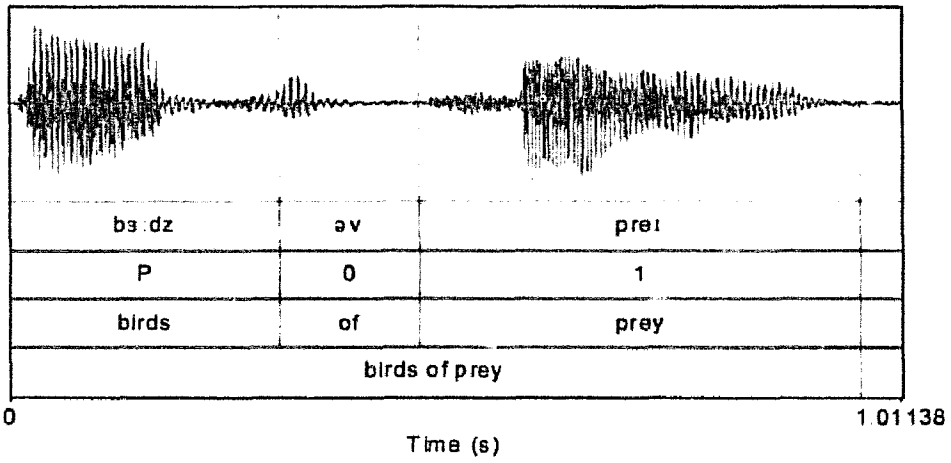
En ce qui concerne les unités lexicalisées de la base de données, il y a plusieurs centaines d'entrées avec au moins trois éléments et elles sont très majoritairement à accentuation double (85% de LPD-Cp et 75% de N-COMPACT). L'hypothèse que nous avons formulée selon laquelle la tendance à l'accentuation double augmente avec le nombre de syllabes peut s'étendre à ce genre de construction : la complexité croissante de la formation tend à augmenter le taux d'entrées à accentuation double. Cela correspond au rythme naturel de la langue, comme nous l'avons vu dans la partie 5.

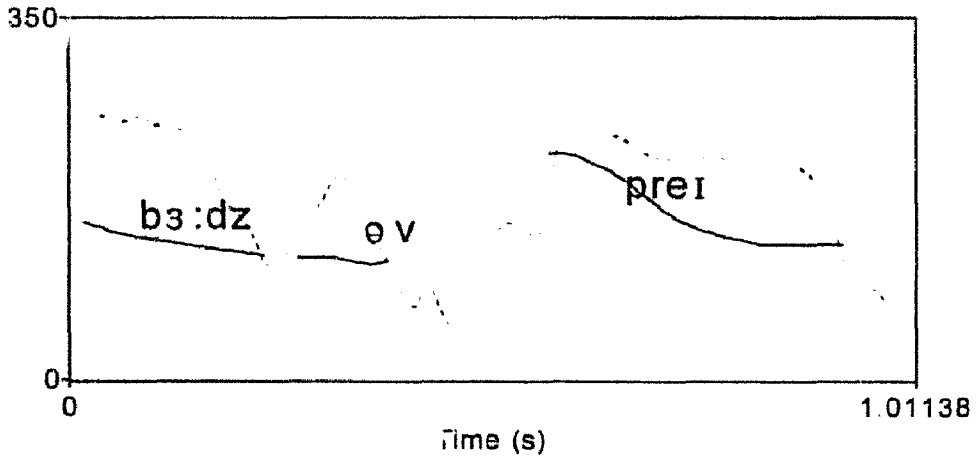
Certaines entrées sont des expressions lexicalisées du type *devil-may-care* ou *beggar-my-neighbour*, dont le schéma accentuel reflète simplement celle du syntagme. Afin de tirer quelques conclusions générales, nous avons étudié certains des éléments intermédiaires qui se répètent : *and, down, of, off, the, to, up*. Pour *off, up, down*, ces éléments font partie du sens de l'ensemble et ne mettent pas les autres éléments en relation. Ils ne correspondent donc pas à la logique que nous essayons de tester mais correspondent aux règles déjà proposées. Pour les autres éléments, ils jouent effectivement ce rôle de mise en relation.

La plupart de ces entrées ont une accentuation double. Pour les entrées en *and*, par exemple, il n'y a que 6 exceptions (sur 43) ; ce sont des unités à quatre éléments qui se terminent par un élément général (*Darby and Joan club*), l'accent primaire portant sur le troisième élément. Nous pouvons rapprocher ces entrées des composés co-ordonnées, et des points cardinaux dont le schéma correspond exactement au comportement accentuel des unités longues du discours : *north north west* /201/. Il y a une proéminence au début et à la fin.

Pour les 84 entrées en *of*, toutes ont un schéma « d » sauf *cost of living index* qui s'explique comme *Darby and Joan club* puisque *index* est très général. C'est *living* qui est mis en relief. De même, les entrées avec *the* et *to* ont toutes une accentuation double. La relation entre les éléments s'interprète comme étant simple. L'exemple lexicalisé du corpus oral, *birds of prey*, corrobore ces résultats, comme le montre les graphiques suivants :

Figure 4.34 : test\birdsof : Oscillographe et courbe de F_0 et I





La F_0 est bien plus élevée sur *prey*, ce qui montre que cette unité a une accentuation double, comme le prévoient les indications du dictionnaire.

Toutefois, elle comporte des caractéristiques particulières car la courbe sur *bird* est très plate. Ceci s'explique parce que *prey* porte non seulement l'accent nucléaire, mais il y a aussi une mise en relief de cet élément en raison de l'opposition entre *birds of prey* et *vulture fund*[94], comme le contexte nous indique «[...] the tactics of some of the vulture funds are threatening to give the birds of prey a bad name».

[94] *Vulture funds* correspond au quatrième extrait sonore relevé, donc à l'accentuation double, L'opposition sous-entendue entre le véritable oiseau et la compagnie dont le comportement est semblable explique peut-être ce schéma inattendu.

Conclusion

Nous avons essayé d'évaluer les règles de l'accentuation des composés ainsi que les indications sur l'accent lexical données par les dictionnaires. Notre méthode a consisté en une analyse des entrées de la base de données et des unités en discours du corpus oral. Les résultats de notre étude confirment que ces règles semblent bien expliquer la plupart des cas que nous avons rencontrés, et les indications de l'accent lexical paraissent fiables, mais nous nous sommes rendu compte de la nécessité d'intégrer les phénomènes du discours pour interpréter les résultats. La dernière rubrique nous a permis de faire le rapprochement entre l'accentuation de certaines unités lexicalisées et le comportement d'unités en discours. Compte tenu de ces constats, nous allons étudier l'opportunité d'un dictionnaire de composés dans notre conclusion.

ANNEXES
du
CHAPITRE 4

Annexe A : Requêtes 530

Annexe B : La structure des tables : statistiques à partir des requêtes 543

Annexe C : Analyse de l'accent lexical et le corpus oral 551

Annexe D : Analyse acoustique : tableaux récapitulatifs des mesures 557

Annexe E : Graphiques correspondant aux extraits sonores 564

Annexe F : Listes d'entrées extraites de la base de données 578

Annexe A : Requêtes

La première annexe du chapitre 4 reprend la programmation des requêtes qui ont été utilisées dans nos analyses ; les résultats font l'objet des parties 3 et 6 du chapitre 4. Les requêtes font partie de la base de données qui se trouve sur le CDROM (\base\COMPOSES00).

Comparaison entre tables : graphie/schéma

COMP-EPD-LPD

SELECT [EPD12-Cp].Graphie, [LPD90-Cp].Graphie

FROM [EPD12-Cp] INNER JOIN [LPD90-Cp] ON [EPD12-Cp].Graphie = [LPD90-Cp].Graphie;

Cette requête permet de mettre en relation deux tables selon le champ choisi, ici « Graphie ». Elle retourne toutes les entrées dont la graphie est commune entre les deux tables. Nous avons créé le même type de requête pour mettre en relation toutes les tables. Cet exemple nous retourne les 535 entrées communes entre EPD-Cp et LPD-Cp.

COMP-EPD-LPD-grx

SELECT [EPD-GRX].Graphie, [LPD-GRX].Graphie, [EPD-GRX].[Catégorie grammaticale], [EPD-GRX].[Type schéma], [LPD-GRX].[Type schéma], [LPD-GRX].Variation

FROM [EPD-GRX] INNER JOIN [LPD-GRX] ON [EPD-GRX].Graphie = [LPD-GRX].Graphie

WHERE ((([EPD-GRX].[Type schéma]) Like "s") AND (([LPD-GRX].[Type schéma]) Like "d"))

ORDER BY [EPD-GRX].Graphie;

Cette requête est semblable à la précédente ; elle crée un lien entre les tables -grx de EPD et LPD, mais en recherchant seulement les entrées avec un type de schéma {s} pour EPD et {d} pour LPD. Elle retourne 78 entrées.

comp-grx-epd-lpd (et ncomp)

SELECT [EPD-GRX].Graphie, [EPD-GRX].[Catégorie grammaticale], [EPD-GRX].[Type Cp], [EPD-GRX].[Type schéma], [LPD-GRX].Graphie, [LPD-GRX].[Type schéma], [LPD-GRX].[Graphie inverse], [N_COMP-GRX].[Type schéma], [N_COMP-GRX].Graphie, [N_COMP-GRX].[Element 1], [N_COMP-GRX].[Element inter], [N_COMP-GRX].[Element 2], [EPD-GRX].[Catégorie grammaticale], [LPD-GRX].Variation, [N_COMP-GRX].[Var schéma], [N_COMP-GRX].[Stress shift]

FROM ([EPD-GRX] LEFT JOIN [LPD-GRX] ON [EPD-GRX].Graphie = [LPD-GRX].Graphie) LEFT JOIN [N_COMP-GRX] ON [EPD-GRX].Graphie = [N_COMP-GRX].Graphie;

Cette requête est similaire mais le sens du lien entre les tables est orienté. Les résultats retournent toutes les entrées de EPD-grx et celles de LPD-grx et NCOMP-grx qui ont une graphie commune avec EPD. Elle sert surtout pour les requêtes suivantes.

comp-schema

SELECT [comp-grx-epd-lpd].[EPD-GRX].Graphie, [comp-grx-epd-lpd].[EPD-GRX].[Type schéma], [comp-grx-epd-lpd].[N_COMP-GRX].[Type schéma]

FROM [comp-grx-epd-lpd]

WHERE ((([comp-grx-epd-lpd].[EPD-GRX].[Type schéma]) Like "d") AND (([comp-grx-epd-lpd].[N_COMP-GRX].[Type schéma]) Like "s"));

A partir de la requête **COMP-EPD-LPD-grx**, nous recherchons les entrées de EPD-grx qui ont un schéma {d} et celles de NCOMP-grx qui ont un schéma {s}. C'est une autre manière d'obtenir les résultats de la requête **COMP-EPD-LPD-grx** (mais sur N-COMPACT et non LPD).

null-epd-lpd

SELECT [comp-grx-epd-lpd].[EPD-GRX].Graphie, [comp-grx-epd-lpd].[LPD-GRX].Graphie, [comp-grx-epd-lpd].[N_COMP-GRX].Graphie, [comp-grx-epd-lpd].[EPD-GRX].[Type schéma], [comp-grx-epd-lpd].[Catégorie grammaticale], [comp-grx-epd-lpd].[Type Cp]

FROM [comp-grx-epd-lpd]

WHERE ((([comp-grx-epd-lpd].[EPD-GRX].Graphie) Is Not Null) AND (([comp-grx-epd-lpd].[LPD-GRX].Graphie) Is Null) AND (([comp-grx-epd-lpd].[N_COMP-GRX].Graphie) Is Null))

ORDER BY [comp-grx-epd-lpd].[EPD-GRX].Graphie;

Enfin, cette requête, qui utilise le résultat de **COMP-EPD-LPD-grx**, permet de rechercher les entrées qui figurent dans EPD mais ni dans LPD ni dans N-COMPACT. Nous cherchons les champs vides de LPD-grx et de NCOMP-grx, c'est-à-dire les entrées de EPD qui ne correspondent pas à une entrée de ces tables. Elle retourne 1879 entrées.

Vérification des tables

DOUBLON-EPD

SELECT [EPD12-Cp].Graphie, Count([EPD12-Cp].Graphie) AS CompteDeGraphie

FROM [EPD12-Cp]

GROUP BY [EPD12-Cp].Graphie

HAVING (((Count([EPD12-Cp].Graphie))>=2));

Cette requête recherche les graphies qui figurent deux fois ou plus dans la table EPD-Cp, donc elle permet de repérer les doublons. Le même type de requête a été créé pour chacune des tables.

EPD-DOUBLE

SELECT [EPD12-Cp].Graphie, [EPD12-Cp].Phonie, [EPD12-Cp].[Catégorie grammaticale], [EPD12-Cp].[Nb syllabes], [EPD12-Cp].Schéma, [EPD12-Cp].[Type schéma], [EPD12-Cp].Source, [DOUBLON-EPD].CompteDeGraphie

FROM [EPD12-Cp] INNER JOIN [DOUBLON-EPD] ON [EPD12-Cp].Graphie = [DOUBLON-EPD].Graphie

WHERE ((([DOUBLON-EPD].CompteDeGraphie)>=2))

ORDER BY [EPD12-Cp].Graphie;

A partir de la requête précédente, cette requête permet de lister les doublons, c'est-à-dire les entrées repérées par **DOUBLON-EPD**. Le même type de requête a été créé pour chacune des tables.

Recherches sur le schéma accentuel

SYLL-SCHEMA

SELECT [LPD90-Cp].Graphie, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Schéma, [LPD90-Cp].[Type schéma], [LPD90-Cp].[Graphie inverse]

FROM [LPD90-Cp]

WHERE ((([LPD90-Cp].[Nb syllabes])="£2") AND (([LPD90-Cp].[Type schéma])="s"));

Cette requête a servi pour étudier la structure de chaque table (le même type de requête a été créé pour chacune) afin de voir la distribution des entrées d'après le nombre de syllabes et le type de schéma. Nous avons recherché successivement les entrées ayant un type de schéma {s} puis {d}. Le tri s'est fait selon le nombre de syllabes.

SHIFT-NC

SELECT N_COMPACT.[Type schéma], N_COMPACT.[Nb syllabes], N_COMPACT.Graphie, N_COMPACT.[Element 1], N_COMPACT.[Element inter], N_COMPACT.[Element 2], N_COMPACT.[Stress shift], N_COMPACT.[Schéma borné], N_COMPACT.[Var schéma]

FROM N_COMPACT

WHERE (((N_COMPACT.[Stress shift]) Like "*SS*"))

ORDER BY N_COMPACT.[Type schéma] DESC , N_COMPACT.[Nb syllabes], N_COMPACT.Graphie;

Cette requête, qui retourne 344 entrées, recherche les cas où le déplacement de l'accent est indiqué. Le codage est particulier à N-COMPACT, table saisie manuellement, donc {*SS*} dans le champ « Stress shift ».

SHIFT-LPD

SELECT [LPD90-Cp].[Type schéma], [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Graphie, [LPD90-Cp].[Catégorie grammaticale], [LPD90-Cp].Phonie, [LPD90-Cp].Variation, [LPD90-Cp].Schéma

FROM [LPD90-Cp]

WHERE ((([LPD90-Cp].Variation) Like "*<*" And ([LPD90-Cp].Variation) Not Like "*><*"))

ORDER BY [LPD90-Cp].[Type schéma] DESC , [LPD90-Cp].[Nb syllabes];

Cette requête est spécifique à LPD-Cp qui contient le code {<} dans le champ « Variation » pour indiquer le déplacement de l'accent. Toutefois, il faut prendre la précaution d'exclure les entrées codées {><} ce qui correspond au triangle avec le point d'exclamation de la version papier, et qui désigne une prononciation non recommandée. Les résultats sont triés selon le type de schéma, puis par nombre de syllabes. La requête retourne 1224 entrées, dont 26 en « s » (qui sont expliquées au chapitre 4). 9 entrées contiennent ces deux codages.

Recherches sur la graphie

Les recherches sur la graphie se sont faites surtout à partir de LPD car les trois graphies existent dans cette table. Les résultats ont parfois soulevé des problèmes de codage, voire des erreurs (voir 4.3.2.2.2, « Etude de la graphie : Structure des tables »). Nous avons donc créé des requêtes complémentaires qui permettent de vérifier les données.

tiret-LPD

SELECT [LPD90-Cp].Graphie, [LPD90-Cp].[Type schéma], [LPD90-Cp].[Catégorie grammaticale], [LPD90-Cp].Phonie, [LPD90-Cp].Variation, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Schéma, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].[Graphie inverse], [LPD90-Cp].[Type Cp], [LPD90-Cp].[N° Voy]

FROM [LPD90-Cp]

GROUP BY [LPD90-Cp].Graphie, [LPD90-Cp].[Type schéma], [LPD90-Cp].[Catégorie grammaticale], [LPD90-Cp].Phonie, [LPD90-Cp].Variation, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Schéma, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].[Graphie inverse], [LPD90-Cp].[Type Cp], [LPD90-Cp].[N° Voy]

HAVING ((([LPD90-Cp].Graphie) Like "*-*") AND (([LPD90-Cp].[Nb syllabes]) Like "£2"));

Ici, cette requête a permis de rechercher les entrées de 2 syllabes avec un trait d'union. Nous avons recherché successivement les entrées ayant 3, 4 ...9 syllabes avec un trait d'union. Nous avons également utilisé la requête de la même manière afin de rechercher l'espace.

var-LPD

SELECT [LPD90-Cp].Graphie, [LPD90-Cp].[Catégorie grammaticale], [LPD90-Cp].Variation, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Schéma, [LPD90-Cp].[Type schéma], [LPD90-Cp].[Graphie inverse], [LPD90-Cp].Critère, [LPD90-Cp].[N° Voy]

FROM [LPD90-Cp]

WHERE ((([LPD90-Cp].Graphie) Like "*,*"));

Cette requête permet d'isoler toutes les entrées avec une variante ; ces entrées étant signalées par la présence d'une virgule dans le champ de la graphie.

tiret-var-LPD

SELECT [LPD90-Cp].Graphie, [LPD90-Cp].Critère, [LPD90-Cp].[Catégorie grammaticale], [LPD90-Cp].Phonie, [LPD90-Cp].Variation, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Schéma, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].[Type schéma], [LPD90-Cp].[Graphie inverse], [LPD90-Cp].[Type Cp], [LPD90-Cp].[N° Voy]

FROM [LPD90-Cp]

GROUP BY [LPD90-Cp].Graphie, [LPD90-Cp].Critère, [LPD90-Cp].[Catégorie grammaticale], [LPD90-Cp].Phonie, [LPD90-Cp].Variation, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Schéma, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].[Type schéma], [LPD90-Cp].[Graphie inverse], [LPD90-Cp].[Type Cp], [LPD90-Cp].[N° Voy]

HAVING ((([LPD90-Cp].Graphie) Like "*-*" And ([LPD90-Cp].Graphie) Like "*,*"));

Cette requête ressemble à la précédente, mais elle ne recherche que les variantes où il y a un trait d'union.

verif-LPD

SELECT [LPD90-Cp].[Graphie inverse], [LPD-GRX].[Graphie inverse], [LPD90-Cp].Graphie, [LPD-GRX].Graphie

FROM [LPD90-Cp] INNER JOIN [LPD-GRX] ON [LPD90-Cp].Graphie = [LPD-GRX].Graphie;

verif-tiret

SELECT [verif-LPD].Graphie AS Expr1, [verif-LPD].[Catégorie grammaticale] AS Expr2, [verif-LPD].[LPD-GRX].[Graphie inverse], [verif-LPD].[Type schéma] AS Expr3

FROM [verif-LPD]

WHERE ((([verif-LPD].[LPD-GRX].[Graphie inverse]) Not Like "*-*") AND (([verif-LPD].[Graphie]) Like "*-*"));

La première de ces deux requêtes crée un lien entre LPD et LPD-grx, en intégrant les graphies et graphies inverses. Elle retourne les entrées ayant la même graphie dans les deux tables, donc logiquement toutes les entrées en compact de LPD-Cp. La deuxième requête utilise le lien créé entre LPD-grx et LPD par la première requête **verif-LPD** et recherche les graphies inverses qui n'auraient pas de trait d'union. S'il y a un trait d'union, alors il y a une erreur.

gr-NC

SELECT N_COMPACT.Graphie, N_COMPACT.[Element 1], N_COMPACT.[Element inter], N_COMPACT.[Element 2], N_COMPACT.[Nb syllabes], N_COMPACT.Schéma, N_COMPACT.[Schéma borné], N_COMPACT.[Type schéma], N_COMPACT.[Type schéma], N_COMPACT.[Var schéma], N_COMPACT.[Stress shift], N_COMPACT.Source

FROM N_COMPACT

GROUP BY N_COMPACT.Graphie, N_COMPACT.[Element 1], N_COMPACT.[Element inter], N_COMPACT.[Element 2], N_COMPACT.[Nb syllabes], N_COMPACT.Schéma, N_COMPACT.[Schéma borné], N_COMPACT.[Type schéma], N_COMPACT.[Type schéma], N_COMPACT.[Var schéma], N_COMPACT.[Stress shift], N_COMPACT.Source

HAVING (((N_COMPACT.[Element 1]) Like "*+*") AND (N_COMPACT.[Element inter]) Like "*+*"))

ORDER BY N_COMPACT.[Nb syllabes];

Cette requête est la seule qui ne concerne pas LPD-Cp pour la graphie. Elle permet de rechercher les entrées de N-COMPACT dont les premiers et seconds éléments sont écrits en compact car nous avons prévu lors de la saisie d'indiquer le type de séparateur dans les champs « Element 1 », « Element inter », « Element 2 » lorsque celui-ci était différent de l'espace. Le {+} représente la graphie en compact, le {-} la graphie avec un trait d'union.

Recherches sur la phonie

SCHWA-NPD

SELECT [NPD-Cp].Graphie, [NPD-Cp].[Catégorie grammaticale], [NPD-Cp].Phonie, [NPD-Cp].Variantes, [NPD-Cp].[Nb syllabes], [NPD-Cp].[Type schéma], [NPD-Cp].[Graphie inverse]

FROM [NPD-Cp]

WHERE ((([NPD-Cp].Phonie) Like "* *@*" And ([NPD-Cp].Phonie) Not Like "*.*@U*" And ([NPD-Cp].Phonie) Not Like "*.*I@*" And ([NPD-Cp].Phonie) Not Like "*.*e@*" And ([NPD-Cp].Phonie) Not Like "*.*U@*") AND (([NPD-Cp].[Type schéma])="s"))

ORDER BY [NPD-Cp].[Nb syllabes], [NPD-Cp].[Graphie inverse];

Cette requête permet de rechercher les entrées de deux syllabes étant accentuée {s} et comportant un schwa dans la deuxième syllabe. Comme NPD utilise le code SAMPA, schwa correspond à {@}, mais il faut également écarter les diphtongues centralisantes /ɪə/, /eə/, et /ʊə/ ainsi que /əʊ/. Nous avons ainsi pu vérifier la portée de l'hypothèse de la non-réduction des éléments monosyllabiques désaccentués des composés à accentuation simple.

SCHWA-EPD

SELECT [EPD12-Cp].Graphie, [EPD12-Cp].Phonie, [EPD12-Cp].[Catégorie grammaticale], [EPD12-Cp].[Nb syllabes], [EPD12-Cp].Schéma, [EPD12-Cp].[Type schéma]

FROM [EPD12-Cp]

WHERE ((([EPD12-Cp].Phonie) Like "*/*6*" And ([EPD12-Cp].Phonie) Not Like "*/*6.*" And ([EPD12-Cp].Phonie) Not Like "*/*U6*" And ([EPD12-Cp].Phonie) Not Like "*/*36*" And ([EPD12-Cp].Phonie) Not Like "*/*I6*") AND (([EPD12-Cp].[Nb syllabes]) Like "£2"));

Cette requête a les mêmes objectifs que la précédente mais comme EPD-Cp utilise un code particulier, schwa correspond à {6}, /ɪ/ à {I}, /e/ à {3} et /ʊ/ à {U}. Cette requête est spécifique à cette table.

Recherches sur la catégorie grammaticale

RECH-catg.-LPD

SELECT [LPD90-Cp].Graphie, [LPD90-Cp].[Catégorie grammaticale], [LPD90-Cp].Variation, [LPD90-Cp].[Type schéma], [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Schéma, [LPD90-Cp].[Graphie inverse], [LPD90-Cp].[Type Cp]

FROM [LPD90-Cp]

WHERE ((([LPD90-Cp].[Catégorie grammaticale]) Like "*Pn*"))

ORDER BY [LPD90-Cp].[Type schéma] DESC , [LPD90-Cp].[Nb syllabes];

Cette requête recherche la séquence {Pn} dans le champ « Catégorie grammaticale ». La séquence à chercher a été successivement {N}, {V} et ainsi de suite. Les résultats ont été triés selon le type de schéma (d'abord {s} puis {d}) puis par le nombre de syllabes.

Recherches à partir d'éléments

RECH-EL-LPD

SELECT [LPD90-Cp].[Type schéma], [LPD90-Cp].Graphie, [LPD90-Cp].[Catégorie grammaticale], [LPD90-Cp].Variation, [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Schéma, [LPD90-Cp].[Graphie inverse], [LPD90-Cp].[Type Cp]

FROM [LPD90-Cp]

WHERE ((([LPD90-Cp].Graphie) Like "*-*-*" And ([LPD90-Cp].Graphie) Not Like "*,*"))

ORDER BY [LPD90-Cp].[Type schéma] DESC , [LPD90-Cp].Graphie;

Cette requête permet de rechercher des unités avec au moins trois éléments, ayant deux traits d'unions. Nous en avons écarté les variantes graphiques qui peuvent comporter deux traits d'union, l'une en graphie principale, l'autre sur la variante. Le tri des entrées se fait selon le type de schéma, puis la graphie.

Cette requête a souvent été utilisée pour rechercher des éléments au début de l'entrée {element*} ou de la fin {*element espace } (car dans LPD le dernier

caractère du champ graphie est un espace). Le même type de requête a été créé pour NPD-Cp et EPD-Cp ainsi que N-COMPACT[1]. Pour cette dernière table, la requête peut être affinée en faisant la recherche sur le champ « Element 1 » ou « Element 2 » ; dans ce cas la séquence {*pan} ne génère que les entrées avec « pan » comme élément final, alors que pour les autres tables, nous avons également celles dont le dernier élément est « span ».

Recherches particulières

Majuscules

```
SELECT N_COMPACT.[Type schéma], N_COMPACT.Graphie,
N_COMPACT.[Nb syllabes], N_COMPACT.[Var schéma],
N_COMPACT.[Stress shift], N_COMPACT.[Schéma borné]

FROM N_COMPACT

WHERE (((N_COMPACT.Graphie) Like "z*") AND
((Comparer([Graphie],"Z"))=True))

ORDER BY N_COMPACT.[Type schéma], N_COMPACT.Graphie;
```

Cette recherche permet de lister les entrées qui commencent par une majuscule. Il faut inclure dans le champ graphie la lettre recherchée, puis, dans le champ « Comparer ([Graphie];],"Z") », la même lettre en majuscules. Sans ces précautions, les recherches ACCESS ne sont pas sensibles à la casse. Dans cet exemple, la lettre recherchée est {Z}.

---

[1] Cette requête a servi pour extraire les données qui se trouvent en Annexe F.

RECH-COMM-NC

SELECT N_COMPACT.[Type schéma], N_COMPACT.[Nb syllabes], N_COMPACT.Graphie, N_COMPACT.[Element 1], N_COMPACT.[Element inter], N_COMPACT.[Element 2], N_COMPACT.[Stress shift], N_COMPACT.[Schéma borné], N_COMPACT.[Var schéma]

FROM N_COMPACT

WHERE (((N_COMPACT.[Stress shift]) Like "*DS*"))

ORDER BY N_COMPACT.[Type schéma] DESC , N_COMPACT.[Nb syllabes], N_COMPACT.Graphie;

Il s'agit d'une requête très semblable à SHIFT-NC car le champ « Stress shift » sert de commentaire[2]. On peut également rechercher la séquence {DS}, comme ici, pour lister les entrées qui relèvent plutôt de syntagmes (discours) ou encore {!} pour lister les entrées qui sont apparues comme intéressantes à commenter lors de la saisie.

RECH-typeCp-LPD

SELECT [LPD90-Cp].[Type schéma], [LPD90-Cp].[Nb syllabes], [LPD90-Cp].Graphie, [LPD90-Cp].[Catégorie grammaticale], [LPD90-Cp].Phonie, [LPD90-Cp].Variation, [LPD90-Cp].[Type Cp], [LPD90-Cp].Schéma

FROM [LPD90-Cp]

WHERE ((([LPD90-Cp].[Type Cp]) Like "CR"))

ORDER BY [LPD90-Cp].[Type schéma] DESC , [LPD90-Cp].[Nb syllabes];

Cette requête permet de lister des composés dont le type est codé dans le champ « Type de Cp », ici {CR} ; les composés réduplicatifs.

---

[2] Cette requête a servi pour extraire les entrées qui se trouvent en Annexe F.

Annexe B: La structure des tables: statistiques à partir des requêtes

Etude du schéma accentuel: toutes les tables

Les chiffres correspondent au nombre d'entrées ayant un schéma « s » ou « d ». Le tableau donne le détail, syllabe par syllabe, en indiquant les pourcentages des entrées de n syllabes par rapport à l'ensemble du dictionnaire, et les pourcentages des entrées ayant un schéma donné par rapport aux entrées de n syllabes. L'accentuation simple est indiquée par « s », l'accentuation double par « d ».

Figure IV.1 : Type d'accentuation par table et par nombre de syllabes

Table / Type d'accentuation	EPD-Cp	NPD-Cp	LPD-Cp	N-COMPACT
Nb entrées	4221	4126	7154	6261
Entrées de 2 syllabes	1766	3374	3946	1256
Entrées de 2 syllabes en % du dictionnaire	41,84%	81,77%	55,16%	20,06%
Nb entrées «s» de 2 syllabes	1318	3237	3273	735
Entrées «s» de 2 syllabes en % du dictionnaire	31,22%	78,45%	45,75%	11,74%
% Entrées «s» des entrées à 2 syllabes	74,63%	95,94%	82,94%	58,52%
Nb entrées «d» de 2 syllabes	448	137	673	521
Entrées «d» de 2 syllabes en % du dictionnaire	10,61%	3,32%	9,41%	8,32%
% Entrées «d» des entrées à 2 syllabes	25,37%	4,06%	17,06%	41,48%
Entrées de 3 syllabes	1759	653	2459	2210
Entrées de 3 syllabes en % du dictionnaire	41,67%	15,83%	34,37%	35,30%
Nb entrées «s» de 3 syllabes	1258	597	1643	1226
Entrées «s» de 3 syllabes en % du dictionnaire	29,80%	14,47%	22,97%	19,58%
% Entrées «s» des entrées à 3 syllabes	71,52%	91,42%	66,82%	55,48%
Nb entrées «d» de 3 syllabes	501	56	816	984
Entrées «d» de 3 syllabes en % du dictionnaire	11,87%	1,36%	11,41%	15,72%
% Entrées «d» des entrées à 3 syllabes	28,48%	8,58%	33,18%	44,52%
Entrées de 4 syllabes	581	92	592	1609
Entrées de 4 syllabes en % du dictionnaire	13,76%	2,23%	8,28%	25,70%
Nb entrées «s» de 4 syllabes	377	69	249	729
Entrées «s» de 4 syllabes en % du dictionnaire	8,93%	1,67%	3,48%	11,64%
% Entrées «s» des entrées à 4 syllabes	64,89%	75,00%	42,06%	45,31%
Nb entrées «d» de 4 syllabes	204	23	343	880
Entrées «d» de 4 syllabes en % du dictionnaire	4,83%	0,56%	4,79%	14,06%
% Entrées «d» des entrées à 4 syllabes	35,11%	25,00%	57,94%	54,69%

Table Type d'accentuation	EPD-Cp	NPD-Cp	LPD-Cp	N-COMPACT
Entrées de 5 syllabes	93	7	120	677
Entrées de 5 syllabes en % du dictionnaire	2,20%	0,17%	1,68%	10,81%
Nb entrées «s» de 5 syllabes	42	4	21	224
Entrées «s» de 5 syllabes en % du dictionnaire	1,00%	0,10%	0,29%	3,58%
% Entrées «s» des entrées à 5 syllabes	45,16%	57,14%	17,50%	33,09%
Nb entrées «d» de 5 syllabes	51	3	99	453
Entrées «d» de 5 syllabes en % du dictionnaire	1,21%	0,07%	1,38%	7,24%
% Entrées «d» des entrées à 5 syllabes	54,84%	42,86%	82,50%	66,91%
Entrées de 6 syllabes	22	0	33	314
Entrées de 6 syllabes en % du dictionnaire	0,52%	0,00%	0,46%	5,02%
Nb entrées «s» de 6 syllabes	7	0	0	87
Entrées «s» de 6 syllabes en % du dictionnaire	0,17%	0,00%	0,00%	1,39%
% Entrées «s» des entrées à 6 syllabes	31,82%		0,00%	27,71%
Nb entrées «d» de 6 syllabes	15	0	33	227
Entrées «d» de 6 syllabes en % du dictionnaire	0,36%	0,00%	0,46%	3,63%
% Entrées «d» des entrées à 6 syllabes	68,18%		100,00%	72,29%
Entrées de 7 syllabes	0	0	3	119
Entrées de 7 syllabes en % du dictionnaire			0,04%	1,90%
Nb entrées «s» de 7 syllabes	0	0	0	29
Entrées «s» de 7 syllabes en % du dictionnaire			0,00%	0,46%
% Entrées «s» des entrées à 7 syllabes			0,00%	24,37%
Nb entrées «d» de 7 syllabes	0	0	3	90
Entrées «d» de 7 syllabes en % du dictionnaire			0,04%	1,44%
% Entrées «d» des entrées à 7 syllabes			100,00%	75,63%
Entrées de 8 syllabes	0	0	1	43
Entrées de 8 syllabes en % du dictionnaire			0,01%	0,69%
Nb entrées «s» de 8 syllabes	0	0	0	8
Entrées «s» de 8 syllabes en % du dictionnaire			0,00%	0,13%
% Entrées «s» des entrées à 8 syllabes			0,00%	18,60%
Nb entrées «d» de 8 syllabes	0	0	1	35
Entrées «d» de 8 syllabes en % du dictionnaire			0,01%	0,56%
% Entrées «d» des entrées à 8 syllabes			100,00%	81,40%
Entrées de 9 syllabes	0	0	0	12
Entrées de 9 syllabes en % du dictionnaire	0	0	0	0,19%
Nb entrées «s» de 9 syllabes	0	0	0	2
Entrées «s» de 9 syllabes en % du dictionnaire	0	0	0	0,03%
% Entrées «s» des entrées à 9 syllabes	0	0	0	16,67%
Nb entrées «d» de 9 syllabes	0	0	0	10
Entrées «d» de 9 syllabes en % du dictionnaire	0	0	0	0,16%
% Entrées «d» des entrées à 9 syllabes	0	0	0	83,33%

table type d'accentuation	EPD12	NPD	LPD90	N-COMP
Entrées de 10 syllabes	0	0	0	4
Entrées de 10 syllabes en % du dictionnaire	0	0	0	0
Nb entrées «s» de 10 syllabes	0	0	0	2
Entrées «s» de 10 syllabes en % du dictionnaire	0	0	0	0
% Entrées «s» des entrées à 10 syllabes	0	0	0	2
Nb entrées «d» de 10 syllabes	0	0	0	0
Entrées «d» de 10 syllabes en % du dictionnaire	0	0	0	2
% Entrées «d» des entrées à 10 syllabes	0	0	0	0
Entrées de 11 syllabes	0	0	0	0
Entrées de 11 syllabes en % du dictionnaire	0	0	0	0
Nb entrées «s» de 11 syllabes	0	0	0	?
Entrées «s» de 11 syllabes en % du dictionnaire	0	0	0	0
% Entrées «s» des entrées à 11 syllabes				
Nb entrées «d» de 11 syllabes				
Entrées «d» de 11 syllabes en % du dictionnaire				
% Entrées «d» des entrées à 11 syllabes				

Etude de la graphie : toutes les tables et variantes LPD

Le tableau donne le détail des variantes graphiques de LPD Pour les tables EPD-Cp, NPD-Cp et N-COMPACT, les chiffres correspondent au nombre total d'entrées ; chaque table correspond à un type de graphie. En revanche, pour LPD-Cp, les trois graphies existent, et certaines entrées donnent deux graphies (ou davantage). La graphie principale correspond donc à la première graphie indiquée, qui est également celle utilisée pour la graphie inverse ; le nombre de graphies principales correspond au nombre d'entrées. Le total des entrées correspondant à la graphie principale ou secondaire dépasse le nombre d'entrées car ces chiffres tiennent compte des variantes qui font partie de la même entrée. Nous constatons que LPD-Cp contient environ 30% d'entrées à trait d'union et 70% d'entrées sans séparateur ; le nombre d'entrées ayant une graphie avec espace est non significatif (0,62%).

Figure IV.2 : Informations sur les variantes graphiques de LPD

Table & Graphie	EPD-Cp trait d'union	NPD-Cp compact	LPD-Cp mixte	LPD-Cp compact	LPD-Cp trait d'union	LPD-Cp non compact	N-COMPACT non compact
Nb entrées	4221	4126	7154				6261
			Graphie principale ou secondaire	5045	2108	44	
			Graphie principale	5032	2086	36	
			Graphie secondaire	13	22	8	

Comparaison entre EPD12 et LPD/N-COMPACT: type de schéma

Le tableau récapitule le nombre d'entrées communes par type de schéma.

Figure IV.3: Schéma accentuel pour les entrées communes EPD/LPD/N-COMPACT

Tables / Type de schéma	EPD-grx	LPD-grx	NCOMP-grx
«s» accentuation simple	1081	1081	
«s» accentuation simple	493		493
«d» accentuation double	439	439	
«d» accentuation double	123		123

Etude de la corrélation entre le schéma accentuel et la graphie : LPD

Le tableau récapitule le type de schéma et le type de graphie par syllabe.

Figure IV.4 : Schéma accentuel et type de graphie, par syllabe : LPD-Cp

LPD90	compact	tiret	non compact
Toutes graphies	5045	2108	44
Graphie principale	5032	2086	36
Graphie secondaire	13	22	8
Entrées de 2 syllabes «s»	2829	444	4
% «s» des entrées de 2 syllabes	92,48%	50,06%	40,00%
Entrées de 2 syllabes «d»	230	443	6
% «d» des entrées de 2 syllabes	7,52%	49,94%	60,00%
Entrées de 3 syllabes «s»	1402	241	9
% «s» des entrées de 3 syllabes	83,01%	31,30%	60,00%
Entrées de 3 syllabes «d»	287	529	6
% «d» des entrées de 3 syllabes	16,99%	68,70%	40,00%
Entrées de 4 syllabes «s»	150	99	3
% «s» des entrées de 4 syllabes	65,22%	27,35%	21,43%
Entrées de 4 syllabes «d»	80	263	11
% «d» des entrées de 4 syllabes	34,78%	72,65%	78,57%
Entrées de 5 syllabes «s»	9	12	2
% «s» des entrées de 5 syllabes	17,31%	17,65%	50,00%
Entrées de 5 syllabes «d»	43	56	2
% «d» des entrées de 5 syllabes	82,69%	82,35%	50,00%
Entrées de 6 syllabes «s»	0	0	0
% «s» des entrées de 6 syllabes	0,00%	0,00%	0,00%
Entrées de 6 syllabes «d»	15	18	1
% «d» des entrées de 6 syllabes	100,00%	100,00%	100,00%
Entrées de 7 syllabes «s»	0	0	0
% «s» des entrées de 7 syllabes	0,00%	0,00%	
Entrées de 7 syllabes «d»	1	2	0
% «d» des entrées de 7 syllabes	100,00%	100,00%	

Etude de la corrélation entre le schéma accentuel et la catégorie grammaticale

Le tableau montre la distribution des entrées par catégorie syntaxique et par syllabe pour LPD-Cp, NPD-Cp et EPD-Cp. L'accentuation simple est indiquée par « s », l'accentuation double par « d ».

Figure IV.5 : Type de schéma par catégorie syntaxique et par syllabe : LPD-Cp, NPD-Cp, EPD-Cp

LPD Catégorie syntaxique /schéma	Total (% table/ catégorie syntaxique)	% £2	% £3	% £4	% £5	% £6	% £7	% £8
N «s»	473 (86,59%)	93,83%	83,27%	57,77%	23,53%	0,00%	0%	0%
N «d»	734 (13,41%)	6,17%	16,73%	42,23%	76,47%	100,00%	100%	100%
V «s»	287 (50,09%)	66,98%	29,95%	29,27%	0,00%	0,00%		
V «d»	286 (49,91%)	33,02%	70,05%	70,73%	100,00%	100,00%		
Q «s»	529 (34,64%)	42,80%	33,1[illegible]%	17,74%	8,33%	0,00%	0%	
Q «d»	998 (65,36%)	57,20%	66,89%	82,26%	91,67%	100,00%	100%	
Av «s»	69 (30,53%)	36,[illegible]5%	32,26%	0,00%	0,00%	0,00%		
Av «d»	157 (69,47%)	63,55%	67,74%	100,00%	100,00%	100,00%		
Pn «s»	19 (79,17%)	87,50%	90,91%	40,00%				
Pn «d»	5 (20,83%)	12,50%	9,09%	60,00%				
Ex «s»	8 (19,05%)	22,22%	12,50%	14,29%				
Ex «d»	34 (80,95%)	77,78%	87,50%	85,71%				
tdmk «s»	135 (88,82%)	92,96%	91,30%	54,55%	0,00%			
tdmk «d»	17 (11,18%)	7,04%	8,70%	45,45%	100,00%			
NPD Catégorie syntaxique /schéma	**Total (% table/ catégorie syntaxique)**	**% £2**	**% £3**	**% £4**	**% £5**	**% £6**	**% £7**	**% £8**
N «s»	3633 (97,27%)	97,72%	96,53%	86,30%	50,00%			
N «d»	102 (2,73%)	2,28%	3,47%	13,70%	50,00%			
V «s»	181 (84,98%)	84,82%	89,47%	66,67%				
V «d»	32 (15,02%)	15,18%	10,53%	33,33%				
Q «s»	238 (85,30%)	89,64%	80,56%	41,67%	100,00%			
Q «d»	41 (14,70%)	10,36%	19,44%	58,33%	0,00%			
Av «s»	62 (44,93%)	49,37%	44,00%	12,50%	0,00%			
Av «d»	76 (55,07%)	50,63%	56,00%	87,50%	100,00%			
Pn «s»	8 (32,00%)	33,33%	33,33%	25,00%				
Pn «d»	17 (68,00%)	66,67%	66,67%	75,00%				
Ex «s»	0 (0,00%)	100,00%						
Ex «d»	3 (100,00%)	0,00%						
EPD Catégorie syntaxique /schéma	**Total (% table/ catégorie syntaxique)**	**% £2**	**% £3**	**% £4**	**% £5**	**% £6**	**% £7**	**% £8**
N «s»	2737 (78,36%)	80,94%	78,71%	74,03%	54,93%	41,18%		
N «d»	756 (21,64%)	19,06%	21,29%	25,97%	45,07%	58,82%		
V «s»	48 (57,83%)	60,87%	58,33%	50,00%	0,00%			
V «d»	35 (2,17%)	39,13%	41,67%	50,00%	100,00%			
Q «s»	250 (36,93%)	45,39%	34,05%	28,30%	11,11%	0,00%		
Q «d»	427 (63,07%)	54,61%	65,95%	71,70%	88,89%	100,00%		
Av «s»	11 (13,92%)	18,18%	8,33%	11,76%	33,33%	0,00%		
Av «d»	68 (86,08%)	81,82%	91,67%	88,24%	66,67%	100,00%		

Etude de l'évolution du lexique

Le tableau récapitule le nombre d'entrées par table où une variante accentuelle ou un déplacement de l'accent est indiqué. L'accentuation simple est indiquée par « s », l'accentuation double par « d ».

Figure IV.6 : Influence des facteurs du déplacement de l'accent et la variation

Les entrées différentes : EPD12 «s» / LPD «d» **91 entrées**	**EPD12**	**LPD**	**EPD15**
Nb entrées indiquant une variante du schéma accentuel	20	31	10
% Entrées susceptibles de variation	21,98%	34,07%	10,99%
Nb entrées indiquant un déplacement de l'accent ou variation selon la fonction	4	62	51
% Entrées susceptibles d'un déplacement de l'accent	4,40%	68,13%	56,04%
Les entrées différentes : EPD12 «d» / LPD «s» **132 entrées**	**EPD12**	**LPD**	**EPD15**
Nb entrées indiquant une variante du schéma accentuel	71	21	11
% Entrées susceptibles de variation	53,79%	15,91%	8,33%
Nb entrées indiquant un déplacement de l'accent ou variation selon la fonction	8	0	0
% Entrées susceptibles d'un déplacement de l'accent	6,06%	0,00%	0,00%

Le tableau permet de comparer les graphies de EPD15 et de LPD pour les entrées ayant un schéma accentuel différent entre EPD12/LPD.

Figure IV.7 : Graphie comparée EPD15/LPD des entrées différentes EPD12/LPD

EPD12 «s» / LPD «d» (91 entrées)	**Graphies EPD15**	**LPD Graphie en compact**	**LPD Graphie avec trait d'union**	**LPD Graphie avec espace**
EPD15 Graphie en compact	4	1	1	2
EPD15 Graphie avec trait d'union	66	11	48	7
EPD15 Graphie avec espace	16	3	1	12
EPD15 Entrées absentes	5	0	4	1
EPD12 «d» / LPD «s» (132 entrées)	Graphies EPD15	LPD Graphie en compact	LPD Graphie avec trait d'union	LPD Graphie avec espace
EPD15 Graphie en compact	63	21	39	3
EPD15 Graphie avec trait d'union	17	0	12	5
EPD15 Graphie avec espace	43	2	3	38
EPD15 Entrées absentes	9	2	2	5

Annexe C : Analyse de l'accent lexical et le corpus oral

Le tableau suivant récapitule les indications des dictionnaires sur l'accent lexical (le cas échéant) des extraits sonores et des éléments entrant dans la composition de ces extraits.

La colonne « Graphie unité/élément » indique non seulement les extraits sonores mais aussi les éléments de la composition des extraits. Le mot *adult* y figure seul, dans l'unité *adult smoker*, et dans l'unité *adult x* qui reprend les indications sur les unités ayant ce mot comme premier élément. Seul *adult smoker*, l'extrait sonore, correspond aux schémas. Le classement s'est fait par ordre alphabétique sur cette colonne afin de faciliter les recherches. La colonne « Graphie dico » indique la graphie donnée par le dictionnaire – par défaut celle de LPD90 – pour les éléments lexicalisés. Puis les colonnes suivantes donnent le schéma accentuel effectivement indiqué par chaque dictionnaire : LPD90 (première édition de LPD), LPD00 (deuxième édition de LPD), EPD15 (15ème édition de EPD), et enfin EPD12 (12ème édition de EPD). Les quatre colonnes suivantes ne concernent que les extraits sonores. Le « Schéma perçu » indique le schéma sous la forme proposée à la fin du chapitre 1 (/1/ accent principal, /2/ accent secondaire, /3/ accent tertiaire, /P/ accent sur le premier élément du composé à accentuation double, /0/ syllabe inaccentuée ou désaccentuée). Puis les trois colonnes suivantes indiquent le type de schéma (1) effectivement perçu, (2) proposé par le dictionnaire, s'il y figure, (3) prévu par l'anglophone dans le cas des néologismes de la source littéraire. Enfin, la dernière colonne est un commentaire sur les entrées des dictionnaires en ce qui concerne l'accentuation ou la graphie dans le cas des éléments lexicalisés.

Figure IV.8 : Extraits sonores, accent perçu et accent lexical : indications des dictionnaires

Graphie unité / élément	Nb syll	Graphie dico LPD90	LPD 90	LPD 00	EPD 15	EPD 12	Schéma perçu	s/d perçu	s/d dico	s/d anglo	Commentaire dico accent
adult		adult	s/d	s/d	s	s					US pref d (LPD00)
adult smoker	4						P0#10	d			
adult x		adult education	abs	abs	d	abs					s sur adult
anti-Dark Arts spell	5						R0#P#1#0	s d		s	
Anti-Dark Force spell	5						R0#P#1#0	s d		d	
anyone	3	anyone	s	s	s	s	P0#0	s	s		
anything	3	anything	s	s	s	s	P0#0	s	s		
armchair	2	armchair	s/d	s/d	d/s	d/s	P#1	d	s/d		EPD12 note var position
armchairs	2	armchair	s/d	s/d	d/s	d/s	P#1	d	s/d		EPD12 note var position
armchairs	2	armchair	s/d	s/d	d/s	d/s	1#0	s	s/d		EPD12 note var position
armchairs	2	armchair	s/d	s/d	d/s	d/s	1#0	s	s/d		EPD12 note var position
autumn-sown oilseed rape	6	voir oilseed					P0#0#P0#1	d s	(s)		
bad debt	2	bad debt	abs	d	abs	abs	1#0	s	d		
bad debts	2	bad debt	abs	d	abs	abs	P#1	d	d		
bad x		bad +	d	d	d	abs					
birds of prey	3	bird of prey	d	d	d	abs	P#0#1	d	d		
bottle top		bottle top	abs	abs	abs	abs					
bottle x		bottle +	s								s sauf bottle green
broom		broom									u var u pref 92% GB (90), EPD12 u = normal, u = unstressed el of cp
broomstick	2						1#0	s		s	
broomsticks	2	broomstick	s	s	s	s	1#0	s	s		LPD90 u var u, EPD12 u var u
broomsticks	2	broomstick	s	s	s	s	1#0	s	s		LPD90 u var u, EPD12 u var u
Charms classroom	3	voir classroom					1#30	s s		(s)	
cigarette		cigarette	d	d	d	d					ts VAR sauf EPD12 85% GB (90), 65% US (00)
cigarette companies	6						P0#10	d			
cigarette x		cigarette holder	s/d	s/d	s	s					EPD15 US d (+ 1 ex sauf 12)

Graphie unité / élément	Nb syll	Graphie dico LPD90	LPD 90	LPD 00	EPD 15	EPD 12	Schéma perçu	s/d perçu	s/d dico	s/d anglo	Commentaire dico accent
classroom		classroom	s	s	ə	abs					LPD u var u, EPD u var u
climate change	3						P0#1	d			
climate change conference	6						P0#0#10	d s			
Dark ages	3						1#30	s			
Dark Arts	2						1#0	s		d	
Dark Arts	2						1#0	s		d	
dark cupboard	3						P#10	d		d	
Dark Forces	3						1#30	s		d	
Dark Magic	3						1#30	s		d	
Dark Magic	3						P#10	d		d	
Dark Side	2						P#1	d		s	
Dark Side	2						P#1	d		s	
Dark Side	2						1#0	s		s	
dark wizard	3						P#10	d		d	
dark wizard	3						1#30	s		s	
dark wizard Grindelwald	6						P#30#100	d s		d	
dark wizards or witches	6						P#P0#0#10	d		d	
Defence Against the Dark Arts	7						0P#0P#0#P#1	d d		d	
Defence Against the Dark Arts	7						0P#0P#0#P#1	d d		d	
Defence Against the Dark Arts	7						0P#0P#0#P#1	d d		d	
distressed country	4						0P#10	d			
distressed investing	5						0P#010	d			
distressed Peruvian debt	7						0P#0P0#1	d d			
distressed securities	6						0P#0100	d			
distressed x		distressed +	abs	abs	abs	abs					
dragon bite	3						10#0	s		s	
dragon dealer	4						10#30	s		s	
dragon egg	3						10#0	s		s	
dragon egg	3						10#0	s		s	
dragon eggs	3						10#0	s		s	
dragon heartstring	4	voir heartstrings					10#3#0	s		s (s)	

Graphie unité / élément	Nb syll	Graphie dico LPD90	LPD 90	LPD 00	EPD 15	EPD 12	Schéma perçu	s/d perçu	s/d dico	s/d anglo	Commentaire dico accent
Dragon liver	4						10#30	s		s	
dragon x		dragonfly	s	s	s	s					EPD12 -
Dragon-breeding	4						10#30	s		s	
Dragon-breeding	4						P0#10	d		s	
dragon's blood	3	dragon's blood	abs	abs	s/d	s	10#0	s	s		EPD12 -
dragon's blood	3	dragon's blood	abs	abs	s/d	s	10#0	s	s		EPD12 -
dragon's head	3						P0#1	d		d	
emissions trading	5						0P0#10	d			
fifteen or sixteen minutes	7						P#0#0#P#0#10	d s			
flying-key room	4	voir room					10#0#0	s s		s d	
genetically modified autumn-sown oilseed rape	14	voir oilseed					0P00#P00#P0#0#P0#1	d s	(s)		
genetically modified crops	7						P00#P00#1	d			
greenhouse		greenhouse	s	s	s	s					
greenhouse gas emissions	6	voir greenhouse					P#0#1#030	s d s			
greenhouse gases	4	greenhouse gas	abs	abs	d	abs	P#0#10	d s	d (s)		
greenhouse x		greenhouse effect	s	s	s	abs					
ground-breaking		groundbreaking	abs	abs	s	abs					
ground-breaking scheme	4	voir groundbreaking					P#30#1	d s	(s)		
hard-earned		hard-earned	abs	d	d	d					LPD00 EPD15&12 SS
hard-earned cash	3	hard-earned cash	abs	abs	d	abs	P#P#1	d d	d (d+SS)		s sur hard-earned
heartstrings		heartstrings	s	s	s	s					seul EPD12 non pl
hocus pocus	4	hocus-pocus	d	d	d	d	P0#10	d	d		
keyhole	2	keyhole	s	s	s	s	1#0	s	s		EPD12 -
kitchen sink	3	kitchen-sink	d	d	d	abs	P0#1	d	d		ts SS
mass production	4	mass production	d	d	d/s	s/d	P#010	d	d		

Graphie unité / élément	Nb syll	Graphie dico LPD90	LPD 90	LPD 00	EPD 15	EPD 12	Schéma perçu	s/d perçu	s/d dico	s/d anglo	Commentaire dico accent
microchip		microchip	s	s	s	abs					
microchip technology	7	voir microchip	abs	abs	abs	abs	P0#0#0100	d s	(s)		
mirror room	3	voir room					10#0	s		s	
modified crops	4						P00#1	d			
mountain bike	3						10#0	s			
New York	2	New York	d	d	d	abs	P#1	d	d		LPD90&00 EPD15 SS
New York company	5						P0#10	d s	(d+SS)		
New York investors	5						P0#010	d s	(d+SS)		
New York x		New York City	d	d	d	abs					ts s sur New York
non-smoker	3	nonsmoker	d	d	d	abs	P#10	d	d		
oilseed		oilseed	abs	s	s	abs					
pillar box	3	pillar box	s	s	s	s	10#0	s	s		EPD15, -
portrait hole	3						10#0	s		s	
portrait hole	3						10#0	s		s	
portrait hole	3						10#0	s		s	
portrait hole	3						10#0	s		s	
publicity-shy		publicity-shy	abs	abs	abs	abs					
publicity-shy Elliots	8						0P0#0#10	d s			
race x		race +	s	s	s	s					s exc race relations
racing		racing	s	s	s	s					
racing bike	3						10#0	s		s	
racing brooms	3						10#0	s		s	
racing x		racing +	abs	abs	abs	abs					
razor blade	3	razor blade	abs	s	s	s	10#0	s	s		
restructured Peruvian debt	8						0P0#0P0#1	d d			
room		room									u var u, LPD pref 82% GB (90) 93% US (00) mais poss u dans cp
school broom	2						P#1?	d		d	
school owls	2						P#1	d		d	
self x		self-	d	d	d	d					
self-stirring	3	self-stirring	abs	abs	abs	abs	1#30	s		d	
self-stirring cauldrons	5						P#30#10	d s		d	
spacecraft	2	spacecraft	s	s	s	abs	1#0	s	s		

Graphie unité / élément	Nb syll	Graphie dico LPD90	LPD 90	LPD 00	EPD 15	EPD 12	Schéma perçu	s/d perçu	s/d dico	r/d anglo	Commentaire dico accent
spring-sown genetically modified oilseed rape	13	see oilseed					P#0#0P00#P00#P0#1	d s	(s)		
staff room	2	voir room					1#0	s		s	
talking bottle top	5	see talking & bottle					P0#P0#1	d d d			
talking bottle top	5	see talking & bottle					P0#10#0	s s d			
talking vodka bottletop	7	see talking & bottle					P0#P0#P0#1	d d d d			
talking x		talking +	abs	s&d	s&d	d					EPD12 ex talking-film!
thirty x		thirty-nine	d	d	abs	abs					LPD90&00 SS
thirty-sixth		thirty-sixth	abs	abs	abs	abs					
thirty-sixth floor	4						P0#1#0	s d			
timetable	3	timetable	s	s	s	s	1#30	s	s		
tower room	3	see room					10#0	s		s	
trophy room	2	see room					10#0	s		s	
trophy room	2	see room					10#0	s		s	
trophy room	2	see room					10#0	s		s	
under x		under +				d					
under-age	3	underage	d	d	d	abs	10#0	s	d		LPD90&00 EPD15 SS & -
under-age smokers	5	see underage					P0#0#10	d s	(d+SS)		
under-age s ing	5	see underage					P0#0#10	d s	(d+SS)		
under-age x		underage drinking	abs	d	abs	abs					LPD00, SS
vodka bottle top	5						P0#P0#1	d d d			
vodka genie	4						P0#)	d			
vodka genie	4						10#30	s			
vodka genie	4						10#30	s			
vulture fund	3						10#0	s			
vulture fund	3						10#0	s			
vulture fund	3						10#0	s			
vulture funds	3						P0#1	d			
washrooms	2	washroom	s	s	s	abs	1#0	s	s		LPD90&00 u var u, EPD15 u var u
workshop	2	workshop	s	s	s	s	1#0	s	s		

Annexe D: Analyse acoustique: tableaux récapitulatifs des mesures

Les tableaux suivants récapitulent les mesures prises pour certains des extraits sonores: ceux du reportage sur le capuchon parlant, dans le répertoire \vodka; et ceux du reportage des essais des OGM dans le répertoire \oilseed. Nous avons scindé les résultats afin qu'ils soient plus lisibles. Ainsi, il y a un premier tableau qui récapitule les résultats par syllabe, comme pour l'échantillonage des extraits en *Jark*. Les mesures par unité (mot, composé, extrait entier) font l'objet d'un deuxième tableau.

Les temps sont tous donnés en secondes (s): le début (T-début), la fin (T-fin), la durée (Durée).

Les mesures de F_0 sont en hertz (Hz), en nombres entiers pour les maxima et minima (F_0Max et F_0Min), et en demitons (dt), avec deux décimales pour la moyenne (F_0Mean) et l'écart type (F_0SD); la pente de F_0 (F_0MAS) est en demitons/seconde (dt/s).

Les mesures de l'intensité sont en décibels (dB), en nombres entiers pour les maxima et minima (Imax et Imin) et avec deux décimales pour la moyenne (I-mean) et l'écart type (I-SD).

Le symbole « # » indique que la mesure n'a pas pu être prise, ou parce que le segment est trop bref (intensité) ou parce que la valeur n'est pas comprise dans l'échelle spécifiée.

Dans la colonne « Phon-syll », la phonie est donnée en utilisant le codage spécifique à PRAAT qui permet d'incorporer les étiquettes sur les graphiques en IPA (voir Annexe B du chapitre 3).

Là où le schéma est indiqué, les mesures pour la syllabe correspondant à l'accent primaire /1/ sont en caractères gras.

Figure IV.9 : Le répertoire\oilseed : mesures pour les extraits sonores, par syllabe

Fichier son	Schéma	Phon-syll	T-début	T-fin	Durée	F_0Max	F_0Min	Imax	Imin	F_0-MAS	F_0-Mean	F_0-SD	I-Mean	I-SD
oilseed1	P	\ct:t	0,000	0,230	0,230	141	117	78	60	29,19	4,95	1,01	73,00	6,62
	0	m	0,230	0,368	0,138	118	112	76	74	11,08	2,47	0,37	75,15	0,61
	0	s\sw\hsn	0,368	0,632	0,264	137	122	79	59	18,02	4,13	0,66	73,68	6,14
	P	\ct\icl	0,632	0,876	0,244	145	119	79	70	17,45	5,09	1,22	76,45	2,49
	0	si:d	0,876	1,095	0,219	127	117	76	58	21,50	3,22	0,50	68,19	6,06
	1	re\icp	1,095	1,336	0,241	176	136	81	61	25,33	7,49	1,68	76,00	5,71
oilseed3	0	d\zh\sw	0,000	0,107	0,107	137	134	78	69	13,11	5,28	0,17	75,59	3,90
	P	net	0,107	0,270	0,162	130	115	78	62	22,29	3,72	0,72	72,72	5,50
	0	\ick	0,270	0,372	0,103	134	124	72	64	24,11	4,28	0,55	67,38	3,63
	0	\li	0,372	0,429	0,057	111	111	#	#	1,49	1,86	0,01	#	#
	P	m\abd	0,429	0,599	0,170	127	108	77	71	21,36	2,86	0,86	75,90	2,07
	0	\ic	0,599	0,662	0,062	120	116	#	#	40,02	2,91	0,31	#	#
	0	fa\icd	0,662	0,945	0,284	148	126	80	57	21,27	6,02	0,79	74,25	7,06
	P	\ct:t	0,945	1,182	0,236	139	114	79	64	36,94	4,23	1,16	73,10	5,35
	0	m	1,182	1,325	0,143	131	118	76	74	18,34	3,95	0,64	75,26	0,75
	0	s\sw\hsn	1,325	1,603	0,278	134	120	79	57	12,03	3,90	0,67	73,71	6,41
	P	\ct\icl	1,603	1,882	0,278	139	114	78	67	16,18	4,35	1,09	75,35	3,10
	0	si:d	1,882	2,084	0,203	132	120	77	60	15,13	3,73	0,60	71,17	5,88
	1	re\icp	2,084	2,408	0,324	175	133	81	55	25,57	7,40	1,73	71,36	10,27
farmtri2	P	fas:m	0,000	0,344	0,344	161	130	81	55	16,68	5,93	1,25	73,36	9,10
	0	tra\iclz	0,344	0,697	0,353	143	128	79	62	16,77	5,53	0,49	73,65	6,19
	0	\swv	0,697	0,809	0,112	117	100	72	64	44,99	1,36	1,01	67,49	3,17
	0	d\zh\sw	0,809	0,884	0,075	#	#	66	62	#	#	#	63,94	2,92
	P	net	0,884	1,075	0,191	136	123	77	59	26,02	4,59	0,57	73,25	6,02
	0	\ick	1,075	1,185	0,110	127	121	66	57	21,22	3,68	0,34	61,19	3,74
	0	li	1,185	1,248	0,063	120	117	#	#	39,31	2,83	0,23	#	#
	P	m\abd	1,248	1,409	0,161	125	112	78	72	15,06	3,25	0,62	76,59	1,92
	0	\ic	1,409	1,465	0,056	99	98	#	#	4,89	-0,26	0,03	#	#
	0	fa\icd	1,465	1,717	0,252	125	114	78	56	12,74	3,45	0,55	71,54	7,37
	1	kr\abps	1,717	2,080	0,363	166	121	82	54	49,34	5,94	1,87	66,43	10,13
modif1	P	m\abd	0,000	0,153	0,153	151	123	80	75	33,26	5,47	1,20	78,48	1,54
	0	\sw	0,153	0,254	0,101	127	109	77	70	44,56	2,93	1,00	73,61	3,01
	0	fa\icd	0,254	0,464	0,210	128	113	78	71	15,70	3,27	0,69	75,45	2,13
	1	kr\abps	0,464	0,815	0,351	176	117	81	55	64,20	6,17	2,44	67,50	8,94
oilseed2	PP	spr\ic\ng	0,000	0,419	0,419	229	133	83	54	51,75	10,72	3,19	71,49	10,76
	0	s\sw\hsn	0,419	0,766	0,347	142	102	76	55	44,21	2,12	1,78	69,10	6,70
	0	d\dh\sw	0,766	0,919	0,153	175	130	77	63	46,09	7,55	2,54	67,78	4,88
	P	net	0,919	1,087	0,168	117	110	76	66	9,71	2,22	0,32	73,39	2,95
	0	\ick	1,087	1,198	0,111	122	108	74	59	42,86	2,36	0,82	66,61	5,86
	0	li	1,198	1,257	0,060	100	100	#	#	1,64	-0,03	0,01	#	#
	P	m\abd	1,257	1,425	0,168	118	110	78	71	10,18	2,51	0,36	75,90	2,00
	0	i	1,425	1,479	0,054	91	90	#	#	8,09	-1,73	0,06	#	#
	0	fa\icd	1,479	1,732	0,253	122	113	79	58	10,17	2,93	0,46	72,95	6,38
	P	\ct\icl	1,732	1,987	0,256	128	107	78	68	14,68	3,05	0,96	75,77	2,77
	0	si:d	1,987	2,209	0,222	121	98	76	58	26,71	1,29	1,24	70,57	5,28
	1	re\icp	2,209	2,531	0,322	167	140	81	56	16,85	7,30	1,16	71,26	9,86

Figure IV.10 : Le répertoire\oilseed : mesures pour les extraits sonores, par unité

Fichier son	Unité	Durée en ms	F_0-MAS	F_0-Mean	F_0-SD	I-Mean	I-SD
oilseed1	extract: autumn-sown oilseed rape	1,336	25,05	4,77	1,89	71,87	6,81
	unit: autumn-sown	0,632	20,60	3,86	1,27	72,01	6,60
	unit: oilseed rape	0,704	28,56	5,47	2,07	72,44	6,93
	mot: oilseed	0,463	19,86	4,26	1,42	71,30	7,01
	mot: autumn	0,368	19,62	3,98	1,44	72,62	5,90
oilseed3	extract: genetically modified autumn-sown oilseed rape	2,408	23,75	4,59	1,56	71,61	7,14
	unit: genetically-modified	0,945	23,01	4,42	1,26	72,96	6,17
	unit: autumn-sown	0,658	20,66	4,02	0,90	71,96	6,59
	unit: oilseed rape	0,805	27,1:	5,37	1,98	70,89	8,20
	mot: genetically	0,429	22,95	4,44	0,73	71,97	6,02
	mot: modified	0,516	24,83	4,36	1,65	73 25	6,45
	mot: autumn	0,379	24,90	4,24	0,94	72,51	5,43
	mot: oilseed	0,481	17,25	4,06	1 04	71,71	6,82
farmtri2	extract: farm trials of genetically modified crops	2,080	20,59	4,38	1,53	69,94	8,15
	unit: farm trials						
	unit: genetically modified	0,834	16,96	3,52	0,86	70,60	7,04
	unit: genetically-modified crops	1,197	24,35	3,90	1,39	68,93	8,32
	mot: genetically	0,439	52,69	4,12	1,15	68,82	7,10
	mot: modified	0,469	42,58	3,16	0,98	71,67	6,68
modif1	extract/unit: modified crops	0,815	32,12	4,54	1,90	69,84	8,38
	mot: modified	0,464	21,83	3,98	1,33	73,16	6,26
oilseed2	extract: spring-sown genetically modi ilseed rape	2 531	27,80	4,40	3,47	70,79	7,55
	unit: spring-sown	0,766	42,95	6,44	5,02	70,02	9,02
	unit: genetically modified	0,966	17,20	2,92	1,54	71,35	5,92
	unit: oilseed rape	0,800	25,67	4,12	2,50	70,80	8,30
	mot: genetically	0,492	25,47	3,36	2,11	69,99	5,88
	mot: modified	0,474	12,19	2,49	0,66	72,13	6,00
	mot: oilseed	0 477	16,12	2,35	1,39	71,29	7,01

Figure IV.11 : Répertoire \vodka : mesures pour les extraits sonores, par syllabe

Fichier son	Schéma	Phon-syll	T-début	T-fin	Durée	F_0Max	F_0Min	Imax	Imin	F_0-MAS	F_0-Mean	F_0-SD	I-Mean	I-SD
Armchai1	P	\as:m	0,000	0,274	0,274	109	98	82	72	12,81	0,94	0,46	78,24	3,44
	1	t\she\swz	0,274	0,662	0,388	155	120	79	61	27,65	6,09	1,32	72,73	6,67
HP-\armcha3	P	as:m	0,000	0,288	0,288	192	169	81	67	15,96	10,43	0,86	74,87	4,62
	1	t\she\sw	0,288	0,619	0,331	208	109	78	63	60,26	10,09	3,90	74,63	4,17
HP-\armcha1	1	\as:m	0,024	0,315	0,291	145	104	79	66	26,06	5,42	1,65	73,59	4,80
	0	t\she\swz	0,315	0,711	0,396	128	104	78	61	13,65	2,48	0,84	73,55	5,13
HP-armcha2	1	\as:m	0,039	0,358	0,319	121	99	84	66	21,93	1,75	0,95	76,04	6,56
	0	t\she\swz	0,358	0,682	0,324	108	75	75	61	28,81	-3,21	1,97	69,12	3,75
Groundbl	P	gra\hsnd	0,000	0,397	0,397	191	133	87	58	27,59	9,83	1,76	78,62	8,45
	3	bre\ick	0,397	0,660	0,264	137	116	83	55	19,15	4,49	0,93	73,37	10,62
	0	\ic\ng	0,660	0,778	0,117	137	127	78	76	19,11	4,90	0,49	77,13	0,93
	1	ski:m	0,778	1,240	0,462	198	102	78	56	45,95	7,22	4,51	71,78	6,77
Kitchen1	P	k\ict\s\	0,000	0,262	0,262	138	132	80	56	35,70	5,17	0,36	67,24	7,36
	0	\swn	0,262	0,355	0,093	144	140	79	78	10,02	6,16	0,20	78,34	0,51
	1	s\ic\ngk	0,355	0,724	0,369	217	126	87	63	55,39	9,82	3,04	76,07	8,23
Massprol	P	m\aes	0,041	0,237	0,196	177	171	76	60	11,81	9,61	0,20	67,87	6,16
	0	pr\sw	0,237	0,383	0,146	222	200	80	55	36,63	12,99	0,71	69,21	10,49
	1	d\vtk	0,383	0,554	0,170	166	152	83	58	15,49	7,69	0,50	70,63	10,40
	0	\shn	0,554	0,675	0,122	#	#-	67	58	#-	#-	#-	63,56	3,76
Microch1	P	ma\ick	0,000	0,340	0,340	178	146	83	56	28,39	8,76	1,07	72,55	8,96
	0	r\sw\hs	0,340	0,466	0,126	140	130	84	80	15,19	5,10	0,43	82,36	1,32
	0	t\sh\cp	0,466	0,718	0,252	126	112	79	56	28,95	2,92	0,74	70,96	8,46
	0	tek	0,718	0,970	0,252	131	118	82	55	31,82	3,77	0,71	67,46	10,72
	1	n\abl	0,970	1,123	0,153	147	142	84	78	7,44	6,51	0,19	81,75	2,03
	0	\swd\zh	1,123	1,279	0,157	138	122	82	63	42,59	4,48	0,82	73,71	7,04
	0	i	1,279	1,500	0,221	126	102	80	55	36,36	1,98	1,32	68,40	10,07
Talkbot2	P	t\ct k	0,000	0,225	0,225	249	179	83	58	47,28	14,09	1,90	74,18	8,45
	0	\ic\ng	0,225	0,423	0,198	169	131	81	60	29,87	5,84	1,53	71,26	5,97
	1	b\abt	0,423	0,614	0,190	221	195	86	60	22,13	13,04	0,78	77,00	9,87
	0	\swl	0,614	0,749	0,135	157	122	83	73	48,78	5,84	1,38	80,79	3,91
	0	t\abp	0,749	0,984	0,235	154	122	83	67	39,98	5,28	1,54	78,14	4,95
Talkbot3	P	t\ct:k	0,000	0,293	0,293	129	109	80	58	36,25	3,16	1,03	68,43	7,98
	0	\ic\ng	0,293	0,536	0,243	115	104	74	57	12,56	1,23	0,55	64,47	5,01
	P	b\abt	0,536	0,724	0,187	189	163	86	59	45,44	9,89	0,86	73,88	10,51
	0	\swl	0,724	0,855	0,131	128	101	78	74	45,49	1,86	1,41	76,90	1,62
	1	t\abp	0,855	1,225	0,371	172	133	86	56	25,02	7,95	1,55	69,77	10,13
Talkbot1	P	t\ct:k	0,000	0,313	0,313	187	152	86	54	28,97	10,01	1,19	73,22	11,26
	0	\ic\ng	0,313	0,514	0,201	191	161	83	73	18,85	10,26	0,88	76,38	3,09
	P	v\abd	0,514	0,736	0,222	149	117	86	62	34,81	5,66	1,35	78,06	8,91
	0	k\sw	0,736	0,933	0,197	131	102	84	59	47,32	2,44	1,56	71,75	9,29
	P	b\abt	0,933	1,129	0,197	124	99	86	60	35,11	2,15	1,16	75,84	9,56
	0	\swl	1,129	1,305	0,176	118	106	78	69	16,69	1,78	0,62	74,64	3,70
	1	t\abp	1,305	1,644	0,338	123	107	83	64	20,13	2,08	0,89	75,20	6,43

Fichier son	Schéma	Phon-syll	T-début	T-fin	Durée	F_0Max	F_0Min	Imax	Imin	F_0-MAS	F_0-Mean	F_0-SD	I-Mean	I-SD
Vodbott1	P	v\abd	0,000	0,215	0,215	304	256	88	66	21,08	18,49	0,89	81,83	7,56
	0	k\sw	0,215	0,372	0,157	292	270	88	60	22,58	17,92	0,51	76,06	11,49
	P	b\abt	0,372	0,593	0,221	268	199	87	64	34,69	14,60	1,64	78,17	8,72
	0	\swl	0,593	0,731	0,137	197	174	82	69	24,07	10,55	0,80	77,86	5,04
	1	t\abp	0,731	1,100	0,369	231	191	85	56	30,44	12,69	1,31	71,53	11,66
Vodgen3	P	v\abd	0,000	0,201	0,201	127	115	85	63	19,00	3,34	0,60	75,68	8,07
	0	k\sw	0,201	0,368	0,167	126	111	80	56	35,58	2,88	0,79	70,35	8,85
	1	d\zhi:n	0,368	0,648	0,280	128	113	80	69	13,61	3,22	0,68	76,19	2,69
	0	i	0,648	0,888	0,239	108	92	78	55	23,27	0,25	0,93	66,78	9,12
Vodgen11		v\abd	0,000	0,171	0,171	210	186	87	73	32,43	11,86	0,77	83,99	4,19
	0	k\sw	0,171	0,364	0,193	175	152	88	61	40,99	8,56	0,92	73,86	11,18
	3	d\zhi:n	0,364	0,654	0,291	138	113	82	64	22,52	4,05	1,25	77,61	4,85
	0	i	0,654	0,791	0,137	103	90	79	73	25,94	-0,63	0,84	77,02	2,30
Vodgen12	1	v\abd	0,000	0,188	0,188	137	119	85	58	30,60	4,54	0,91	75,41	9,94
	0	k\sw	0,188	0,283	0,095	132	122	78	72	34,42	4,14	0,55	75,87	2,80
	3	d\zhi:n	0,283	0,496	0,212	125	107	77	61	32,42	2,70	0,95	71,66	5,55
	0	i	0,496	0,560	0,064	107	104	?	?	39,28	0,94	0,23	?	?
workshol	1	w\er:k	0,000	0,316	0,316	178	152	86	57	16,14	9,06	1,07	75,57	10,33
	t	\sh\ab\?g	0,316	0,574	0,259	133	115	81	58	28,27	3,58	0,87	74,96	6,03

Figure IV.12 : Répertoire \vodka : mesures pour les extraits sonores, par unité

Unités	Unités	Durée	F_0-MAS	F_0-Mean	F_0-SD	I-Mean	I-SD
Armchai1	extract/unit: armchairs	0,662	25,57	3,39	2,74	73,70	7,00
HP-\armcha3	extract/unit: armchair	0,619	37,66	10,11	2,47	73,38	5,78
HP-\armcha1	extract/unit: armchairs	0,686	15,41	3,87	1,87	72,23	6,46
HP- armcha2	extract/unit: armchairs	0,643	20,10	-1,23	2,87	71,38	7,10
Groundb1	extract: ground-breaking scheme	1,240	32,10	7,17	3,43	73,93	8,32
	unit: ground-breaking	0,778	23,26	7,32	2,87	75,65	8,88
	mot: breaking	0,381	20,09	4,70	0,82	73,32	9,10
Kitchen1	extract/unit: kitchen sink	0,724	31,87	8,00	2,78	73,17	8,12
	mot: kitchen	0,355	12,07	6,07	0,60	70,04	7,70
Massprol	extract/unit: mass production	0,634	24,95	9,82	2,01	67,45	8,61
	mot: production	0,438	38,12	10,05	2,40	68,78	9,14
Microch1	extract/unit: microchip technology	1,500	19,03	5,49	2,48	71,75	9,34
	unit: microchip	0,718	19,60	6,52	2,63	72,38	8,60
	mot: micro	0,466	19,23	7,65	1,88	73,47	8,92
	mot: technology	0,783	19,95	4,46	1,82	72,13	9,75
Talkbot2	extract/unit: talking bottle top	0,984	40,13	8,80	4,00	73,54	8,59
	unit: bottle top	0,561	34,40	8,06	3,77	75,68	8,55
	mot: talking	0,423	33,89	9,45	4,33	72,36	7,56
	mot: bottle	0,326	43,62	9,63	3,72	76,66	8,90
Talkbot3	extract/unit: talking bottle top	1,225	32,50	4,71	3,61	68,36	8,75
	unit: bottle top	0,689	34,62	6,69	3,59	70,29	9,66
	mot: talking	0,536	14,66	1,96	1,17	73,04	8,75
	mot: bottle	0,318	39,08	5,91	4,20	73,04	8,75
Talkbot1	extract/unit: talking vodka bottle top	1,644	20,22	5,71	3,68	73,21	8,61
	unit: bottle top	0,711	32,79	2,57	2,34	73,88	7,03
	mot: talking	0,514	22,67	10,22	0,92	74,24	8,80
Vodbott1	mot: vodka	0,418	28,54	4,76	1,86	72,62	10,35
	mot: bottle	0,372	50,52	2,93	2,78	73,77	7,71
	extract/unit: vodka bottle top	1,100	26,55	14,74	3,11	74,16	10,39
	unit: bottle top	0,728	26,29	12,85	1,95	73,01	10,27
	mot: vodka	0,372	17,89	18,40	0,87	76,39	11,15
	mot: bottle	0,359	26,70	12,99	2,31	76,39	8,01
Vodgen3	extract/unit: vodka genie	0,888	16 43	2,34	1,45	70,88	8,40
	mot: vodka	0,368	16,75	3,17	0,73	70,72	9,52
	mot: genie	0,520	17,43	1,90	1,58	72,08	7,50
Vodgen11	extract/unit: vodka genie	0,791	25,80	5,48	4,73	75,95	7,63
	mot: vodka	0,364	24,01	10,87	1,87	76,07	10,15
	mot: genie	0,428	27,06	2,14	2,45	77,04	4,27
Vodgen12	extract/unit: vodka genie	0,560	17,30	3,27	1,22	70,78	8,16
	mot: vodka	0,283	19,09	4,40	0,84	71,86	9,95
	mot: genie	0,276	17,07	2,43	0,76	71,37	5,64
worksho1	extract/unit: workshop	0,574	21,82	7,28	2,84	74,76	8,64

Annexe E : Graphiques correspondant aux extraits sonores

Les graphiques suivants correspondent aux extraits sonores cités dans le chapitre 4. L'oscillogramme est présenté sous la forme d'un TextGrid qui permet de montrer la segmentation en syllabes et les différents niveaux d'analyse pour lesquels nous avons pu prendre des mesures : la syllabe, avec la phonie ; le schéma accentuel en utilisant la notation qui a été proposée à la fin du chapitre 1 ; les mots ; l'unité ; l'extrait entier (qui correspond souvent à l'unité, mais pas toujours).

Les courbes de la fréquence fondamentale (F_0) et de l'intensité (correspondant à « I » dans les légendes des figures) sont représentées sur le même graphique. Le F_0 correspond à la ligne continue, l'intensité à la ligne brisée. L'échelle de F_0 est indiquée à gauche du graphique et se situe entre 0 et 350 Hz ; l'échelle de l'intensité se situe entre 40 et 100 dB.

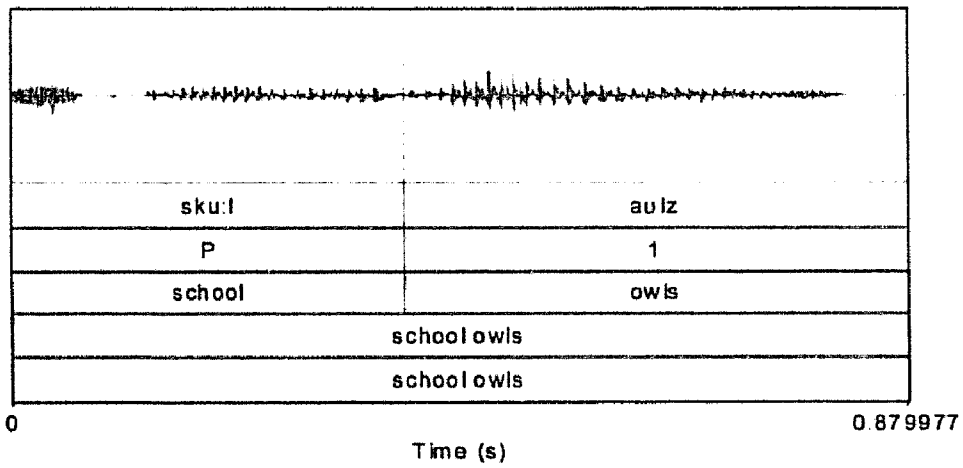
Le temps est toujours indiqué en bas, pour l'oscillogramme et pour les courbes de F_0 et de l'intensité ; il est en secondes.

Les graphiques sont présentés par ordre alphabétique, selon le nom du répertoire suivi du nom du fichier. Ces extraits se trouvent également sur le CDROM[3].

---

[3] Les extraits sonores se trouvent dans le répertoire «corporal, dans le sous-répertoire indiqué ci-après.

Figure IV.13 : broom\schowl1 : Oscillogramme et courbes de F_0 et I



L'oscillogramme permet de déceler une légère pause après [l] et avant la voyelle du deuxième élément. Cet extrait est commenté en 4.5.1.2.2 « L'accentuation double ».

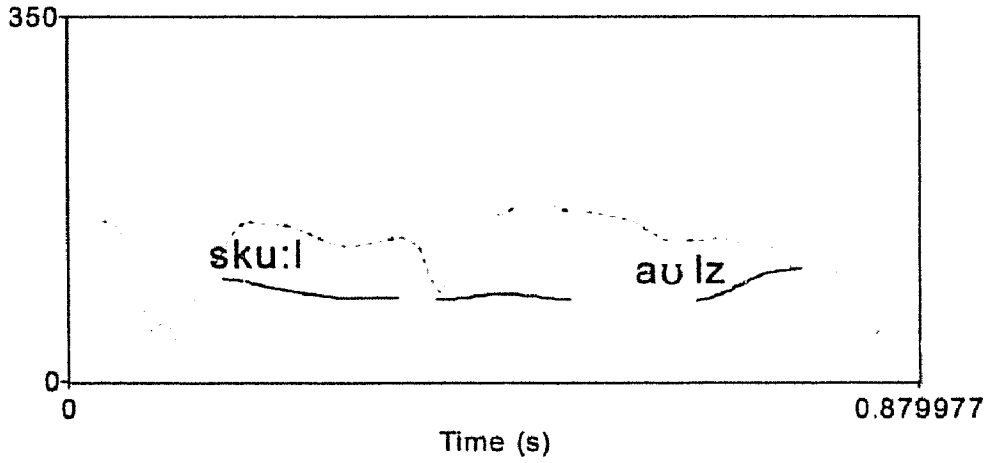
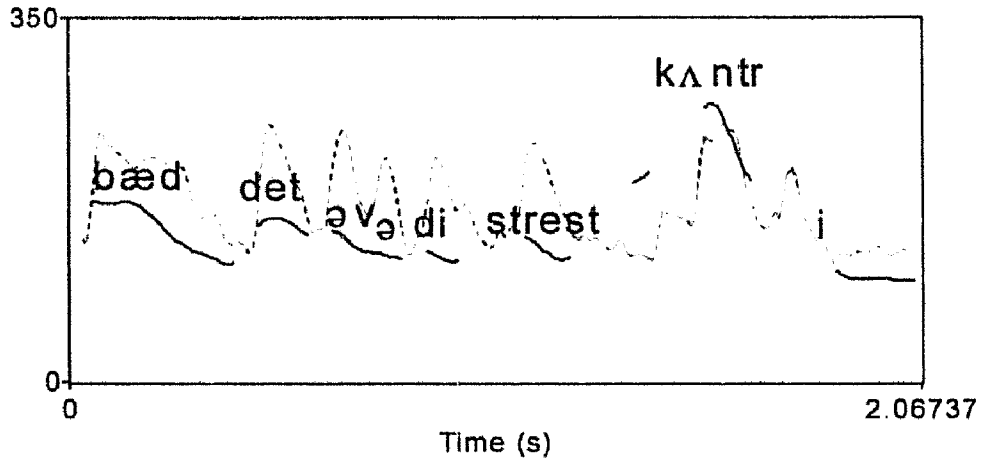
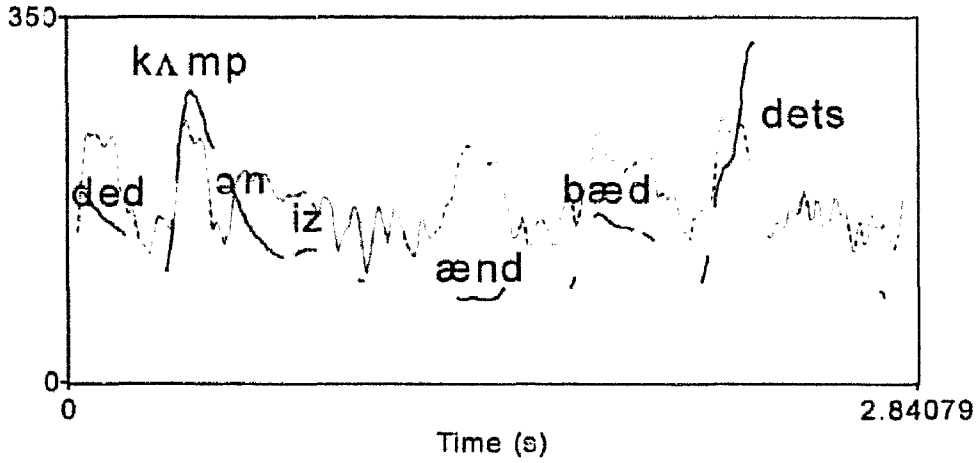


Figure IV.14 : debt\baddebt1: Courbes de F_0 et I



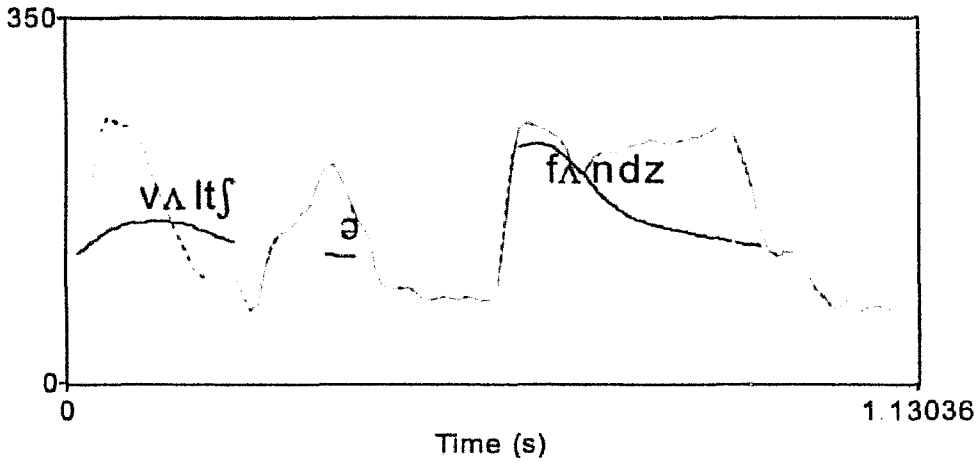
Il y a une proéminence sur *bad* et *country*, c'est-à-dire sur les extrémités, ce qui soutient l'hypothèse de « *Phrase Edge Prominence* ». Notons que le statut d'accentuation simple de *bad debt* se discute car il y a une petite chute sur le deuxième élément qui reflète celle du premier. Toutefois, il est vrai que pour l'accentuation double, le mouvement mélodique s'amplifie généralement sur le deuxième élément ce qui n'est pas le cas ici à cause de la position de *debt* dans l'extrait. C'est pourquoi nous avons conclu qu'il y a une accentuation simple. Cet extrait est commenté en 4.5.3.2, « *Phrase Edge Prominence* ou le déplacement de l'accent ».

Figure IV.15 : debt\dcadcom1 : Courbes de F_0 et I



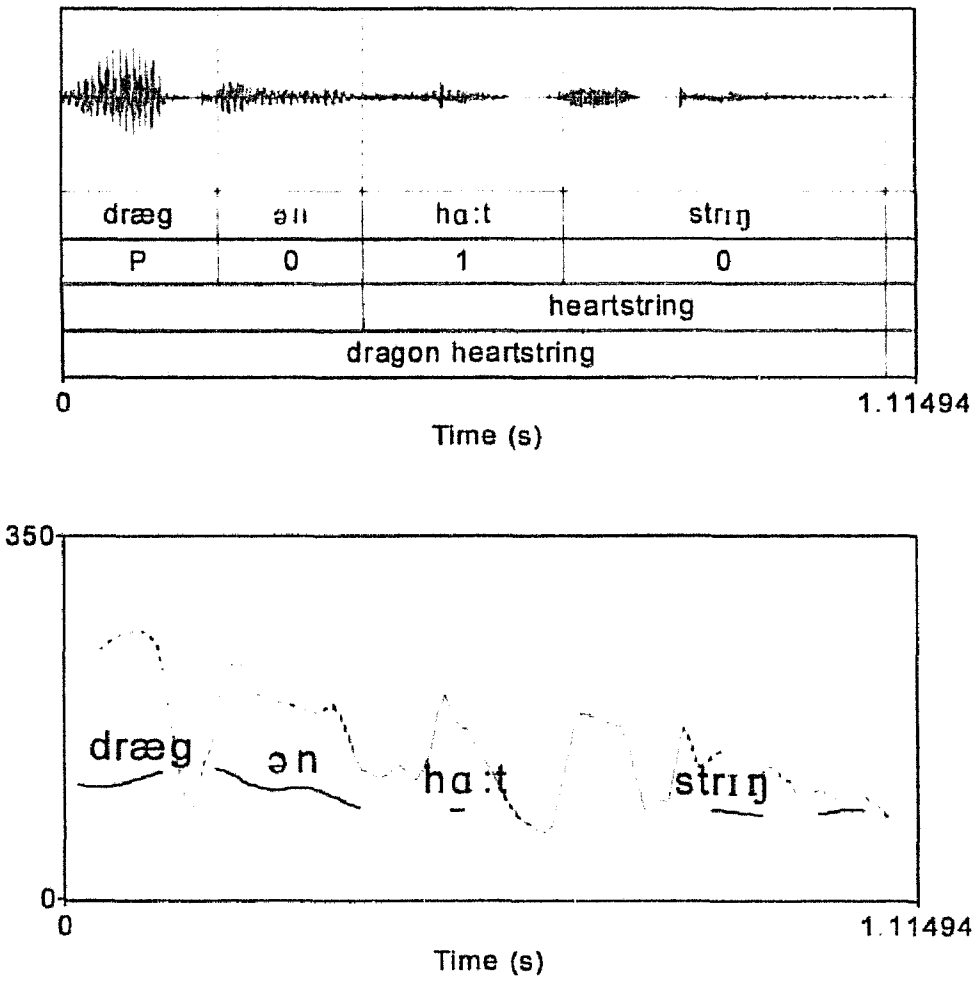
Cette courbe présente des anomalies en raison de l'envol du F_0 sur *debts*. Toutefois, la mise en relief de *companies* et *debts* apparaît clairement, ce qui soutient l'hypothèse de « *Phrase Edge Prominence* ». Cet extrait est commenté en 4.5.3.2, « *Phrase Edge Prominence* ou le déplacement de l'accent ».

Figure IV.16 : debt\vultfun4 : Courbes de F_0 et I



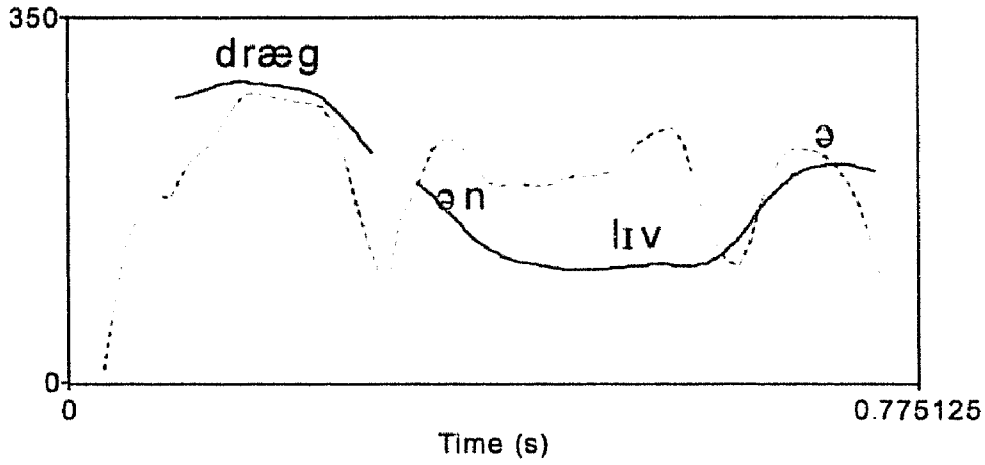
L'accentuation double apparaît clairement dans cet extrait car la chute sur *vulture* est reprise et amplifiée sur le deuxième élément *funds*. Cet extrait est comparé à d'autres où nous trouvons une accentuation simple et il est commenté en 5.3.4, « La variation ».

Figure IV.17 : dragon\dragheart1 : Oscillogramme et courbes de F_0 et I



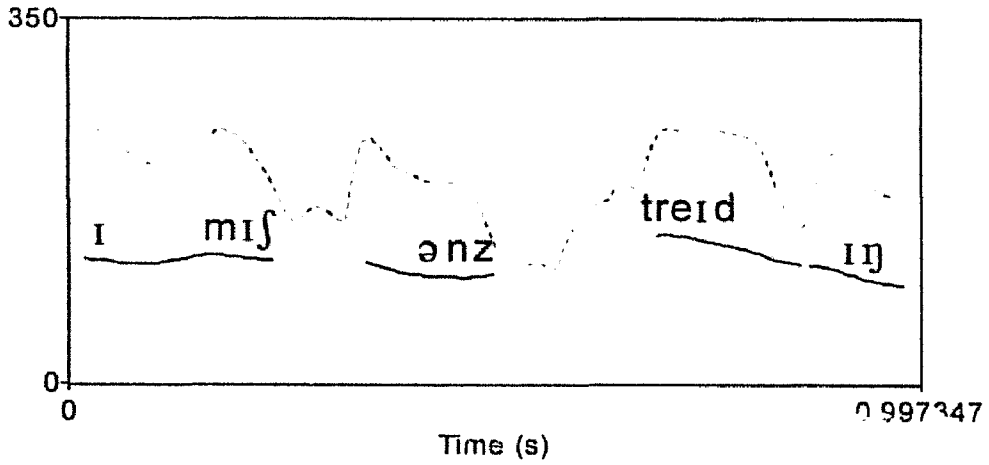
Cet extrait est un composé radical, et sert d'exemple pour l'accentuation simple. Il y a une proéminence sur *dragon*, avec une valeur plus élevée du F_0 et de l'intensité, puis sur les autres éléments, ces valeurs restent basses. Il est cité en 4.6, « L'analyse acoustique, l'analyse de l'accent lexical et les règles d'accentuation des composés ».

Figure IV.18 : dragon\dragliver1 : Courbes de F_0 et I



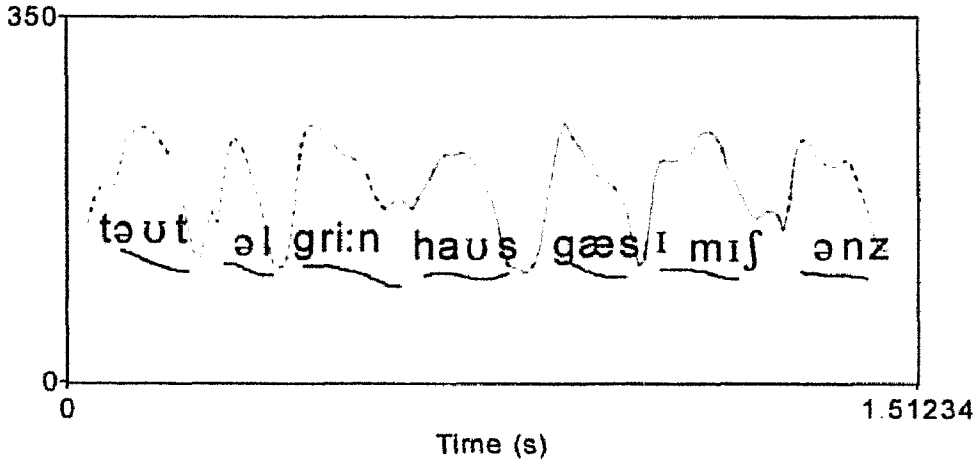
Cet extrait a une accentuation simple, comme le montre la chute sur le premier élément. La petite remontée sur le deuxième indique une poursuite de l'information. Notons la bonne corrélation entre les valeurs élevées du F_0 et de l'intensité qui montre la proéminence sur *dragon*. Cet extrait est commenté en 4.5.1.2.2, « Interprétation du contour intonatif ».

Figure IV.19 : emiss\emiss1 : Courbes de F_0 et I



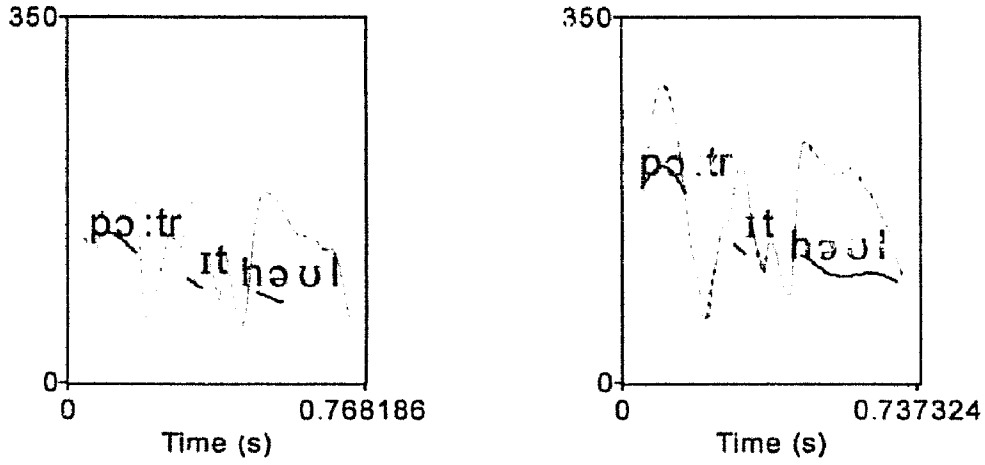
Il y a une petite chute sur le premier élément, et celle-ci est amplifiée sur le deuxième élément de cette unité, *trading*. L'accentuation double s'explique par le fait que *emissions* est l'ancienne information et il y a donc une mise en relief de *trading*. Cet extrait est cité en 4.5.1.2.3.1, « Les contextes non neutres : Focus ».

Figure IV.20 : emiss\emiss2 : Courbes de F_0 et I



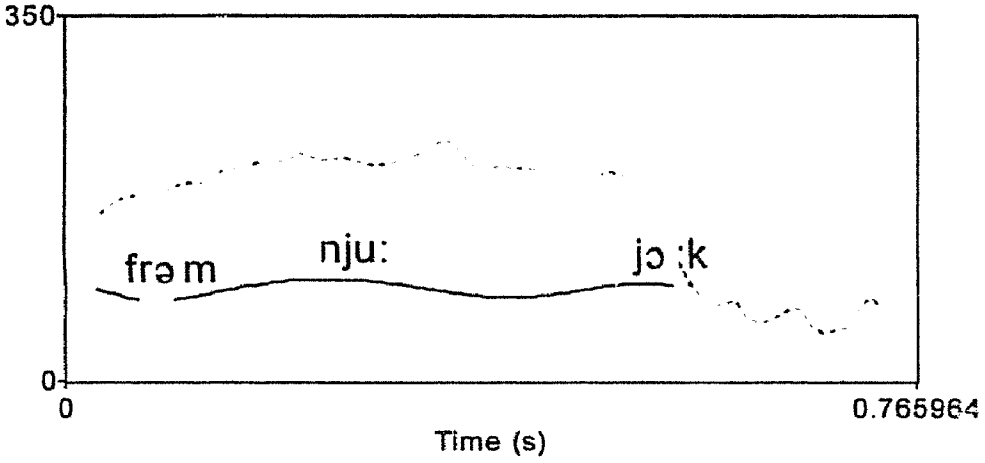
Nous constatons la mise en relief de *total*, en début de l'unité, mais il joue un rôle d'adjectif qualificatif. L'accent nucléaire se trouve sur *gas*. Il y a deux courbes semblables sur *greenhouse*, puis sur *gas emissions*. L'extrait est cité en 4.5.1.2.3.1, « Les contextes non neutres : Focus ».

Figure IV.21 : hole\portrh3 et hole\portrh4 : Courbes de F_0 et I



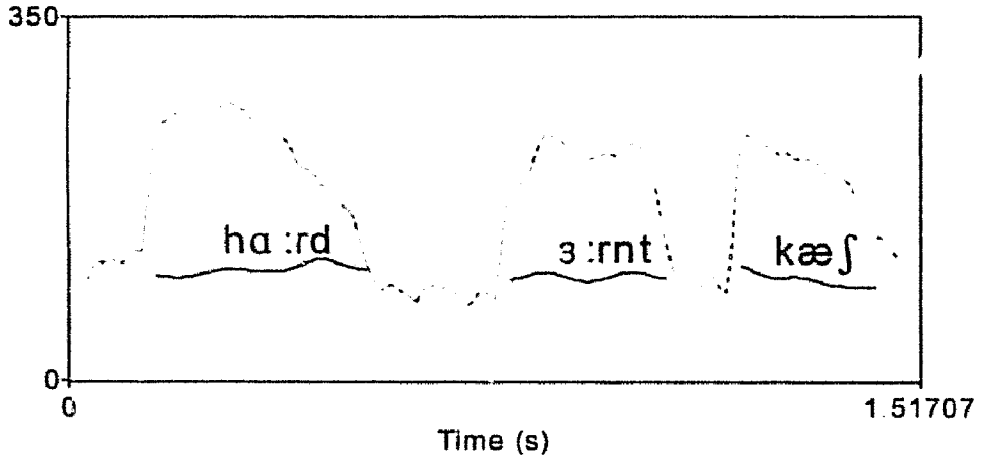
Ces deux extraits présentent les mêmes caractéristiques d'accentuation simple avec une chute du F_0 et de l'intensité sur le premier élément. Toutefois, le graphique de droite montre que ce mouvement est amplifié. Ceci s'explique par le contexte où le locuteur trahit son émotion ; l'extrait est commenté en 4.5.1.2.3.4, « Les contextes non neutres : Les facteurs psychologiques ».

Figure IV.22 : newyork\newyork2 : Courbes de F_0 et I



Dans cet extrait, on devine à peine la double accentuation avec une très légère baisse sur chacun des éléments. Il n'y a pas de véritable mise en relief et la courbe est très plate. Voir les commentaires en 4.5.3.3, « Les composés isolés et les composés dans une construction cumulative ».

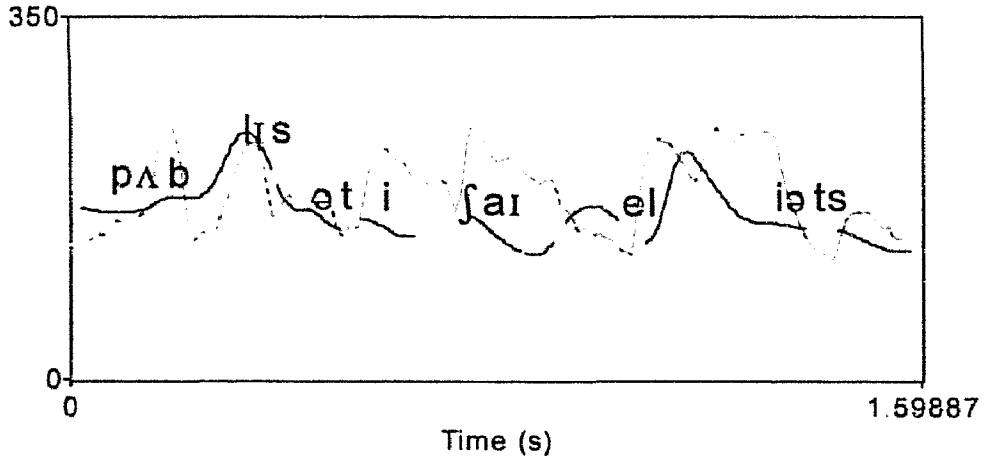
Figure IV.23 : shift\hardear1 : Courbes de F_0 et I



La baisse de l'intensité entre *hard* et *earned* montre clairement qu'il y a une pause entre ces éléments. Nous constatons également le débit lent de l'extrait. Contrairement aux attentes, il n'y a pas de déplacement de l'accent : les trois

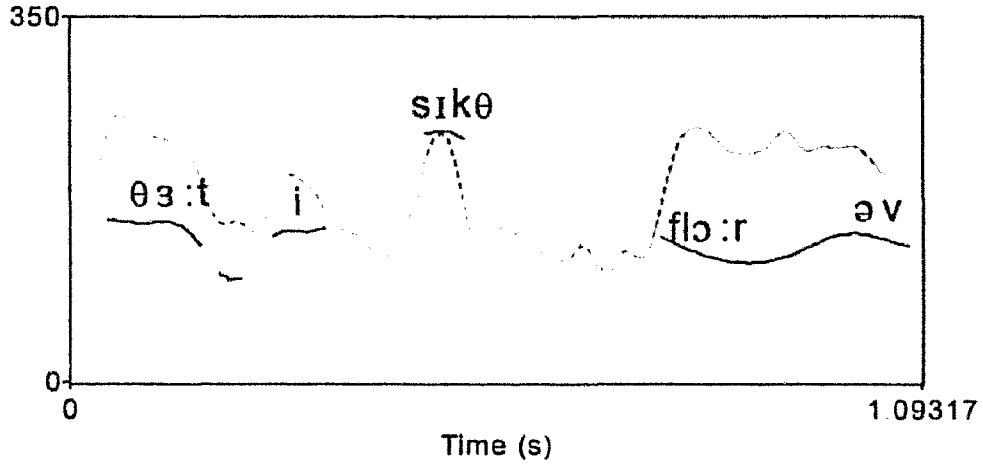
éléments se prononcent à peu près de la même manière. Voir les remarques en 4.5.3.1, « Le déplacement de l'accent ».

Figure IV.24 : shift\publici : Courbes de F_0 et I



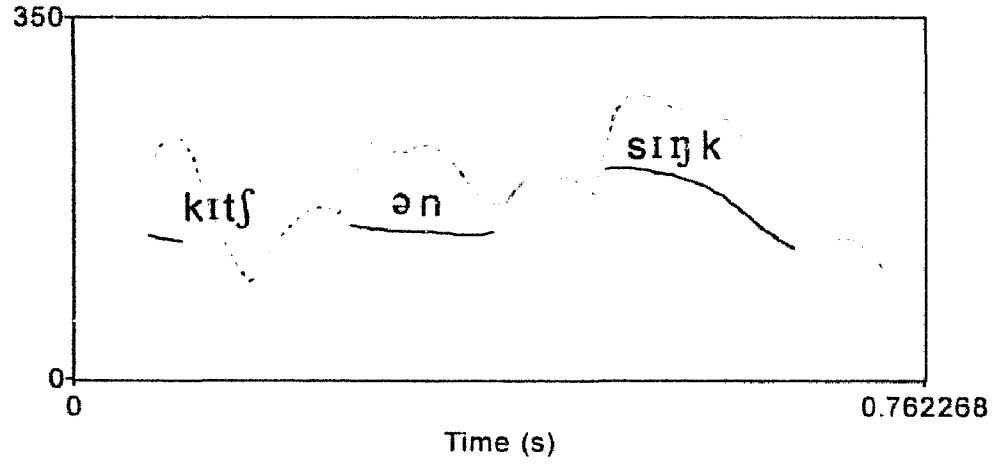
Dans cet exemple, la proéminence sur la deuxième syllabe de *publicity* ressort nettement en raison de la valeur élevée de F_0. En revanche, *shy* correspond à une valeur basse : le déplacement de l'accent sur le premier élément de cet adjectif composé est donc évident. Cet extrait est cité en 4.5.3.1, « Le déplacement de l'accent ».

Figure IV.25 : shift\thirty1 : Courbes de F_0 et I



La proéminence correspond à *sixth* qui est le noyau de cette unité. Il n'y a pas de déplacement de l'accent. Voir les commentaires en 4.5.3.2, « *Phrase Edge Prominence* ou le déplacement de l'accent ».

Figure IV.26 : vodka\kitchen1: Courbes de F_0 et I



Cet extrait est cité en 4.5.1.2.2, « L'accentuation double ». Le second élément comporte une chute, et il y a une bonne corrélation entre les valeurs élevées du F_0 et de l'intensité sur *sink*, qui fait ressortir la proéminence. Cette

amplification du signal, plutôt qu'une simple reprise, s'explique par le contour « rise-fall » où le locuteur exprime sa surprise.

Annexe F: Listes d'entrées extraites de la base de données

Dans cette annexe, nous pré entons des listes d'entrées que nous avons extraites de la base de données à partir de certaines requêtes. Il y a deux ensembles: les entrées de N-COMPACT qui ont été relevées à la saisie comme étant susceptibles de nous intéresser; et les entrées comportant un élément qui intervient dans la construction des unités des extraits sonores. Nous avons sélectionné un échantillonnage correspondant aux éléments qui se sont révélés les plus intéressants au niveau de l'analyse: *break, fly, green* et *work.*

Pour chacun de ces ensembles, nous avons retenu les informations suivantes: type de schéma (codé «s» pour une entrée à accentuation simple ou «d» pour une entrée à accentuation double); graphie (y compris les variantes graphiques de LPD et les symboles spécifiques[4]); et la catégorie syntaxique (intitulé «Catgr»). Cette dernière colonne ne contient que les codes que nous avons utilisés pour nos analyses: «N» (nom), «Q» (adjectif), «V» (verbe), «tdmk» (nom de marque) ... Il ne s'applique pas à la table N-COMPACT. Nous n'avons pas jugé utile d'inclure les indications sur la phonie, ni la graphie inverse.

Pour le premier ensemble qui concerne N-COMPACT uniquement, nous avons inclus les informations précises portant sur le schéma accentuel lui-même, avec le schéma en variante, entre parenthèses, le cas échéant.

---

[4] Par exemple, la barre verticale qui signale l'emplacement, pour la version papier du dictionnaire, des indications concernant les terminaisons.

Entrées N-COMPACT présentant un intérêt particulier

Le tableau suivant contient les entrées extraites à l'aide de la requête **shift-NC** (voir Annexe A). La partie gauche du tableau contient les entrées à accentuation simple, la partie droite les entrées à accentuation double. Ces entrées sont commentées en 4.6.1.

Figure IV.27 : Sélection d'entrées de N-COMPACT

Type schéma	Graphie	Schéma (variante)	Type schéma	Graphie	Schéma (variante)
s	butterfly bush	100#0	d	barley wine	20#1
s	Central nervous system	20#10#30	d	bay rum	2#1
s	central processing unit	20#100#30	d	bed room slippers	2#0#10
s	clay pigeon shooting	2#10#30	d	cotton candy	20#10
s	correlation coefficient	2010#0030	d	covering letter	200#10
s	double's match	10#0	d	desert rat	20#1
s	flying boat	10#0	d	flat spin	2#1
s	flying machine	10#03	d	flying buttress	20#10
s	flying officer	10#300	d	flying colours	20#10
s	flying squad	10#0	d	flying doctor	20#10
s	freezing compartment	10#030	d	flying fish	20#1
s	frequency curve	100#3	d	flying fox	20#1
s	front man	1#0	d	flying picket	20#10
s	giant killer	10#30	d	flying saucer	20#10
s	green belt	1#0	d	flying start	20#1
s	harvest mite	10#0	d	fruit salad	2#10
s	harvest mouse	10#0	d	giant panda	20#10
s	high jump	1#0	d	greasy spoon	20#1
s	high life	1#0	d	grizzly bear	20#1 (10#0)
s	high point	1#0	d	harvest festival	20#100
s	high road	1#0	d	harvest home	20#1
s	holy water	10#30 (20#10)	d	harvest moon	20#1
s	hot rod	1#0	d	high priestess	2#10
s	laughing gas	10#0	d	horse chestnut	2#10 (US 1#30)
s	liege man	1#0	d	hot dog	2#1 (US 1#0)

s	life raft	1#0	d	laughing jackass	20#10
s	life story	1#30	d	liege lord	2#1
s	light bulb	1#0	d	light ale	2#1
s	little people	10#30	d	little finger	20#10
s	local time	10#0	d	little woman	20#10
s	long johns	1#0	d	long haul	2#1
s	long jump	1#0	d	long suit	2#1
s	long jumper	1#30	d	long ton	2#1
s	long shot	1#0	d	long vac	2#1
s	me generation	1#0030	d	long vacation	2#010
s	moving van	10#0	d	long wave	2#1 (1#0)
s	new town	1#0 (2#1)	d	market value	20#10 (10#30)
s	old girl	1#0	d	pith helmet	2#10
s	on side	1#0	d	running jump	20#1
s	real estate	1#03	d	sausage roll	20#1
s	real estate agent	1#00#30	d	seller's market	20#10
s	sausage dog	10#0	d	sharp practice	2#10
s	sausage meat	10#0	d	sixth sense	2#1
s	school board	1#0	d	sleeping partner	20#10
s	set theory	1#30	d	speaking clock	20#1
s	sharp end	1#0	d	spinning jenny	20#10
s	sixth form	1#0	d	stable door	20#1
s	solar system	10#30	d	stormy petrel	20#10
s	speaking tube	10#0	d	surface tension	20#10
s	spinning wheel	10#0	d	trial run	20#1
s	stable boy	10#0	d	tuna salad	20#10
s	train set	1#0	d	twin bed	2#1
s	train spotter	1#30	d	vegetable marrow	200#10
s	trial court	10#0			
s	tuna fish	10#0			
s	twin set	1#0			
s	women's movement	10#30			
s	women's studies	10#30			

Entrées correspondant à la recherche d'un élément

Les entrées des tableaux suivants contiennent un élément donné. L'extraction s'est faite d'après la requête **RECH-EL-table** (voir Annexe A). Elles sont classées par type de schéma, puis par ordre alphabétique. Les trois premières colonnes correspondent aux entrées de EPD (« Type schéma », « Graphie », « Catgr »), les trois suivantes à NPD (les mêmes intitulés), les trois suivantes à LPD (les mêmes intitulés), les deux dernières à N-COMPACT (« Type schéma » et « Graphie »). Ces entrées sont commentées en 4.6.3.

Figure IV.28 : Entrées contenant l'élément *break*

EPD			NPD			LPD			N-COMPACT	
Type schéma	Graphie	Catgr	Type schéma	Graphie	Catgr	Type schéma	Graphie	Catgr	Type schéma	Graphie
s	break-away	N	s	backbreaker	N-Cp	s	backbreaker/s	N	s	breaking point
s	breakfast-set	N	s	backbreaking	Q-Cp	s	breakaway	N	s	circuit breaker
s	break-neck	Q	s	breakdown	N-Cp	s	breakdown	N-	s	coffee break
s	day-break	N	s	breakfast	NV-Cp	s	breakfront	N	s	station break
s	house-breaker	N	s	breakfront	N-Cp	s	break-in	N	s	tea break
s	house-breaking	N	s	breakneck	NQ-Cp	s	breakneck	NQ	s	wedding breakfast
s	jaw-breaker	N	s	breakout	N-Cp	s	breakout	N	d	bed and breakfast
s	law-breaker	N	s	breakthrough	N-Cp	s	breakthrough	N	d	breaking and entering
s	law-breaking	N	s	breakup	N-Cp	s	breakup	N	d	continental breakfast
s	oath-breaker	N	s	breakwater	N-Cp	s	breakwater	N	d	dog's breakfast
s	oath-breaking	N	s	daybreak	N-Cp	s	daybreak	N	d	nervous break down
s	stone-breaker	N	s	heartbreak	N-Cp	s	firebreak	N		
s	strike-breaker	N	s	heartbreaking	Q-Cp	s	heartbreak	N		
d	break-through	N	s	horsebreaker	N-Cp	s	heartbreaking	Q		
d	break-up	N	s	jailbreak	N-Cp	s	housebreaker/s	N		
			s	tiebreak	N-Cp	s	icebreaker	N		
			s	windbreak	N-Cp	s	jailbreak	N		
			s	wordbreak	N-Cp	s	jawbreaker	N		
						s	law-breaker	N		
						s	outbreak	N		
						s	record-breaker/s	N		
						s	safebreaker	N		
						s	stonebreaker	N		
						s	strikebreaker	N		
						s	strikebreaking	N		
						s	tiebreak	N		
						s	tiebreaker	N		
						s	windbreak	N		
						s	windbreaker, W~	N-±tdmk//		
						s	wordbreak	N		
						d	break-even, breakeven	V		

Figure IV.29 : Entrées contenant l'élément *fly*

EPD			NPD			LPD			N COMPACT	
Type schéma	Graphie	Catgr	Type schéma	Graphie	Catgr	Type schéma	Graphie	Catgr	Type schéma	Graphie
s	blow-fly	N	s	barfly	N-Cp	s	bar|fly	N	s	butterfly bush
s	caddis-fly	N	s	blackfly	N-Cp	s	black|fly	N	s	butterfly stroke
s	dragon-fly	N	s	blowfly	N-Cp	s	blow|fly	N	s	butterfly valve
s	fly-blow	N	s	botfly	N-Cp	s	bot|fly	N	s	flying boat
s	fly-blown	Q	s	butterfly	N-Cp	s	catch|fly	N	s	flying machine
s	fly-bomb	N	s	catchfly	N-Cp	s	dragonfly	N	s	flying officer
s	fly-catcher	N	s	firefly	N-Cp	s	firefly	N	s	flying squad
s	fly-fishing	N	s	flyback	N-Cp	s	fly|leaf	N	s	fruit fly
s	flying-man	NY	s	flyblown	Q-Cp	s	flyaway	QN	s	hover fly
s	flying-men	NZ	s	flyby	N-Cp	s	flyback	N	s	ichneumon fly
s	flying-officer	N	s	flyleaf	N-Cp	s	flyblown	Q	s	sand fly
s	fly-leaf	N	s	flyman	N-Cp	s	flyby	N	s	tsetse fly
s	fly-line	N	s	flymen	N-Cp	s	fly-by-night	Q	s	warble fly
s	fly-paper	N	s	flypast	N-Cp	s	flycatcher	N	d	fly half
s	fly-sheet	N	s	flyscreen	N-Cp	s	fly-drive	N	d	flying buttress
s	fly-swatter	N	s	flysheet	N-Cp	s	fly-fish	N	d	flying colours
s	gall-fly	N	s	flyspeck	N-Cp	s	Flymo	N±tdmk	d	flying doctor
s	house-fly	N	s	flytrap	N-Cp	s	flyover	N	d	flying fish
s	hover-fly	N	s	flyway	N-Cp	s	flypaper	N	d	flying fox
s	kite-flying	N	s	flyweight	N-Cp	s	flypast	N	d	flying picket
s	may-fly	N	s	flywheel	N-Cp	s	flyscreen	N	d	flying saucer
			s	flywhisk	N-Cp	s	flysheet	N	d	flying start
			s	gadfly	N-Cp	s	flyspeck	N	d	Venus's fly trap
			s	greenfly	N-Cp	s	flyswatter	N		
			s	horsefly	N-Cp	s	flytrap	N		
			s	housefly	N-Cp	s	flyweight	N		
			s	mayfly	N-Cp	s	flywheel	N-		
			s	sandfly	N-Cp	s	flywhisk	N		
			s	shoofly	N-Cp	s	gad|fly	N		
			s	whitefly	N-Cp	s	green|fly	N		
			d	highflyer	N-Cp	s	horse|fly	N		
						s	house|fly	N		
						s	may|fly	N		
						s	shoofly	N		
						s	white|fly	N		
						d	high-flier, high-flyer	N		
						d	high-flying	N		
						d	overfly	V		

Figure IV.30 : Entrées contenant l'élément *green*

EPD			NPD			LPD			N-COMPACT	
Type schéma	Graphie	Catgr	Type schéma	Graphie	Catgr	Type schéma	Graphie	Catgr	Type schéma	Graphie
s	bottle-green	QDG	s	evergreen	NQ-Cp	s	evergreen	NQ	s	bowling green
s	bowling-green	N	s	greenback	N-Cp	s	greenfly	N	s	green belt
s	green-eyed	Q	s	greenfinch	N-Cp	s	greenhouse	N	s	green house effect
s	green-room	N	s	greenfly	N-Cp	s	greenback	N	s	Greenwich Mean Time
s	jack-in-the-green	N	s	greengrocer	N-Cp	s	greenfinch	N	d	bice green
s	putting-green	N	s	greenhorn	N-Cp	s	greengrocer	N-	d	blue -green algae
d	grass-green	Q	s	greenhouse	N-Cp	s	greengrocery	N	d	bottle green
d	pea-green	Q	s	greenroom	N-Cp	s	greenhorn	N	d	green bean
d	sea-green	NQ	s	greensand	N-Cp	s	Greenpeace	N	d	green fingers
			s	greenshank	N-Cp	s	greenroom	N	d	green light
			s	greenstick	N-Cp	s	greensand	N	d	green Paper
			s	greenstone	N-Cp	s	greenshank	N	d	green pepper
			s	greenstuff	N-Cp	s	Greensleeves	N	d	green tea
			s	greensward	N-Cp	s	greenstick	N	d	green thumb
			s	greenwood	N-Cp	s	greenstone	N	d	green-eyed monster
			d	greengage	N-Cp	s	greenstuff	N	d	Greenwich Village
						s	greensward	N	d	Gretna Green
						s	greenwood, G-	N	d	Hither Green
						s	wintergreen	N	d	lime green
						d	blue-green	Q	d	pea green
						d	green-eyed	N	d	sage green
									d	sea green
									d	Stepney Green

Figure IV.31 : Entrées contenant l'élément *work*

EPD			NPD			LPD			N-COMPACT	
Type schéma	Graphie	Catgr	Type schéma	Graphie	Catgr	Type schéma	Graphie	Catgr	Type schéma	Graphie
s	basket-work	N	s	artwork	Q-Cp	s	artwork	Q	s	dirty work
s	clock-work	N	s	beadwork	N-Cp	s	basketwork	N	s	guest worker
s	dye-works	N	s	bookwork	N-Cp	s	beadwork	N	s	old -boy network
s	fancy-work	N	s	breastwork	N-Cp	s	bodywork	N	s	part work
s	field-work	N	s	brickwork	N-Cp	s	bookwork	N	s	road works
s	flue-work	N	s	bridgework	N-Cp	s	breastwork	N	s	sanitation worker
s	glass-work	N	s	brightwork	N-Cp	s	brickwork	N	s	shift worker
s	head-work	N	s	brushwork	N-Cp	s	bridgework	N	s	shift working
s	house-work	N	s	casework	N-Cp	s	brightwork	N	s	social work
s	lattice-work	N	s	chainwork	N-Cp	s	brushwork	N	s	working party
s	net-work	N	s	clockwork	N-Cp	s	busywork	N	d	back stage workers
s	night-work	N	s	coachwork	N-Cp	s	casework	N	d	full-time work
s	open-work	N	s	coursework	N-Cp	s	caseworker	N	d	life work
s	out-worker	N	s	daywork	N-Cp	s	clockwork	N	d	life's work
s	piece-work	N	s	deskwork	N-Cp	s	coachwork	N	d	local area network
s	spade-work	N	s	earthwork	N-Cp	s	coursework	N	d	piece of work
s	stone-work	N	s	fieldwork	N-Cp	s	deskwork	N	d	pioneering work
s	team-work	N	s	fireworks	N-Cp	s	donkeywork	N	d	public works
s	time-work	N	s	footwork	N-Cp	s	dyeworks	N-	d	worked up
s	trellis-work	N	s	framework	N-Cp	s	earthwork	N	d	working class
s	wicker-work	N	s	fretwork	N-Cp	s	fancywork	N	d	working day
s	wonder-worker	N	s	frostwork	N-Cp	s	fieldwork	N	d	working knowledge
s	work-bag	N	s	gasworks	N-Cp	s	firework	N	d	working order
s	work-basket	N	s	glassworks	N-Cp	s	footwork	N	d	working week
s	work-box	N	s	groundwork	N-Cp	s	framework	N		
s	work-day	N	s	guesswork	N-Cp	s	fretwork	N		
s	work-room	N	s	hackwork	N-Cp	s	gasworks	N		

s	work-shy	Q	s	handicraftwork	N-Cp	s	glassworks	N-
s	work-table	N	s	handwork	N-Cp	s	groundwork	N
d	hand-worked	Q	s	homework	NQ-Cp	s	guesswork	N
d	life-work	N	s	hornwork	N-Cp	s	hackwork	N
			s	housework	N-Cp	s	handiwork	N
			s	ironwork	N-Cp	s	handwork	N
			s	legwork	N-Cp	s	homework	QN
			s	needlework	N-Cp	s	housework	N
			s	network	NV-Cp	s	ironwork	N
			s	outwork	N-Cp	s	legwork	N
			s	paintwork	N-Cp	s	masterwork	N
			s	patchwork	N-Cp	s	metalwork	N
			s	piecework	N-Cp	s	metalworker/s	N
			s	rockwork	N-Cp	s	needlework	N
			s	saltworks	N-Cp	s	network	NV
			s	schoolwork	N-Cp	s	openwork	N
			s	scrollwork	N-Cp	s	outwork	N-o
			s	shiftwork	N-Cp	s	paintwork	N
			s	spadework	N-Cp	s	paperwork	N
			s	steelworks	N-Cp	s	patchwork	N
			s	stonework	N-Cp	s	piecework	N
			s	teamwork	N-Cp	s	pokerwork	N
			s	timework	N-Cp	s	saltworks	N
			s	waterworks	N-Cp	s	schoolwork	N
			s	waxwork	N-Cp	s	scrollwork	N
			s	woodwork	N-Cp	s	shiftwork	N
			s	woolwork	N-Cp	s	spadework	N
			s	workaday	Q-Cp	s	steelworker	N
			s	workbag	N-Cp	s	steelworks	N-
			s	workbench	N-Cp	s	stonework	N
			s	workbook	N-Cp	s	teamwork	N
			s	workbox	N-Cp	s	timework	N
			s	workday	N-Cp	s	waterworks	N
			s	workfare	N-Cp	s	waxwork	N
			s	workforce	N-Cp	s	wickerwork	N
			s	workhorse	N-Cp	s	woodwork	N
			s	workhouse	N-Cp	s	workhouse	N
			s	workload	N-Cp	s	workaday	QN
			s	workman	N-Cp	s	workbag	N
			s	workmanly	Q-Cp	s	workbasket	N
			s	workmate	N-Cp	s	workbench	N
			s	workmen	N-Cp	s	workbook	N
			s	workout	N-Cp	s	workbox	N
			s	workplace	N-Cp	s	workday	N
			s	workroom	N-Cp	s	workfare	N
			s	workshop	N-Cp	s	workforce	N
			s	workshy	Q-Cp	s	workhorse	N
			s	workspace	N-Cp	s	work-in	N
			s	worktop	N-Cp	s	workload	N

		s	workmate	N	
		s	workout	N	
		s	workpeople	N	
		s	workplace	N	
		s	workroom	N	
		s	work-sharing	N	
		s	workshop	N	
		s	workshy	Q	
		s	workspace	N	
		s	workstation	N	
		s	work-study	N	
		s	worktable	N	
		s	worktop	N	
		d	hard-working	Q	
		d	overwork	V-	
		d	overwork	N-	
		d	worker-priest	N	
		d	working-class	Q	
		d	working-out	N	
		d	work-to-rule	N	

CONCLUSION

Conclusion

Dans notre conclusion, nous voudrions d'abord rappeler les acquis de la recherche, puis nous terminerons par une discussion des considérations relatives à un éventuel dictionnaire des composés.

Les acquis des résultats de la recherche

Ce travail nous a permis d'évaluer les règles d'accentuation des composés qui se sont avérées efficaces. L'opposition entre « fonction », associée à l'accentuation simple, et « nature », associée à l'accentuation double, explique un très grand nombre de cas. Nous trouvons aussi la mise en relief du premier élément lorsqu'il y a un contraste implicite, lorsqu'il y a une sous-catégorisation, ou bien lorsque le dernier élément est prévisible (par exemple, s'il a une portée générale). Les deux éléments portent un accent, donc l'accentuation est double, lorsque la relation entre eux est simple (quand le premier indique la nature du second). Ce schéma se trouve également dans les cas de localisation du premier élément par rapport au second ou encore lorsqu'il s'agit d'une véritable appellation. Nous avons formulé l'hypothèse d'une opposition entre « appellation », qui est associée à un schéma à accentuation double, et « sous-catégorisation », associée à un schéma à accentuation simple.

Le tableau suivant récapitule les résultats de nos recherches en ce qui concerne ces règles, mais aussi les tendances que nous avons remarquées lors de l'analyse de l'accent lexical. Toutefois, les catégories syntaxiques n'ont été incluses que lorsque les résultats montrent une tendance très nette vers un type d'accentuation.

Figure C1 : Les règles et les tendances de l'accentuation des composés

Accentuation simple	Accentuation double
Le premier élément indique la fonction du second.	Le premier élément indique la nature du second ; le sens de l'ensemble s'interprète de manière linéaire.
Le premier élément est mis en relief car la relation entre les éléments est complexe et ne peut pas s'interpréter de manière linéaire.	
Il y a une mise en relief d'une caractéristique représentative.	Le premier élément sert de localisation du second.
Le premier élément implique un contraste, et le composé fait partie d'une série. Il y a une sous-catégorisation.	Le premier élément sert d'appellation.
Il y a une mise en relief du premier élément lorsque le second est prévisible (par exemple, s'il a une portée générale).	
Tendance : Les composés ayant peu de syllabes, et ayant une formation simple.	Tendance : Les composés ayant plusieurs syllabes ou ayant une formation complexe.
Tendance : Les composés appartenant à la catégorie syntaxique du nom	Tendance : Les composés appartenant à la catégorie syntaxique de l'adjectif, à l'état isolé.

Nous devons préciser que ces explications se chevauchent souvent : le type d'accentuation peut s'expliquer de deux manières, comme nous l'avons vu dans nos analyses (4.6). L'interprétation de certaines entrées peut également poser quelques problèmes. En effet, il n'est pas toujours aisé de différencier les cas de sous-catégorisation où il y a un contraste implicite, associés à l'accentuation simple, et les cas où il y a une localisation, associés à l'accentuation double. Pourtant, les schémas accentuels de ces entrées sont rarement sujets à variation, ce qui sous-entend qu'il n'existe qu'une manière de les interpréter.

En ce qui concerne la variation, elle affecte un certain nombre d'unités, ce qui suppose que l'accent peut changer, ou bien selon le contexte syntaxique, ou bien simplement selon le gré du locuteur. Dans le premier cas, il s'agit de

l'influence du contexte de discours que nous avons étudiée en 4.5 et qui influence de manière sensible les schémas accentuels. Dans le second cas, il s'agit d'une variation libre, qui n'est pas en relation avec le discours, mais nous avons noté que les différences de l'accentuation du composé peuvent correspondre à des différences d'interprétation.

Les tendances indiquées dans le tableau concernent la longueur et la complexité des unités, ainsi que leur type de formation et la catégorie syntaxique à laquelle elles appartiennent. Toutefois, ce ne sont que des tendances et non des règles. Plus le composé est long, plus il est complexe, plus nous avons de chances qu'il porte une accentuation double. Ces tendances sont sans doute à rapprocher des schémas habituels de la langue, et de la préférence naturelle de l'anglophone pour les séquences courtes. La proportion importante d'entrées de deux ou trois syllabes dans les dictionnaires en témoigne. L'analyse des extraits sonores a permis d'étudier non seulement les cas de composés lexicalisés mais aussi des unités de discours plus longues dans lesquelles ces tendances se sont vérifiées.

Un dictionnaire de composés

Ayant étudié les composés qui figurent dans des dictionnaires de prononciation, nous nous posons la question de l'opportunité d'un dictionnaire spécifique pour les rassembler. Quels seraient les objectifs d'un tel ouvrage ? Il nous semble qu'il devrait comporter deux types d'indications : celles qui concernent la prononciation et l'accentuation du composé ; puis celles qui concernent sa formation, ce qui permettrait de faire des recherches sur la prononciation tout en comprenant les associations sémantiques sous-entendues.

L'élaboration d'un ouvrage de ce genre pose trois grandes questions : le classement des entrées, les critères d'inclusion, et enfin, le choix des indications concernant la prononciation et le schéma accentuel. Etudions d'abord les problèmes que peut poser le classement des entrées.

Le classement des composés

Le classement des composés demande beaucoup de réflexion. Dans le chapitre 3, nous avons soulevé le problème de l'homophonie qui fait que les dictionnaires de prononciation mettent les entrées ayant le même premier élément ensemble, même si cet élément a plusieurs sens. Nous avons donné l'exemple de *train* qui signifie « train ferroviaire » ou « entraînement ». Il nous semble gênant de ne pas les distinguer. Les dérivés posent des problèmes semblables : si le premier élément a des formes en *-ed* ou *-ing*, celles-ci seront ordonnées d'après un classement alphabétique strict qui ne permet pas de rassembler tous ceux formés à partir du radical, puis ceux en *-ed*, puis ceux en *-ing*. Cette question se pose aussi pour les seconds éléments, tels que *room* qui, nous l'avons vu, a deux sens : celui de la « place » et celui de la « pièce ». Il faudrait les distinguer.

Nous reprenons la suggestion de Tournier (1988) que nous avons cité dans le chapitre 3 (3.2.3.2) qui propose que l'on classe les composés selon leur second élément car celui-ci est le référent dans la plupart des cas, tout en prévoyant un renvoi au premier. Ceci permettrait de faire des rapprochements entre différentes formations.

Les critères d'inclusion dans le dictionnaire

Nous devons aussi déterminer quels seraient les critères d'inclusion des entrées. Deux problèmes se posent : le premier est celui des indications concernant les terminaisons flexionnelles, l'inclusion ou non d'un mot au pluriel (quand celui-ci ne représente pas de terminaison mais une graphie à part), l'inclusion ou non des dérivés. Nous avons constaté que certains dictionnaires écartent tous les dérivés, mais incluent les pluriels irréguliers. Il nous semble que le lexique est le lieu où l'on inclut tout ce que l'on ne peut pas déduire à partir de règles. Comme le disent Roeper & Siegel (1978 : 211) :

> [...] there is a crucial distinction to be made here between the syntax and the lexicon. Syntax operates without a memory for sentences; we must generate each sentence we say. Lexical rules describe systematic entries in a mental lexicon to which we can refer once a word has been created. Aronoff therefore calls lexical rules "once-only" rules.

Pour cette raison, il est justifiable de n'inclure que les pluriels irréguliers, et si le schéma accentuel de toute une catégorie de composés relève de règles fiables, alors l'inclusion de la catégorie n'est pas non plus nécessaire. Toutefois, nous sommes tenue de garder l'objectif suivant : le dictionnaire est à l'intention de non-anglophones qui ont besoin de connaître la prononciation d'un composé. Nous allons donc vouloir en mettre, mais lesquels ?

Les critères d'inclusion des entrées concernent aussi le choix des mots, d'usage plus ou moins fréquent, ayant des prononciations plus ou moins prévisibles, et des sens qui peuvent ou non se deviner à partir des éléments qui les constituent. La question de la lexicalisation de ces entrées est soulevée par Maniez (2002 : 173) :

> [...] des adjectifs comme *high-ranking* ou *government-sponsored*, qui sont d'emploi très fréquent ne sont pas lexicalisés sans doute parce que des gloses telles que *of high rank* et *sponsored by the government* rendent ces formes composées immédiatement interprétables.

Nous avons très souvent fait remarquer dans le chapitre 4 l'arbitraire qui gouverne l'inclusion ou l'omission de certaines entrées selon le dictionnaire. Si certaines entrées font leur apparition pour des raisons d'usage (*greenhouse effect*) et d'autres se perdent pour la même raison (*talking-film*), les explications de ces choix ne sont pas toujours évidentes : *greenhouse gas* ne figure que dans EPD15, *bad debt* dans LPD00, *staff room* et *racing bike* dans aucun des deux. Comme nous l'avons signalé au début du chapitre 2, la composition est un processus très productif et c'est pourquoi il est difficile d'établir des critères objectifs et stables dans un domaine qui est si créatif, et instable. Ces difficultés ont été signalées par Lutstorf (1960 : 144-5) :

> The possible number of combinations of the *stone-wall* and *mountain-path* type, which can be made up freely at any time, is almost infinite and no dictionary can give more than a comparatively small number of them. This does not matter much, since their meaning can always be inferred from the meaning of the elements. It rests within the lexicographer's own discretion how many and which of them he wants to enter in his dictionary.

Tout dictionnaire a des contraintes de place : une sélection, forcément arbitraire dans une certaine mesure, s'impose.

Les indications concernant la prononciation et le schéma accentuel

Prise en compte du discours et de la variation

En ce qui concerne la prononciation prévue des entrées lexicales, nous avons souligné la nécessité de prendre en compte le discours dans la mesure du possible. Pour Boisson (1980), le contexte du discours explique la grande

instabilité des adjectifs composés; Kingdon (1958b: 147) a aussi conscience du problème:

> [...] in sorting compounds into stress-types it is important to get at the lexical stress, which in some cases is obscure because the compound happens nearly always to be used in contexts which cause a stress adjustment.

En effet, l'indication qu'une entrée est susceptible de subir un déplacement de l'accent s'est avérée très pertinente dans notre analyse des extraits sonores; il est difficile de prévoir d'autres types de changements dans la mesure où les effets de courbes intonatives et des contextes émotionnels ont une influence sur la manière dont les corrélats acoustiques doivent s'interpréter plutôt qu'un impact sur le schéma lui-même.

En revanche, il est vrai que certaines entrées ont un schéma instable, sans que cette variation soit associée à un contexte syntaxique; il est donc souhaitable de l'indiquer de façon claire et cohérente. Une de nos difficultés pour tirer des conclusions à partir des variantes est la manière dont elles sont indiquées: ce sont les entrées de N-COMPACT qui fournissent les indications les plus claires car la version papier du dictionnaire LPD donne l'entrée deux fois en toutes lettres avec un schéma différent. Dans la saisie de cette table, nous avons indiqué une seule entrée mais avec un schéma dans le champ « Schéma » et l'autre correspondant à la variante dans le champ « Variation ». Cette notation représente peut-être une solution efficace.

Indications de la phonie

Enfin, en ce qui concerne les indications de la phonie et de l'accent des éléments isolés, il est évident que le statut indépendant d'un tel dictionnaire

entraînerait une redondance des informations qui n'est pas envisageable. Les dictionnaires EPD et LPD d'aujourd'hui ne marquent que le schéma accentuel sur les composés dans les sous-entrées, en utilisant les symboles IPA de l'accent devant la graphie de la syllabe accentuée. La phonie complète de chaque élément se trouve sous l'entrée indépendante. Cette solution nous semble satisfaisante et pourrait s'étendre pour comprendre tous les composés (dans les entrées principales et les sous-entrées). Il reste à déterminer le nombre de degrés d'accent à indiquer, ainsi que la forme que doit prendre cette notation.

Notation du schéma accentuel

A la fin du chapitre 1, nous avons proposé une notation adaptée aux composés où nous utilisons /1/ pour l'accent nucléaire potentiel et /P/ à la place de /2/ pour le premier élément des composés à accentuation double. Ayant expérimenté cette méthode dans l'analyse acoustique, nous restons persuadée qu'il est nécessaire de distinguer l'accent nucléaire des autres accents. Cependant, cette différenciation ne se fait qu'en discours car – nous en avons eu des exemples – si le composé ne se trouve pas à la fin de l'unité tonale, il ne porte pas cet accent ; les deux accents seraient donc /P/. Il faut bien insister sur le fait qu'il s'agit d'un accent nucléaire potentiel.

En revanche, il semble que la différence entre l'accent secondaire (rythmique, marqué /2/) et l'accent du premier élément du composé à accentuation double (que nous avons marqué /P/) n'a pas de fondement sur le plan des corrélats acoustiques. Si nous avions fait cette distinction, c'est parce que le placement de ces accents ne se fait pas de la même manière, ni pour les mêmes raisons : un

accent secondaire est positionné par des contraintes rythmiques, qui sont sa raison d'être, alors que l'accent du premier élément d'un composé est beaucoup plus proche du battement rythmique du début d'un syntagme ; il se place en fonction de l'accentuation de l'élément à l'état isolé, et son existence signale un certain type de relation entre les éléments. Toutefois, il n'est peut-être pas nécessaire de préserver cette notation, qui ne facilite pas l'interprétation des schémas. Le symbole /R/ avait aussi été proposé pour remplacer /2/ lorsque l'accent primaire était noté /P/ ; cette notation est superflue.

Pour les composés à accentuation simple, nous avions débattu du bien fondé de l'accent /3/ que Wells a introduit dans LPD90 et qu'il a abandonné dans l'édition suivante. Ce débat pose deux questions : la notation des deuxièmes éléments monosyllabiques désaccentués, et celle des éléments plurisyllabiques désaccentués. Pour les monosyllabes, nous avons remarqué que la non-réduction du deuxième élément des composés à accentuation simple est une règle fiable ne présentant que quelques exceptions dont nous avons discuté dans le chapitre 4. Le fait que la voyelle ne se réduise pas ne nous semble pas une raison suffisante de marquer un accent, car si nous le faisions, cela supposerait que l'on indique également l'accent sur des verbes en *-ate*, par exemple. C'est l'accent tertiaire de Lilly & Viel (1998), non de Wells (voir 1.4.1.5). Toutefois, il faut noter que EPD15 marque /2/ sur les composés non compacts dont le second élément désaccentué est monosyllabique.

En effet, il nous semble que le statut de la voyelle non accentuée de ce type de composé est différent des mots en *–ate*, parce que cette non-réduction joue un

rôle d'identifieur sur le plan sémantique. Il s'agit de la désaccentuation de la syllabe et non d'une syllabe inaccentuée. Pour ces raisons, nous pensons que le fait de marquer le monosyllabe désaccentué peut se justifier, mais nous préférons distinguer cet accent de l'accent /2/. Contrairement à l'opposition /P/ et /2/ où nous n'avons pas trouvé de corrélats acoustiques distincts, les valeurs du F_0, et souvent de l'intensité, ne fonctionnent pas du tout de la même manière. Les courbes s'aplatissent après la courbe souvent descendante sur l'accent primaire, alors que pour l'accent /2/ nous avons des pics et une courbe parallèle à celle de la syllabe nucléaire. Néanmoins, du fait de la convention générale des dictionnaires de ne pas marquer l'accent sur les monosyllabes, il nous a semblé qu'il n'était pas nécessaire d'innover sur ce point. En revanche, pour les éléments plurisyllabiques, il existe une proéminence relative d'une des syllabes de l'élément qu'il faut indiquer. Pour les raisons que nous avons citées ci-dessus, nous préférons employer l'accent /3/ plutôt que l'accent /2/ pour marquer cette proéminence.

Le rôle du lexicographe

En étudiant les questions qui se posent pour la création d'un dictionnaire de composés, nous apprécions le rôle essentiel du lexicographe qui sélectionne les entrées, leur classement, les prononciations. Nous avons tenté de mettre en relief des faits gênants dans l'actuelle présentation des composés, mais la tâche est délicate. Nous n'avons pas oublié les difficultés que nous avons eues à trouver des critères de sélection pour notre base de données et la question de la définition de ce qu'est un composé reste encore ouverte. La conclusion de Lutstorf qui a fait une analyse acoustique des composés est la suivante (Lutstorf, 1960 : 156) :

> Compound stress patterns as given by dictionaries must be regarded as reflections of the lexicographer's own opinion or usage. Foreign students will of course do well to follow such examples. But in the matter of compound stressing vacillations are bound to occur; no wonder therefore that different dictionaries appear to disagree. They can indeed not do better as long as the present situation which we have described prevails.

La nôtre est moins pessimiste. Même s'il est vrai que les indications d'un dictionnaire ne peuvent pas être parfaites, car des entrées manquent, et les locuteurs varient, nous avons vu que les règles d'accentuation permettent d'expliquer les différents schémas, et l'analyse des extraits sonores a permis de vérifier ces hypothèses, tout en démontrant la nécessité d'une prise en compte du discours pour interpréter les résultats. Les indications des dictionnaires de prononciation essaient justement de faire cela.

Il est évident qu'un dictionnaire des seuls composés n'est pas une solution réaliste, mais ces entrées pourraient constituer une annexe du dictionnaire de prononciation, classée selon les critères que nous avons proposés. Ce classement permettrait de mettre en relief les liens complexes et variés qui existent entre les différents éléments. Si une version électronique du dictionnaire comportait en plus une indication de la frontière morphologique, des recherches pourraient se faire indifféremment à partir du premier ou du deuxième élément. Elle constituerait de la sorte un outil d'analyse puissant.

GLOSSAIRE

Glossaire sélectif

Accent lexical : schéma accentuel prévu par le dictionnaire.

Accent nucléaire : accent porté par le noyau intonatif de l'unité.

Accent primaire : accent /1/, potentiellement l'accent nucléaire (position isolée), indiqué par / ˈ/ en IPA.

Accent secondaire : accent /2/, accent rythmique porté par l'une des deux premières syllabes d'un mot, lorsque l'accent /1/ se situe au delà, indiqué par /ˌ/ en IPA.

Accent tertiaire : accent /3/ porté par la syllabe accentuée du deuxième élément plurisyllabique d'un composé à accentuation simple, utilisé dans LPD90 et indiqué /ₒ/ en IPA.

Accentuation double : type de schéma accentuel d'un composé lorsque les deux éléments portent un accent, l'accent primaire étant sur le deuxième : abréviation « d ».

Accentuation simple : type de schéma accentuel d'un composé lorsque le premier élément porte l'accent primaire, abréviation « s ».

ACCESS : base de données du pack Office Pro 2000 de Microsoft.

Composé à trait d'union : composé dont la graphie comporte un trait d'union.

Composé compact : composé dont la graphie n'a aucun séparateur.

Composé non compact : composé dont la graphie comporte un espace.

Composé réduplicatif : composé où le deuxième élément reprend le premier, parfois en changeant un segment (vocalique ou consonantique).

Construction cumulative : unité lexicale du discours qui est formée à partir de trois éléments au moins (pouvant servir dans d'autres unités semblables).

Corpus d'analyse : l'ensemble des entrées de la base de données, soumises à l'analyse.

Corpus oral d'analyse : les extraits sonores effectivement soumis à l'analyse.

Déclinaison : la baisse de la fréquence fondamentale à la fin de l'unité tonale ; « declination » en anglais.

EXCEL : tableur du pack Office Pro 2000 de Microsoft.

MPW : Macintosh Programmer's Workshop : éditeur de programmes permettant de faire des requêtes à partir de fichiers de données sur Macintosh Apple.

PRAAT : logiciel d'analyse acoustique (Boersma & Weenik, 1996).

BIBLIOGRAPHIE THÉMATIQUE

Bibliographie thématique

1 Sources primaires

1.1 Dictionnaires

Jones, Daniel, 1963, *An English Pronouncing Dictionary*, 12th edition, London, J. M. Dent & Sons Ltd., 539 p.

Jones, Daniel, 1997, *An English Pronouncing Dictionary*, 15th edition, edited by Peter Roach & James Hartman, Cambridge, Cambridge University Press, 559 p.

Wells, J. C., 1990, *Longman Pronouncing Dictionary*, Harlow, Longman, 802 p.

Wells, J. C., 2000, *Longman Pronouncing Dictionary*, Harlow, Longman, 869 p.

1.2 Documents oraux

Ensemble des extraits sonores de l'Agrégation : session 2000

Reporters, émission de BBC World, janvier 2001

Reporters, émission de BBC World, 21 mars 2001

Reporters, émission de BBC World, 28 mars 2001

Rowling, J. K., 1997, *Harry Potter and the Philosopher's Stone*, Bloomsbury, London, 223 p.

Rowling, J. K., 1997, *Harry Potter and the Philosopher's Stone*, lu par Stephen Fry, version CDROM, Cover to Cover.

2 Ouvrages généraux et de référence

2.1 Linguistique générale

Benveniste, Emile, 1974, *Problèmes de linguistique générale II*, Paris, Gallimard, 286 p.

Bouscaren, C. et al., 1995, *Le Petit Ophrys*, Paris, Ophrys, 794 p.

Crystal, David, 1995, *The Cambridge Encyclopaedia of the English Language*, London, Cambridge University Press, BCA, 489 p.

Fowler, H. W., 1996, *The New Fowler's Modern English Usage*, ed. R. W. Burchfield, 3ème edition, Oxford, Clarendon Press, 864 p.

Fowler, H. W., 1952, *A Dictionary of Modern English Usage*, 1ère édition réditée, Oxford, Clarendon Press, 742 p.

Jespersen, O., 1961-65, *A Modern English Grammar on Historical Principles*, vol.VI, London, Allen & Unwin / Copenhagen, Ejnar Munksgaard, 570 p.

Mossé, G., 1945, *Manuel de l'anglais du moyen-âge. I. Vieil anglais*, Paris, Aubier-Montaigne, 383 p.

Nesfield, J.-C., 1939, *Manual of English Grammar and Composition*, London, Macmillan & Co. Ltd., 423 p.

2.2 Dictionnaires

Chambers 21st Century Dictionary, 1996, Chambers Harrap Publishers Ltd., 1654 p.

Chambers English Dictionary, 1988, 7th edition, Chambers/Cambridge University Press, 1793 p.

Concise Oxford Dictionary of Current English, 1996, 9th edition, edited by Della Thompson, London, BCA, 1673 p.

CUVOALD, 3rd edition 1974, Computer version by Roger Mitton, Birkbeck College, London, Oxford Text Archive, ftp://ota.ox.ac.uk/pub/ota/public/dicts/710.

Jones, Daniel, 1967, *An English Pronouncing Dictionary*, 13th edition, edited by A. C. Gimson, London, J. M. Dent & Sons Ltd., 544 p.

Jones, Daniel, 1977, *An English Pronouncing Dictionary*, 14th edition, edited by Susan Ramsaran, London and Melbourne, J. M. Dent & Sons Ltd., 576 p.

2.3 Manuels de phonologie et de phonétique de l'anglais

Deschamps, Alain, Duchet, Jean-Louis, Fournier, Jean-Michel, O'Neil, Michael, 2000, *Manuel de phonologie de l'anglais*, Paris, Didier Erudition CNED, 232 p.

Duchet, Jean-Louis, [1981] 1998, *La phonologie*, Que sais-je?, Paris, Presses Universitaires de France, 5ème édition, 127 p.

Gimson, A. C., 1980, *An Introduction to the Pronunciation of English*, London, Ed. Arnold, 3rd & 4th editions, 360 p.

Ginésy, M., 1995, *Mémento de phonétique anglaise*, Paris, Nathan université, 288 p.

Guierre, Lionel, 1984, *Drills in English Stress Patterns*, 4ème éd., Paris, A. Colin Longman, 128 p.

Ladefoged, Peter, 1982, *A Course in Phonetics*, Second edition, Fort Worth, Harcourt Brace, 302 p.

Ladefoged, Peter, 1993, *A Course in Phonetics*, Third edition, Fort Worth, Harcourt Brace, 308 p.

Lilly, Richard & Viel, Michel, [1975] 1993, *Initiation raisonnée à la phonétique de l'anglais*, Paris, Hachette Supérieur, 206 p.

Lilly, Richard & Viel, Michel, [1977] 1993, *La prononciation de l'anglais. Règles phonologiques et exercices de transcription*, Paris, Hachette Supérieur, 128 p.

Pountney, M. A., [1972] 1998, *A Practical Approach to English Phonetics*, Département d'anglais, Université de Limoges, 69 p.

Pountney, M. A., 1999, *A Practical Approach to English Phonetics & Phonology*, Limoges, PULIM, 76 p.

Roach, Peter, 1992, *Introducing Phonetics*, London, Penguin, 128 p.

Viel, Michel, [1981] 1994, *La Phonétique de l'anglais*, Que sais-je?, Paris, Presses Universitaires de France, 5ème édition, 128 p.

3 Ouvrages et articles spécialisés

3.1 Phonétique acoustique

3.1.1 Phonétique acoustique : ouvrages

Beckman, M., 1986, *Stress and non-stress accent*, Dordrecht, Foris Publications 1986, 239 p.

Bucellato, P., 1985, *Contribution à l'étude des rapports entre réalisations accentuelles et formes intonatives en anglais à partir d'une analyse expérimentale*, thèse, Strasbourg, 299 p.

Dujardin, S., 2001, *L'emphase dans le discours spontané anglais : corrélats acoustiques et prosodiques*, thèse, Doctorat nouveau régime, mais non publiée, Université Aix-Marseille I, 391 p.

Johnson, K., 1997, *Acoustic and Auditory Phonetics*, Oxford, Blackwell, 169 p.

Ladefoged, Peter, 1962, *Elements of Acoustic Phonetics*, Chicago, University of Chicago Press, 120 p.

Ladefoged, Peter, 1971, *Preliminaries to Linguistic Phonetics*, Chicago, University of Chicago Press, 122 p.

Ladefoged, Peter, 2001, *Vowels and Consonants*, Oxford, Blackwell, 191 p.

3.1.2 Phonétique acoustique : recueils de publications et articles

Allen, George, 1972. "The location of English stress beats in English: An experimental study", *Language and Speech*, vol. 15, part 1 pp. 72-102.

Bailly, G., Benoît, C., ed., 1992, *Talking Machines: Theories, Models and Designs*, Amsterdam, North-Holland Elsevier Science Publishers B. V., 350 p.

Beckman, M. & Edwards, J., 1994, "Articulatory evidence for differentiating stress categories", in Keating, P. A., (eds.) *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, pp. 1-33.

Beckman, M. & Pierrehumbert J., 1992, "Comments on chapters 14 and 15", in Docherty, G. & Ladd, D.R., eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, Cambridge, Cambridge University Press, p. 387.

van den Berg, R., Gussenhoven, C., Rietveld, T., 1992, "Downstep in Dutch: implications for a model", in Docherty, G. & Ladd, D.R., eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, Cambridge, Cambridge University Press, pp. 335-359.

Brown, W. & McGlone, R., 1974, "Aerodynamic and acoustic study of stress in sentence productions", *Journal of the Acoustical Society of America*, vol. 56, N° 3, pp. 971-974.

Campbell, N., 1992, "Syllable-based segmental duration", in Bailly, G., Benoît, C., ed., 1992, *Talking Machines: Theories, Models and Designs*, Amsterdam, North-Holland Elsevier Science Publishers B. V., pp. 211-224.

Collier, R., 1992, "A Comment on the Prediction of Prosody", in Bailly, G., Benoît, C., ed., 1992, *Talking Machines: Theories, Models and Designs*, Amsterdam, North-Holland Elsevier Science Publishers B. V., pp. 205-208.

Diana, A., 1994, "Panorama des recherches sur la perception de la prosodie: de la perception auditive à la compréhension", in J.-L. Duchet, J.-M. Fournier, J. Humbley et p. Larreya (éds), *7ème Colloque d'avril sur l'anglais oral 1994 : les contraintes en phonologie*, ALOES, Université de Paris-Nord, pp. 25-62.

Docherty, G & Ladd, D.R., eds., 1992, *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, Cambridge, Cambridge University Press, 462 p.

Fry, D. B., 1955, "Duration and intensity as physical correlates of linguistic stress", *Journal of the Acoustical Society of America* 27/4,: pp. 765-768

Fry, D. B., 1958, "Experiments in the perception of stress", *Language and Speech* 1: pp. 126-152.

Gussenhoven, C. & Rietveld, A., 1988, "Fundamental frequency declination in Dutch: testing three hypotheses", Journal of Phonetics, 16, pp. 355-369.

Keating, P. A., (eds.), 1994, *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, 462 p

Ladd, D. R., 1992, "An Introduction to Intonational Phonology", in Docherty, G. & Ladd, D.R., eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, Cambridge, Cambridge University Press, pp. 321-368.

Ladd, D. R., 1993, "Notes on the Phonology of Prominence", *Working papers 41, Proceedings of the ESCA workshop on prosody*, University of Lund, Sweden, pp. 10-15.

Lieberman, Phillip, 1960, "Some acoustic correlates of word stress in American English". *Journal of the Acoustical Society of America*. 22, pp. 451-454.

Nolan, F.. 1994, "Phonetic correlates of syllable affiliation", in Keating, P. A , (eds.) *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, pp. 160-167.

Sawusch, J. & Turk, A., "The processing of duration and intensity clues to prominence", *Journal of the Acoustical Society of America*, Vol. 99, N° 6, pp. 3782-3790.

Shattuck-Hufnagel, S, 1994, "Stress shift as early placement of pitch accents: a comment on Beckman & Edwards", in Keating, P. A., (eds.) *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, pp. 34-42.

Sluijter, A. & van Heuven, V., 1993, "Perceptual clues of linguistic stress: intensity revisited", *Working papers 41, Proceedings of the ESCA workshop on prosody*, University of Lund, Sweden, pp. 246-249.

Terken, J., 1991, "Fundamental frequency and perceived prominence of accented syllables", *Journal of the Acoustical Society of America* 894/1: 1768-1776.

Turk A., 1994, "Articulatory phonetic clues to syllable affiliation: gestural characteristics of bilabial stops", in Keating, P. A., (eds.) *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, pp. 107-135.

3.2 Phonétique et phonologie

3.2.1 Phonétique et phonologie générales

3.2.1.1 Phonétique et phonologie générales: ouvrages

Chomsky, Noam & Halle, Morris, 1968, *The Sound Pattern of English*, New York, Harper & Row, 470 p.

Garde, Paul, 1968, *L'Accent*, Paris, Presses Universitaires de France, 128 p.

Goldsmith, John A., 1990, *Metrical Phonology*, Oxford, Basil Blackwell, 376 p.

Gussenhoven, Carlos & Jacobs, Haike, 1998, *Understanding Phonology*, Arnold, London, 286 p.

O'Connor, J. D., 1991 (1973), *Phonetics*, Penguin, 128 p.

Roca, Iggy, 1994, *Generative Phonology*, London, Routledge, 307 p.

Selkirk, E., 1982, *The Syntax of Words*, MIT Press, Cambridge, MA, 471 p.

Selkirk, E., 1984, *Phonology and Syntax: The Relation between Sound and Structure*, MIT Press, Cambridge, 476 p.

3.2.1.2 Phonétique et phonologie générales : recueils de publications & articles

Blevins, J., 1995 "The Syllable in Phonological Theory", in Goldsmith, John A., ed., *The Handbook of Phonological Theory*, Oxford, Blackwell, pp. 200-235

Goldsmith, John A., ed., 1995, *The Handbook of Phonological Theory*, Oxford, Blackwell, 986 p.

Goyvaerts, D. L. and Pullum, G. K., ed., 1975, *Essays on the Sound Pattern of English*, Gand, Story-Scientia, 580 p.

Liberman & Prince, 1977, "On stress and linguistic rhythm", *Linguistic Enquiry* 8, pp. 249-336.

Nespoulos, J.-L., ed., 1993, *Tendances actuelles en linguistique générale*, Neuchâtel : Delachaux et Niestlé, 199 p

Paradis, C., 1993, "Phonologie générative multilinéaire", in Nespoulos, J.-L., 1993 ed. *Tendances actuelles en linguistique générale*, Neuchâtel : Delachaux et Niestlé, pp. 11-57.

Troubetzkoy, N. S., 1949, "Oristique, Etude de la fonction délimitative", in *Principes de phonologie*, trad. J. Cantineau, Paris, Klincksieck, 396 p.

3.2.2 Phonétique et phonologie de l'anglais

3.2.2.1 Phonétique et phonologie de l'anglais : ouvrages

Duchet, Jean-Louis, [1991] 1994, *Code de l'anglais oral*, Paris, Ophrys, 2ème édition, 124 p.

Fudge, E., 1984, *English Word Stress*, London, Allen & Unwin, 240 p.

Giegerich, Heinz J., 1992, *English Phonology : an Introduction*, Cambridge, CUP, 333 p.

Gray, M., 2001, *Place et rôle de l'intonation dans l'apprentissage de l'anglais*, Thèse, Doctorat nouveau régime, mais non publiée, Université Paris XII, 688 p.

Guierre, Lionel, 1979, *L'accentuation en anglais contemporain. Eléments pour une synthèse*, Thèse de doctorat d'Etat, Université Paris VII, 878 p.

Halle, Morris & Keyser, Samuel Jay, 1971, *English Stress*, Harper & Row, New York, 206 p.

Hammond, Michael, 1999, *The Phonology of English*, Oxford, OUP, 368 p.

Huart, Ruth, 2002, *Grammaire orale de l'anglais*, Ophrys, Paris, 122 p.

Jensen, J., 1993, *English Phonology. Current Issues in Linguistic Theory*, Vol. 99, Amsterdam/Philadelphia, John Benjamins Publishing Company, 251 p.

Kingdon, Roger, 1958a, *The Groundwork of English Intonation*, London, Longman, 272 p.

Kingdon, Roger, 1958b, *The Groundwork of English Stress*, London, Longman, 224 p.

Montreuil J.-P., 2001, *La Phonologie de l'anglais*, Rennes, Presses Universitaires de Rennes, 168 p.

Nicaise, A. et Gray, G., 1998, *L'Intonation de l'anglais*, Paris, Nathan, 128 p.

Poldauf, I., 1984, *English Word Stress*, Oxford, Pergamon, 169 p.

Roach, Peter, [1983] 1991, *English Phonetics and Phonology*, Cambridge, Cambridge University Press, 262 p.

Zumstein, F., thèse en cours, Doctorat nouveau régime, Université de Poitiers, à paraître.

3.2.2.2 Phonétique et phonologie de l'anglais : recueils de publications & articles

Abercrombie, D. & al., 1964, In Honour of Daniel Jones, Papers contributed on the occasion of his eightieth birthday 12-9-1961, Longman, 474 p.

Abercrombie, D., 1964, "Syllable Quantity and Enclitics in English", in Abercrombie, D. & al., *In Honour of Daniel Jones. Papers contributed on the occasion of his eightieth birthday 12-9-1961*, Longman, pp. 216-222.

Bolinger, Dwight, 1972, "Accent is Predictable (If You're a Mind-Reader)", in *Language*, n° 48, pp. 633-644.

Boulle, Jacques, 1984, "Suffixes latins et accentuation anglaise : aperçus sur une évolution historique", in M. Cling & J. Humbley (éds), *2ème Colloque d'avril sur l'anglais oral 1984: le suprasegmental 1984*, ALOES, Université de Paris-Nord, pp. 207-216.

Cooke, R., 1994, "Developing awareness of word stress in ESP students. Constraints and suggestions", in J.-L. Duchet, J.-M. Fournier, J. Humbley et P. Larreya (éds), *7ème Colloque d'avril sur l'anglais oral 1994 : les contraintes en phonologie*, ALOES, Université de Paris-Nord, pp. 189-204.

Crépin, André, 1982, "La syllabe en anglais", in M. Cling & J. Humbley (éds), *Colloque d'avril sur l'anglais oral 1982, réalités et didactique*, ALOES, Université de Paris-Nord, pp. 55-62.

Duchet, J.-L. & Fournier J.-M., 1988, "Isomorphisme et productivité dans l'accentuation et la prononciation des mots dérivés anglais", in M. Cling & J. Humbley (éds), *Quatrième colloque d'avril sur l'anglais oral 1988: le suprasegmental II*, ALOES, Université de Paris-Nord, pp. 111-122.

Duchet, J.-L.,1994, "Eléments pour une histoire de l'accentuation lexicale en anglais", *Etudes anglaises*, 47/2, Paris Didier, pp. 161-170.

Duchet, J.-L., 1993, "La coupe syllabique en anglais", *Etudes de linguistique et de littérature en l'honneur d'André Crépin*, Reinelle-Verlag-Greifswald, Wodan, pp. 137-145.

Duchet, J.-L., 1993, "Les marques de la motivation. Des noms d'animaux aux relations dérivationnelles", *Faits de Langues n° 1*, pp. 241-249.

Duchet, J.-L., October 1991, "John Wells, Longman Pronunciation Dictionary", *English Today 28*, pp. 52-54.

Duchiron, Emmanuelle, 2000, *L'accent tertiaire en anglais, étude de concept*, dossier de phonologie de maîtrise, Université de Poitiers, 21 p.

Durand, J., 1998, "Les traits phonologiques et le débat articulation / audition", in *Actes du colloque de décembre 1998 : Phonétique et Phonologie de l'anglais*, Pau, PUP, pp. 56-70.

Fournier, J.-M., 1992, "Théorie et enseignement : l'accent en 10 questions", in J.-L. Duchet, J.-M. Fournier, J. Humbley et p. Larreya (éds), *6ème Colloque d'avril sur l'anglais oral 1992*, ALOES, Université de Paris-Nord, pp. 73-97.

Fudge, E., 1994, "Higher level prosodic units and the rhythm of spoken English", in J.-L. Duchet, J.-M. Fournier, J. Humbley et p. Larreya (éds), *7ème Colloque d'avril sur l'anglais oral 1994 : les contraintes en phonologie*, ALOES, Université de Paris-Nord, pp. 7-24.

Selkirk, E, 1995, "Sentence prosody: Intonation, Stress and Phrasing", in Goldsmith, John A., ed. 1995, *The Handbook of Phonological Theory*, Oxford, Blackwell, pp. 562-566.

Taylor, D., 1981, "Non-native speakers and the rhythm of English", *International Review of Applied linguistics in Language Teaching*, Vol. 19, pp. 218-226.

Wells, J. C., 2000, "Pronunciation Preferences in British English: A New Survey", *Dixième colloque d'avril sur l'anglais oral*, Villetaneuse, à paraître.

3.2.3 Phonétique et phonologie : la composition

3.2.3.1 Phonétique et phonologie : ouvrages sur la composition

Boisson, Claude, 1980, *L'accentuation des composés en anglais contemporain*, Thèse de doctorat d'Etat, Université Paris VII, 824 p.

Huart Friedlander, Ruth, 1984, *La Composition nominale en anglais: opérations de repérage et accentuation*, Thèse de 3ème cycle, Université Paris VII, 256 p.

Lutstorf, H., 1960, *The Stressing of Compounds in Modern English: A Study in Experimental Phonetics*, Thèse, Université de Zurich, 159 p.

Thun, Nils, 1963, *Reduplicative Words in English. A Study of Formations of the Types 'Tick-tick', 'Hurly-burly' and 'Shilly-shally'*, Lund, Carl Bloms Boktryckeri, 347 p.

Tsiranana, J. H., 1981, *Recherches sur la formation et accentuation des composés quasi-morphématiques simples et suffixés en anglais contemporain*, Thèse de doctorat de 3ème cycle, Université Paris VII, 420 p.

3.2.3.2 Phonétique et phonologie : recueils de publications et articles sur la composition

Bauer, L., 1978, "Stress in Compounds: a Rejoinder", *English Studies*, 64, 1, pp. 47-53.

Bauer, L., 1998, "When is a sequence of two nouns a compound in English?", *English Language and Linguistics*, CUP, Vol. 2, part 1, pp. 65-86.

Boisson, Claude, 1984, "L'accentuation des composés anglais", in M. Cling & J. Humbley (éds), *2ème Colloque d'avril sur l'anglais oral : le suprasegmental 1984*, ALOES, Université de Paris-Nord, pp. 165-178.

Cinque, G., 1993, "A null theory of phrase and compound stress", *Linguistic Inquiry*, 24, 2, pp. 239-297.

Gautier, André (ed.), 1989, *Explorations en linguistique anglaise*, Berne, Peter Lang, 248 p.

Guierre, Lionel, 1990, "Mots composés anglais et agrégats consonantiques", in J.-L. Duchet, J.-M. Fournier, J. Humbley et p. Larreya (éds), *5ème Colloque d'avril sur l'anglais oral 1990*, ALOES, Université de Paris-Nord, pp. 59-72.

Huart Friedlander, Ruth, 1979, "Fonction de l'accent dans les noms composés anglais", *Cahiers Charles V, n° 1: Linguistique*, pp. 7-15.

Huart Friedlander, Ruth, 1986, "L'Accentuation des génitifs dits 'génériques'", *Troisième colloque d'avril sur l'anglais oral: De la communication à la théorie*, Villetaneuse, pp. 199-210.

Huart Friedlander, Ruth, "Nouveau regard sur les noms composés" in Gautier, André (ed.), 1989, *Explorations en linguistique anglaise*, Berne, Peter Lang, pp. 25-90.

Gibbon, D & Richter, R. (ed.), *Intonation, Accent and Rhythm*, Berlin, Walter de Gruyter, 350 p.

Ladd, D. R. Jr, 1984, "English Compound Stress", in Gibbon, D & Richter, R. (ed.), *Intonation, Accent and Rhythm*, Berlin, Walter de Gruyter, pp. 253-266.

Sampson, Rodney, 1980, "Stress in English N+N phrases: a further complicating factor", *English Studies*, 61, pp. 264-270.

3.3 Lexicologie et morphologie

3.3.1 Lexicologie et morphologie générales : ouvrages et articles

3.3.1.1 Lexicologie et morphologie générales: ouvrages

Adams, V., 1973, *An Introduction to Modern English Word Formation*, London, Longman, English Language Series, 238 p.

Bauer, L., 1983, *English Word-formation*, Cambridge, C.U.P., 311 p.

Bauer, L., 1988, *Introducing Linguistic Morphology*, Edinburgh: Edinburgh University Press, 272 p.

Lyons, J., 1977, *Semantics*, Vol. 2, Cambridge, CUP, 897 p.

Marchand, Hans, 1969, *The Categories and Types of Present-Day English Word Formation*, Second Edition, Munich: C. H. Beck'sche Verlagsbuchhandlung, 545 p.

Paillard, M., 2000, *Lexicologie contrastive anglais-français*, Paris, Ophrys, 199 p.

Picoche, J., 1992, *Précis de lexicologie française*, Poitiers, Nathan université, 191 p.

Tournier, Jean, 1985, *Introduction descriptive à la lexicogénèse de l'anglais contemporain*, Champion Slatkine, Paris-Genève, 517 p.

Tournier, Jean, 1988, *Précis de lexicologie anglaise*, Nathan, Paris, 208 p.

Trevian, I., 2000, "Typologie et morphologie, des constructions savantes, ou un fonds en constant accroissement du lexique anglais", *Dixième colloque d'avril sur l'anglais oral*, Villetaneuse, à paraître.

3.3.1.2 Lexicologie et morphologie générales: recueils de publications et articles

Bauer L. & A. Renouf, 2001, "A Corpus-Based Study of Compounding in English", *Journal of English Linguistics*, Sage Publications, Vol. 29 / No. 2, pp. 101-123

Villiard, P. 1993, "Morphologie : tendances actuelles", in Nespoulos, J.-L., 1993 ed. *Tendances actuelles en linguistique générale*, Neuchâtel : Delachaux et Niestlé, pp. 49-82.

3.3.2 Lexicologie et morphologie : la composition

3.3.2.1 Recueils de publications et articles sur la composition

Albrespit, Jean, 1998, "Quelques procédés de formation des agrégats nominaux en vieil-anglais", in *Recherches en linguistique étrangère XIX*, vol. 29, Mélanges Jean Tournier, Annales littéraires de l'Université de Franche-Comté, pp. 15-29.

Boucher, Paul, 1992, "Teaching Compounds in English: an Application of the Charlie Brown Principle" in *L'Ordre des mots*, CIEREC, Université de Saint-Etienne, pp. 33-50.

Boucher, Paul, 1998, "Les composés déverbaux en anglais: problèmes d'apprentissage", *Journée de Travail du CELEX: Travaux en Cours*, Angers, le 28/02/98.

Corbin, D., 1992, "Hypothèses sur les frontières de la composition nominale", *Cahiers de grammaire*, 17, pp. 27-55.

Corbin, D., 1997, "Locutions, composés, unités polylexématiques : lexicalisation et mode de construction", in Martins-Baltar M. (ed.) *La locution, entre langue et usages*, Fontenay-aux Roses, ENS Editions, pp. 53-102

Downing, P., 1977, "On the Creation and Use of English Compound Nouns", *Language*, n° 53, pp. 810-842.

Gross, G. 1988, "Degré de figement des noms composés", *Langages* 90, pp. 57-72

Kuiper, Koenraad, 1999, "Compounding by adjunction and its empirical consequences", *Language Sciences*, 21, pp. 407-422.

Lieber, R., "Argument Linking and Compounds in English", *Linguistic Inquiry*, 14, pp. 251-286.

Maniez, F., "Les mots composés à trait d'union dans la presse américaine : une étude diachronique basée sur l'usage de corpus", *Cahiers du Ciel*, Paris VII, pp. 162-175.

Paillard, M., 2002, "Quelques exemples de présentation des faits de langue", *Rapport de l'Agrégation interne*, à paraître.

Roeper, T. & Siegel, M., 1978, "A Lexical Transformation for Verbal Compounds", *Linguistic Inquiry*, Vol. 9, N° 2, pp. 199-260.

4 Logiciels

4.1 Programmes et données

AUDACITY, http://audacity.sourceforge.net

Guierre, L., New Pronouncing Dictionary, http://www.mshs.univ-poitiers.fr/Forell/Guierre/index.htm, site créé en août 2002.

MPW. Mackintosh Programmer's Workshop, août 1998, ftp://ftp.apple.com/developer/Tool_Chest/Core_Mac_OS_Tools/MPW_etc./

PRAAT, avril 2000 et septembre 2001, http://www.fon.hum.uva.nl/praat ou http://www.praat.org,

4.2 Documentation et aides

Boersma, P. & Weenik, D., 1996, "PRAAT: a system for doing phonetics by computer", *Report of the Institute of Phonetic Sciences of the University of Amsterdam*, p. 132.

Buchanan, T. & Eddy, C., 1999, *Microsoft Access 2000*, Paris, CampusPress, 460 p.

Wells, John, 1995, SAMPA: "Computer readable phonetic alphabet", http://www.phon.ucl.ac.uk/home/sampa/home.htm.

Wells, John, 1995, "Computer-coding the IPA: a proposed extension of SAMPA", http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm.

Wells, J. C., Barry, W., Grice, M, Fourcin, A., Gibbon, D., 1992, "Standard computer-compatible transcription", Esprit project 2589 (SAM), Doc. no. SAM-UCL-037, London, Phonetics and Linguistics Department, UCL.

Wood, S., 2000. *Praat beginner's manual*, http://www.ling.lu.se/persons/Sidney/

Zumstein, Franck, 1998, "SFI&GI, une interface de recherche sur des bases de données", in *Actes du colloque de décembre 1998 : Phonétique et Phonologie de l'anglais*, Pau, PUP, pp. 137-157.

Zumstein, Franck, 2000, *Les Corpus lexico-phonétiques numérisés et leur contribution à l'étude de la variation accentuelle et phonétique avec une étude particulière des mots en <-CiCe#>*, T.E.R. de D.E.A., Université de Poitiers, 94 p.

RÉFÉRENCES BIBLIOGRAPHIQUES

Références bibliographiques

Abercrombie, D. & al., 1964, In Honour of Daniel Jones, Papers contributed on the occasion of his eightieth birthday 12-9-1961, Longman, 474 p.

Abercrombie, D., 1964, "Syllable Quantity and Enclitics in English", in Abercrombie, D. & al., *In Honour of Daniel Jones, Papers contributed on the occasion of his eightieth birthday 12-9-1961*, Longman, pp. 216-222.

Adams, V., 1973, *An Introduction to Modern English Word Formation*, London, Longman, English Language Series, 238 p.

Albrespit, Jean, 1998, "Quelques procédés de formation des agrégats nominaux en vieil-anglais", in *Recherches en linguistique étrangère XIX*, vol. 29, Mélanges Jean Tournier, Annales littéraires de l'Université de Franche-Comté, pp. 15-29.

Allen, George, 1972. "The location of English stress beats in English: An experimental study", *Language and Speech*, vol. 15, part 1, pp. 72-102.

AUDACITY, http://audacity.sourceforge.net

Bailly, G., Benoît, C., ed., 1992, *Talking Machines: Theories, Models and Designs*, Amsterdam, North-Holland Elsevier Science Publishers B. V., 350 p.

Bauer, L., 1978, "Stress in Compounds: a Rejoinder", *English Studies*, 64, 1, pp. 47-53.

Bauer, L., 1983, *English Word-formation*, Cambridge, C.U.P., 311 p.

Bauer, L., 1988, *Introducing Linguistic Morphology*, Edinburgh: Edinburgh University Press, 272 p.

Bauer, L., 1998, "When is a sequence of two nouns a compound in English?", *English Language and Linguistics*, CUP, Vol. 2, part 1, pp. 65-86.

Bauer L. & A. Renouf, 2001, "A Corpus-Based Study of Compounding in English", *Journal of English Linguistics*, Sage Publications, Vol. 29 / No. 2, pp. 101-123

Beckman, M., 1986, *Stress and non-stress accent*, Dordrecht, Foris Publications 1986, 239 p.

Beckman, M. & Edwards, J., 1994, "Articulatory evidence for differentiating stress categories", in Keating, P. A., (eds.) *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, pp. 1-33.

Beckman, M. & Pierrehumbert J., 1992, "Comments on chapters 14 and 15", in Docherty, G. & Ladd, D.R., eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, Cambridge, Cambridge University Press, p. 387.

Benveniste, Emile, 1974, *Problèmes de linguistique générale II*, Paris, Gallimard, 286 p.

van den Berg, R., Gussenhoven, C., Rietveld, T., 1992, "Downstep in Dutch: implications for a model", in Docherty, G. & Ladd, D.R., eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, Cambridge, Cambridge University Press, pp. 335-359.

Blevins, J., 1995 "The Syllable in Phonological Theory", in Goldsmith, John A., ed., *The Handbook of Phonological Theory*, Oxford, Blackwell, pp. 200-235

Boersma, P. & Weenik, D., 1996, "PRAAT: a system for doing phonetics by computer", *Report of the Institute of Phonetic Sciences of the University of Amsterdam*, p. 132.

Boisson, Claude, 1980, *L'accentuation des composés en anglais contemporain*, Thèse de doctorat d'Etat, Université Paris VII, 824 p.

Boisson, Claude, 1984, "L'accentuation des composés anglais", in M. Cling & J. Humbley (éds), *2ème Colloque d'avril sur l'anglais oral : le suprasegmental 1984*, ALOES, Université de Paris-Nord, pp. 165-178.

Bolinger, Dwight, 1972, "Accent is Predictable (If You're a Mind-Reader)", in *Language*, n° 48, pp. 633-644.

Boucher, Paul, 1992, "Teaching Compounds in English: an Application of the Charlie Brown Principle" in *L'Ordre des mots*, CIEREC, Université de Saint-Etienne, pp. 33-50.

Boucher, Paul, 1998, "Les composés déverbaux en anglais: problèmes d'apprentissage", *Journée de Travail du COLEX: Travaux en Cours*, Angers, le 28/02/98.

Boulle, Jacques, 1984, "Suffixes latins et accentuation anglaise : aperçus sur une évolution historique", in M. Cling & J. Humbley (éds), *2ème Colloque d'avril sur l'anglais oral 1984: le suprasegmental 1984*, ALOES, Université de Paris-Nord, pp. 207-216.

Bouscaren, C. et al., 1995, *Le Petit Ophrys*, Paris, Ophrys, 794 p.

Brown, W. & McGlone, R., 1974, "Aerodynamic and acoustic study of stress in sentence productions", *Journal of the Acoustical Society of America*, vol. [illegible], N° 3, pp. 971-974.

Bucellato, P., 1985, *Contribution à l'étude des rapports entre réalisations accentuelles et formes intonatives en anglais à partir d'une analyse expérimentale*, thèse, Strasbourg, 299 p.

Buchanan, T. & Eddy, C., 1999, *Microsoft Access 2000*, Paris, CampusPress, 460 p.

Campbell, N., 1992, "Syllable-based segmental duration", in Bailly, G., Benoît, C., ed., 1992, *Talking Machines: Theories, Models and Designs*, Amsterdam, North-Holland Elsevier Science Publishers B. V., pp. 211-224.

Chambers 21st Century Dictionary, 1996, Chambers Harrap Publishers Ltd., 1654 p.

Chambers English Dictionary, 1988, 7th edition, Chambers/Cambridge University Press, 1793 p.

Chomsky, Noam & Halle, Morris, 1968, *The Sound Pattern of English*, New York, Harper & Row, 470 p.

Cinque, G., 1993, "A null theory of phrase and compound stress", *Linguistic Inquiry*, 24, 2, pp. 239-297.

Collier, R., 1992, "A Comment on the Prediction of Prosody", in Bailly, G., Benoît, C., ed., 1992, *Talking Machines: Theories, Models and Designs*, Amsterdam, North-Holland Elsevier Science Publishers B. V., pp. 205-208.

Concise Oxford Dictionary of Current English, 1996, 9th edition, edited by Della Thompson, London, BCA, 1673 p.

Cooke, R., 1994, "Developing awareness of word stress in ESP students. Constraints and suggestions", in J.-L. Duchet, J.-M. Fournier, J. Humbley et P. Larreya (éds), *7ème Colloque d'avril sur l'anglais oral 1994 : les contraintes en phonologie*, ALOES, Université de Paris-Nord, pp. 189-204.

Corbin, D., 1992, "Hypothèses sur les frontières de la composition nominale", *Cahiers de grammaire*, 17, pp. 27-55.

Corbin, D., 1997, "Locutions, composés, unités polylexématiques : lexicalisation et mode de construction", in Martins-Baltar M. (ed.) *La locution, entre langue et usages*, Fontenay-aux Roses, ENS Editions, pp. 53-102

Crépin, André, 1982, "La syllabe en anglais", in M. Cling & J. Humbley (éds), *Colloque d'avril sur l'anglais oral 1982, réalités et didactique*, ALOES, Université de Paris-Nord, pp. 55-62.

Crystal, David, 1995, *The Cambridge Encyclopaedia of the English Language*, London, Cambridge University Press, BCA, 489 p.

CUVOALD, 3rd edition 1974, Computer version by Roger Mitton, Birkbeck College, London, Oxford Text Archive, ftp://ota.ox.ac.uk/pub/ota/public/dicts/710.

Deschamps, Alain, Duchet, Jean-Louis, Fournier, Jean-Michel, O'Neil, Michael, 2000, *Manuel de phonologie de l'anglais*, Paris, Didier Erudition CNED, 232 p.

Diana, A., 1994, "Panorama des recherches sur la perception de la prosodie: de la perception auditive à la compréhension", in J.-L. Duchet, J.-M. Fournier, J. Humbley et p. Larreya (éds), *7ème Colloque d'avril sur l'anglais oral 1994 : les contraintes en phonologie*, ALOES, Université de Paris-Nord, pp. 25-62.

Docherty, G. & Ladd, D.R., eds., 1992, *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, Cambridge, Cambridge University Press, 462 p.

Downing, P., 1977, "On the Creation and Use of English Compound Nouns", *Language*, n° 53, pp. 810-842.

Duchet, J.-L., October 1991, "John Wells, Longman Pronunciation Dictionary", *English Today 28*, pp. 52-54.

Duchet, J.-L., 1993, "La coupe syllabique en anglais", *Etudes de linguistique et de littérature en l'honneur d'André Crépin*, Reineke-Verlag-Greifswald, Wodan, pp. 137-145.

Duchet, J.-L., 1993, "Les marques de la motivation. Des noms d'animaux aux relations dérivationnelles", *Faits de Langues n° 1*, pp. 241-249.

Duchet, J.-L.,1994, "Eléments pour une histoire de l'accentuation lexicale en anglais", *Etudes anglaises*, 47/2, Paris, Didier, pp. 161-170.

Duchet, Jean-Louis, [1991] 1994, *Code de l'anglais oral*, Paris, Ophrys, 2ème édition, 124 p.

Duchet, Jean-Louis, [1981] 1998, *La phonologie*, Que sais-je?, Paris, Presses Universitaires de France, 5ème édition, 127 p.

Duchet, J.-L. & Fournier J.-M., 1988, "Isomorphisme et productivité dans l'accentuation et la prononciation des mots dérivés anglais", in M. Cling & J. Humbley (éds), *Quatrième colloque d'avril sur l'anglais oral 1988: le suprasegmental II*, ALOES, Université de Paris-Nord, pp. 111-122.

Duchiron, Emmanuelle, 2000, *L'accent tertiaire en anglais, étude de concept*, dossier de phonologie de maîtrise, Université de Poitiers, 21 p.

Dujardin, S., 2001, *L'emphase dans le discours spontané anglais, corrélats acoustiques et prosodiques*, thèse, Doctorat nouveau régime, mais non publiée, Université Aix-Marseille I, 391 p.

Durand, J., 1998, "Les traits phonologiques et le débat articulation / audition", in *Actes du colloque de décembre 1998 : Phonétique et Phonologie de l'anglais*, Pau, PUP, pp. 56-70.

Extraits sonores de l'Agrégation : session 2000

Fournier, J.-M., 1992, "Théorie et enseignement : l'accent en 10 questions", in J.-L. Duchet, J.-M. Fournier, J. Humbley et p. Larreya (éds), *6ème Colloque d'avril sur l'anglais oral 1992*, ALOES, Université de Paris-Nord, pp. 73-97.

Fowler, H. W., 1952, *A Dictionary of Modern English Usage*, 1ère édition réditée, Oxford, Clarendon Press, 742 p.

Fowler, H. W., 1996, *The New Fowler's Modern English Usage*, ed. R. W. Burchfield, 3ème edition, Oxford, Clarendon Press, 864 p.

Fry, D. B., 1955, "Duration and intensity as physical correlates of linguistic stress", *Journal of the Acoustical Society of America* 27/4,: pp. 765-768

Fry, D. B., 1958, "Experiments in the perception of stress", *Language and Speech* 1: pp. 126-152.

Fudge, E., 1984, *English Word Stress*, London, Allen & Unwin, 240 p.

Fudge, E., 1994, "Higher level prosodic units and the rhythm of spoken English", in J.-L. Duchet, J.-M. Fournier, J. Humbley et p. Larreya (éds), *7ème Colloque d'avril sur l'anglais oral 1994 : les contraintes en phonologie*, ALOES, Université de Paris-Nord, pp. 7-24.

Garde, Paul, 1968, *L'Accent*, Paris, Presses Universitaires de France, 128 p.

Gautier, André (ed.), 1989, *Explorations en linguistique anglaise*, Berne, Peter Lang, 248 p.

Gibbon, D & Richter, R. (ed.), *Intonation, Accent and Rhythm*, Berlin, Walter de Gruyter, 350 p.

Giegerich, Heinz J., 1992, *English Phonology. an Introduction*, Cambridge, CUP, 333 p.

Gimson, A. C., 1980, *An Introduction to the Pronunciation of English*, London, Ed. Arnold, 3rd & 4th editions, 360 p.

Ginésy, M., 1995, *Mémento de phonétique anglaise*, Paris, Nathan université, 288 p.

Goldsmith, John A., 1990, *Metrical Phonology*, Oxford, Basil Blackwell, 376 p.

Goldsmith, John A., ed., 1995, *The Handbook of Phonological Theory*. Oxford, Blackwell, 986 p.

Goyvaerts, D. L. and Pullum, G. K., ed., 1975, *Essays on the Sound Pattern of English*, Gand, Story-Scientia, 580 p.

Gray, M., 2001, *Place et rôle de l'intonation dans l'apprentissage de l'anglais*, Thèse, Doctorat nouveau régime, mais non publiée, Université Paris XII. 688 p.

Gross, G. 1988, "Degré de figement des noms composés", *Langages* 90, pp. 57-72

Guierre, L., New Pronouncing Dictionary, http://www.mshs.univ-poitiers.fr/Forell/Guierre/index.htm, site créé en août 2002.

Guierre, Lionel, 1979, *L'accentuation en anglais contemporain. Eléments pour une synthèse*, Thèse de doctorat d'Etat, Université Paris VII, 878 p.

Guierre, Lionel, 1984, *Drills in English Stress Patterns*, 4ème éd., Paris, A. Colin Longman, 128 p.

Guierre, Lionel, 1990, "Mots composés anglais et agrégats consonantiques", in J.-L. Duchet, J.-M. Fournier, J. Humbley et p. Larreya (éds), *5ème Colloque d'avril sur l'anglais oral 1990*, ALOES, Université de Paris-Nord, pp. 59-72.

Gussenhoven, C. & Rietveld, A., 1988, "Fundamental frequency declination in Dutch: testing three hypotheses", Journal of Phonetics, 16, pp. 355-369.

Gussenhoven, Carlos & Jacobs, Haike, 1998, *Understanding Phonology*, Arnold, London, 286 p.

Halle, Morris & Keyser, Samuel Jay, 1971, *English Stress*, Harper & Row, New York, 206 p.

Hammond, Michael, 1999, *The Phonology of English*, Oxford, OUP, 368 p.

Huart Friedlander, Ruth, 1979, "Fonction de l'accent dans les noms composés anglais", *Cahiers Charles V, n° 1: Linguistique*, pp. 7-15.

Huart Friedlander, Ruth, 1984, *La Composition nominale en anglais: opérations de repérage et accentuation*, Thèse de 3ème cycle, Université Paris VII, 256 p.

Huart Friedlander, Ruth, 1986, "L'Accentuation des génitifs dits 'génériques'", *Troisième colloque d'avril sur l'anglais oral: De la communication à la théorie*, Villetaneuse, pp. 199-210.

Huart Friedlander, Ruth, 1989, "Nouveau regard sur les noms composés" in Gautier, André (ed.), *Explorations en linguistique anglaise*, Berne, Peter Lang, pp. 25-90.

Huart, Ruth, 2002, *Grammaire orale de l'anglais*, Ophrys, Paris, 122 p.

Jensen, J., 1993, *English Phonology, Current Issues in Linguistic Theory*, Vol. 99, Amsterdam/Philadelphia, John Benjamins Publishing Company, 251 p.

Jespersen, O., 1961-65, *A Modern English Grammar on Historical Principles*, vol.VI, London, Allen & Unwin / Copenhagen, Ejnar Munksgaard, 570 p.

Johnson, K., 1997, *Acoustic and Auditory Phonetics*, Oxford, Blackwell, 169 p.

Jones, Daniel, 1963, *An English Pronouncing Dictionary*, 12th edition, London, J. M. Dent & Sons Ltd., 539 p.

Jones, Daniel, 1967, *An English Pronouncing Dictionary*, 13th edition, edited by A. C. Gimson, London, J. M. Dent & Sons Ltd., . 44 p.

Jones, Daniel, 1977, *An English Pronouncing Dictionary*, 14th edition, edited by Susan Ramsaran, London and Melbourne, J. M. Dent & Sons Ltd., 576 p.

Jones, Daniel, 1997, *An English Pronouncing Dictionary*, 15th edition, edited by Peter Roach & James Hartman, Cambridge, Cambridge University Press, 559 p.

Keating, P. A., (eds.), 1994, *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, 462 p.

Kingdon, Roger, 1958a, *The Groundwork of English Intonation*, London, Longman, 272 p.

Kingdon, Roger, 1958b, *The Groundwork of English Stress*, London, Longman, 224 p.

Kuiper, Koenraad, 1999, "Compounding by adjunction and its empirical consequences", *Language Sciences*, 21, pp. 407-422.

Ladd, D. R. Jr, 1984, "English Compound Stress", in Gibbon, D & Richter, R. (ed.), *Intonation, Accent and Rhythm*, Berlin, Walter de Gruyter, pp. 253-266.

Ladd, D. R., 1992, "An Introduction to Intonational Phonology", in Docherty, G. & Ladd, D.R., eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, Cambridge, Cambridge University Press, pp. 321-368.

Ladd, D. R., 1993, "Notes on the Phonology of Prominence", *Working papers 41, Proceedings of the ESCA workshop on prosody*, University of Lund, Sweden, pp. 10-15.

Ladefoged, Peter, 1962, *Elements of Acoustic Phonetics*, Chicago, University of Chicago Press, 120 p.

Ladefoged, Peter, 1971, *Preliminaries to Linguistic Phonetics*, Chicago, University of Chicago Press, 122 p.

Ladefoged, Peter, 1982, *A Course in Phonetics*, Second edition, Fort Worth, Harcourt Brace, 302 p.

Ladefoged, Peter, 1993, *A Course in Phonetics*, Third edition, Fort Worth, Harcourt Brace, 308 p.

Ladefoged, Peter, 2001, *Vowels and Consonants*, Oxford, Blackwell, 191 p.

Liberman & Prince, 1977, "On stress and linguistic rhythm", *Linguistic Enquiry* 8, pp. 249-336.

Lieber, R., "Argument Linking and Compounds in English", *Linguistic Inquiry*, 14, pp. 251-286.

Lieberman, Phillip, 1960, "Some acoustic correlates of word stress in American English". *Journal of the Acoustical Society of America*, 22, pp. 451-454.

Lilly, Richard & Viel, Michel, [1975] 1993, *Initiation raisonnée à la phonétique de l'anglais*, Paris, Hachette Supérieur, 206 p.

Lilly, Richard & Viel, Michel, [1977] 1993, *La prononciation de l'anglais. Règles phonologiques et exercices de transcription*, Paris, Hachette Supérieur, 128 p.

Lutstorf, H., 1960, *The Stressing of Compounds in Modern English: A Study in Experimental Phonetics*, Thèse, Université de Zurich, 159 p.

Lyons, J., 1977, *Semantics*, Vol. 2, Cambridge, CUP, 897 p.

Maniez, F., 2002, "Les mots composés à trait d'union dans la presse américaine : une étude diachronique basée sur l'usage de corpus", *Cahiers du Ciel*, Paris VII, pp. 162-175.

Marchand, Hans, 1969, *The Categories and Types of Present-Day English Word Formation*, Second Edition, Munich: C. H. Beck'sche Verlagsbuchhandlung, 545 p.

Montreuil J.-P., 2001, *La Phonologie de l'anglais*, Rennes, Presses Universitaires de Rennes, 168 p.

Mossé, G., 1945, *Manuel de l'anglais du moyen-âge. I. Vieil anglais*, Paris, Aubier-Montaigne, 383 p.

MPW, Mackintosh Programmer's Workshop, août 1998, ftp://ftp.apple.com/developer/Tool_Chest/Core_Mac_OS_Tools/MPW_etc /

Nesfield, J.-C., 1939, *Manual of English Grammar and Composition*, London, Macmillan & Co. Ltd., 423 p.

Nespoulos, J.-L., ed., 1993, *Tendances actuelles en linguistique générale*, Neuchâtel : Delachaux et Niestlé, 199 p.

Nicaise, A. et Gray, G., 1998, *L'Intonation de l'anglais*, Paris, Nathan, 128 p.

Nolan, F., 1994, "Phonetic correlates of syllable affiliation", in Keating, P. A., (eds.) *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, pp. 160-167.

O'Connor, J. D., 1991 (1973), *Phonetics*, Penguin, 128 p.

Paillard, M., 2000, *Lexicologie contrastive anglais-français*, Paris, Ophrys, 199 p.

Paillard, M., 2002, "Quelques exemples de présentation des faits de langue", *Rapport de l'Agrégation interne*, à paraître.

Paradis, C., 1993, "Phonologie générative multilinéaire", in Nespoulos, J.-L., 1993 ed. *Tendances actuelles en linguistique générale*, Neuchâtel : Delachaux et Niestlé, pp. 11-57.

Picoche, J., 1992, *Précis de lexicologie française*, Poitiers, Nathan université, 191 p.

Poldauf, I., 1984, *English Word Stress*, Oxford, Pergamon, 169 p.

Pountney, M. A., [1972] 1998, *A Practical Approach to English Phonetics*, Département d'anglais, Université de Limoges, 69 p.

Pountney, M. A., 1999, *A Practical Approach to English Phonetics & Phonology*, Limoges, PULIM, 76 p.

PRAAT, http://www.fon.hum.uva.nl/praat ou http://www.praat.org, avril 2000 et septembre 2001

Reporters, émission de BBC World, 21 mars 2001

Reporters, émission de BBC World, 28 mars 2001

Reporters, émission de BBC World, janvier 2001

Roach, Peter, [1983] 1991, *English Phonetics and Phonology*, Cambridge, Cambridge University Press, 262 p.

Roach, Peter, 1992, *Introducing Phonetics*, London, Penguin, 128 p.

Roca, Iggy, 1994, *Generative Phonology*, London, Routledge, 307 p.

Roeper, T. & Siegel, M., 1978, "A Lexical Transformation for Verbal Compounds", *Linguistic Inquiry*, Vol. 9, N° 2, pp. 199-260.

Rowling, J. K., 1997, *Harry Potter and the Philosopher's Stone*, Bloomsbury, London, 223 p.

Rowling, J. K., 1997, *Harry Potter and the Philosopher's Stone*, lu par Stephen Fry, version CDROM, Cover to Cover.

Sampson, Rodney, 1980, "Stress in English N+N phrases: a further complicating factor", *English Studies*, 61, pp. 264-270.

Sawusch, J. & Turk, A., "The processing of duration and intensity clues to prominence", *Journal of the Acoustical Society of America*, Vol. 99, N° 6, pp. 3782-3790.

Selkirk, E., 1982, *The Syntax of Words*, MIT Press, Cambridge, MA, 471 p.

Selkirk, E., 1984, *Phonology and Syntax: The Relation between Sound and Structure*, MIT Press, Cambridge, 476 p.

Selkirk, E, 1995, "Sentence prosody: Intonation, Stress and Phrasing", in Goldsmith, John A., ed., 1995, *The Handbook of Phonological Theory*, Oxford, Blackwell, pp. 562-566.

Shattuck-Hufnagel, S, 1994, "Stress shift as early placement of pitch accents: a comment on Beckman & Edwards", in Keating, P. A., (eds.) *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, pp. 34-42.

Sluijter, A. & van Heuven, V., 1993, "Perceptual clues of linguistic stress: intensity revisited", *Working papers 41, Proceedings of the ESCA workshop on prosody*, University of Lund, Sweden, pp. 246-249.

Taylor, D., 1981, "Non-native speakers and the rhythm of English", *International Review of Applied linguistics in Language Teaching*, Vol. 19, pp. 218-226.

Terken, J., 1991, "Fundamental frequency and perceived prominence of accented syllables", *Journal of the Acoustical Society of America* 894/1: 1768-1776.

Thun, Nils, 1963, *Reduplicative Words in English. A Study of Formations of the Types 'Tick-tick', 'Hurly-burly' and 'Shilly-shally'*, Lund, Carl Bloms Boktryckeri, 347 p.

Tournier, Jean, 1985, *Introduction descriptive à la lexicogénèse de l'anglais contemporain*, Champion-Slatkine, Paris-Genève, 517 p.

Tournier, Jean, 1988, *Précis de lexicologie anglaise*, Nathan, Paris, 208 p.

Trevian, I., 2000, "Typologie et morphologie, des constructions savantes, ou un fonds en constant accroissement du lexique anglais", *Dixième colloque d'avril sur l'anglais oral*, Villetaneuse, à paraître.

Troubetzkoy, N. S., 1949, "Oristique, Etude de la fonction délimitative". in *Principes de phonologie*, trad. J. Cantineau, Paris, Klincksieck, 396 p.

Tsiranana, J. H., 1981, *Recherches sur la formation et accentuation des composés quasi-morphématiques simples et suffixés en anglais contemporain*, Thèse de doctorat de 3ème cycle, Université Paris VII, 420 p.

Turk A., 1994, "Articulatory phonetic clues to syllable affiliation: gestural characteristics of bilabial stops", in Keating, P. A., (eds.) *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press, pp. 107-135.

Viel, Michel, [1981] 1994, *La Phonétique de l'anglais*, Que sais-je?, Paris, Presses Universitaires de France, 5ème édition, 128 p.

Villiard, P. 1993, "Morphologie : tendances actuelles", in Nespoulos, J.-L., 1993 ed. *Tendances actuelles en linguistique générale*, Neuchâtel : Delachaux et Niestlé, pp. 49-82.

Wells, J. C., 1990, *Longman Pronouncing Dictionary*, Harlow, Longman, 802 p.

Wells, John, 1995, SAMPA: "Computer readable phonetic alphabet", http://www.phon.ucl.ac.uk/home/sampa/home.htm.

Wells, John, 1995, "Computer-coding the IPA: a proposed extension of SAMPA", http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm.

Wells, J. C., 2000, "Pronunciation Preferences in British English: A New Survey", *Dixième colloque d'avril sur l'anglais oral*, Villetaneuse, à paraître.

Wells, J. C., 2000, *Longman Pronouncing Dictionary*, Harlow, Longman, 869 p.

Wells, J. C., Barry, W., Grice, M, Fourcin, A., Gibbon, D., 1992, "Standard computer-compatible transcription", Esprit project 2589 (SAM), Doc. no. SAM-UCL-037, London, Phonetics and Linguistics Department, UCL.

Wood, S., 2000, *Praat beginner's manual*, http://www.ling.lu.se/persons/Sidney/

Zumstein, F., thèse en cours, Doctorat nouveau régime, Université de Poitiers, à paraître.

Zumstein, Franck, 1998, "SFI&GI, une interface de recherche sur des bases de données", in *Actes du colloque de décembre 1998 : Phonétique et Phonologie de l'anglais*, Pau, PUP, pp. 137-157.

Zumstein, Franck, 2000, *Les Corpus lexico-phonétiques numérisés et leur contribution à l'étude de la variation accentuelle et phonétique avec une étude particulière des mots en <-CiCe#>*, T.E.R. de D.E.A., Université de Poitiers, 94 p.

INDEX DES NOTIONS

A

ACCENT ADJACENT (*STRESS CLASH*) : 66-7, 76, 79, 81-82, 367-8, 378, 404, 442, 473-4, 525, 597, 598

ACCENT D'INSISTANCE (EMPHASE) : 38, 216, 296, 321, 324, 326-7, 457, 467

ACCENT DE CONTRASTE : 33, 36, 38, 383, 466

ACCENT DE PHRASE : 36-8, 216, 225-6, 228, 281, 373-4, 466-7

ACCENT LEXICAL : 9-10, 17, 28, 38-40, 71, 80, 91, 101, 198, 211, 232, 284, 334, 340, 343-6, 367, 369, 374, 378, 386-8, 392, 425-6, 436, 440, 445, 447, 476, 483-5, 528, 529, 551-2, 569, 589, 595

ACCENT NUCLEAIRE : 35-6, 43, 47, 90, 94, 111, 376, 378, 451, 455-6, 463-464, 476, 480, 524, 527, 572, 596

ACCENT PRIMAIRE : 6, 26, 37, 40-1, 43, 47, 49-53, 56-7, 60-61, 65, 70, 73-4, 76, 78-9, 81-82, 84, 86-90, 92-95, 97, 116-7, 123-4, 133, 160, 177-80, 194, 196, 244, 245, 248, 314, 319, 322-324, 326, 334, 344, 358, 367-8, 378, 442, 473-4, 525, 558, 597-8

ACCENT RYTHMIQUE : 37, 89, 94-95, 124, 195

ACCENT SECONDAIRE : 6, 31, 40, 43, 46, 48-53, 56-7, 70, 74-6, 78, 81-2, 86-90, 93-5, 97, 116-7, 124, 160, 177, 180-1, 194, 245, 248, 319, 344-5, 368-9, 392, 473, 524, 551,596, 598

ACCENT SILENCIEUX (*SILENT STRESS*) : 37, 38

ACCENT TERTIAIRE : 6, 40-1, 43, 46, 54-7, 70, 86-8, 93, 97, 173, 177-9, 183, 194, 244-5, 248, 256-7, 319, 344, 369, 551,597, 598

ACCENT, DEFINITION DE : 9, 17, 21-2, 25, 27-9, 44, 58, 308

ACCENTUATION DOUBLE : 84-85, 87, 89, 91, 94-5, 97, 140-1, 153, 160, 180-1, 195, 232, 245, 252, 254, 283-4, 314-5, 319, 322, 334, 336, 344-5, 367-9, 373-5, 380-6, 390, 392-5, 406, 413, 416-7, 420, 422, 424, 433, 436, 438-9, 445, 454-62, 464, 472-474, 477-8, 480-3, 485-92, 496, 498-501, 503, 505-6, 510, 512-20, 522-527, 543, 547, 549-51, 565-66, 568, 571, 576, 578, 579, 589-91, 596

ACCENTUATION SIMPLE : 78, 84-5, 87, 89, 92, 95-6, 98, 112, 116, 124, 126-7, 140, 153, 170, 173, 177-9, 184, 191, 195, 232, 244, 248, 252-3, 283-4, 306, 314, 319, 322, 334-5, 345, 367-8, 373-5, 378, 380-5, 390, 393-5, 406-9, 415, 420, 422-4, 428, 433, 435, 439, 445, 452, 454, 456-8, 460-2, 464-5, 468, 472, 477, 478-83, 485-92, 494, 496, 498, 500-1, 503-5, 508-12, 515-7, 519, 521, 523, 539, 543, 547, 549-50, 566, 568-70, 573, 578-9, 589-90, 597

ACOUSTIQUE, ANALYSE : 11-12, 16, 30, 38, 48, 83, 210, 218, 226, 227, 228, 231, 279, 293, 301, 327, 374, 380, 386, 396, 423, 439-40, 447, 472, 483-4, 569, 596, 598

AMALGAME : 138-9, 156, 186, 212, 256

AMPLITUDE : 303, 324-5, 448

ARBRE METRIQUE : 66, 68, 312-3, 315

ASSIMILATION : 110

B

BASE DE DONNEES (ACCESS) : 3, 6, 11-12, 165, 197, 231, 238, 283, 293, 331-2, 336-8, 341, 348, 350-1, 354, 367-8, 375, 382, 388-9, 391-3, 399-400, 403, 409, 412, 424, 432, 448, 484-5, 502, 505, 507, 520, 524-5, 528, 529-30, 541, 578, 598

C

CHAMP : 151, 167, 178, 186, 197, 219, 331-2, 336-47, 350, 368, 389-90, 392, 399, 503, 530, 532, 534, 536, 538, 540-2, 595

CHUTE MELODIQUE : 26, 32, 34, 40, 43, 50, 84, 89, 90, 93-5, 97

COMPOSE CO-ORDONNE (*DVANDVA*) : 284, 371, 485, 491, 503-4, 507, 512, 518, 525

COMPOSE ETRANGER : 150, 186, 212, 403

COMPOSE EXOCENTRIQUE (*BAHUVRIHI*) : 284, 371, 485, 507, 513, 516, 521

COMPOSE HISTORIQUE : 72, 78, 130, 143, 145, 149, 152, 167, 186, 212, 256-8, 403, 411, 516

COMPOSE RADICAL (ENDOCENTRIQUE) : 74, 284, 371, 485, 507-8, 510, 515, 569

COMPOSE REDUPLICATIF : 223, 229, 284, 347, 372, 485, 504, 507, 513, 517, 542

COMPOSE VERBAL : 59, 284, 371, 485, 507, 515-6, 519

COMPOSE, DEFINITION DE : 32, 103, 107, 112-14, 131, 185, 195, 391, 412

COMPOUND STRESS RULE : 6, 59-62, 64-65, 111, 374, 378

COMPREHENSION : 8, 15, 17, 30-1, 33-4, 373, 390

CONCORDANCE : 207, 220, 233, 258, 263, 279, 340-1, 453, 460-1

CONFLIT ACCENTUEL : 67, 476

CONSTRUCTION CUMULATIVE : 214, 216, 219-20, 223-4, 227, 230, 232, 284, 441, 446, 477, 485, 489, 507, 515, 523, 524, 574

CONTRAINTE RYTHMIQUE : 15, 51, 73, 81-2, 92, 94, 97, 476, 597

CORRELAT ACOUSTIQUE : 9, 13, 17-8, 20-3, 53-4, 83, 210, 214, 221, 226-9, 210, 214, 221, 226-9, 280, 293-5, 297, 299-300, 302, 304, 307, 316-7, 321, 327, 340, 347, 354, 356-7, 448, 450, 452, 456, 463-4, 468, 470, 481, 484, 524, 529, 557, 595-6, 598

CORRELAT ARTICULATOIRE : 20-22

COUPE SYLLABIQUE : 158, 175, 176, 309-312, 318, 449, 454, 456

CYCLE TRANSFORMATIONNEL : 59, 62, 64

D

DECLINAISON DE F_0 : 297, 299-300, 322, 324, 455-6, 458, 467, 518

DEGRE D'ACCENT : 20, 24, 39, 41, 45-7, 49, 53-5, 62-3, 70, 74, 77, 83, 91, 93, 96, 380, 476, 596

DEPLACEMENT D'ACCENT (*STRESS SHIFT*) : 30, 33, 65-7, 69, 81-3, 90-1, 132, 196, 217, 225-6, 315, 336, 346, 369, 386-7, 417, 421, 423, 425, 434-6, 438, 445-7, 457, 473-5, 478, 480, 488, 511, 517, 524, 534, 550, 566-7, 574-6, 595

DESACCENTUATION : 36, 76, 93, 141, 173, 344, 378, 380, 385-6, 415, 463, 514, 598

DISCOURS CONTINU : 9-10, 15, 17, 48, 51, 79-82, 101, 154, 211, 215, 231, 278, 281, 293, 315, 334, 338, 369, 374, 385-6, 423, 434, 436, 440, 445, 450, 472, 524, 528, 596

DISCOURS, DEBIT : 81, 206, 218, 447, 475, 480, 483, 574

DISCOURS, POSITION SYNTAXIQUE : 450, 453, 459, 472-3, 477, 479, 483, 590, 595

DURÉE : 23, 24, 205, 217-8, 220, 294, 296, 299, 303-5, 315-8, 320-2, 325-7, 357, 448-449, 452, 454, 456-7, 521, 522, 557

E

EXCEL : 252, 253, 319, 331, 340, 347, 399

F

FIGEMENT : 106, 140

FOCUS : 296, 378, 463, 487

FORCE RELATIVE : 21, 27, 28, 29, 30, 92, 303

FORME CONTRACTEE : 133, 155, 167, 186, 188

FREQUENCE FONDAMENTALE : 6, 226, 294-304, 315-8, 320-30, 357, 361-2, 448-9, 452-62, 464-9, 471, 473-8, 480, 482, 497, 509, 511, 518, 522, 524, 526, 557, 559-76, 598

G

GENITIF : 103, 188, 480, 521

GRAPHIE A TRAIT D'UNION : 10, 107, 116, 124-7, 134, 165-6, 171, 180-2, 187, 194, 214, 236-8, 240, 256-7, 311, 333, 335, 337, 342, 353, 396-8, 400, 404, 407, 419-20, 426, 428, 432, 438-9, 535-8, 546, 550

GRAPHIE COMPACTE : 107, 109-10, 166, 168, 396-400, 404-8, 422, 429, 431-2, 435, 437-9, 537-8, 546, 548, 550

GRAPHIE INVERSE : 157, 159, 161-3, 197, 231-2, 335, 342-3, 399, 546, 578

GRAPHIE NON COMPACTE : 86, 108, 142, 153, 166, 168-9, 179, 188, 196, 232, 283, 337-8, 348, 352, 368, 396-7, 406-7, 432, 438-9, 546, 548, 597

GRILLE METRIQUE : 66-7, 75-6, 313-5, 387

H

HAUTEUR MELODIQUE : 19-21, 25, 27, 30, 34, 37, 46-8, 50-1, 54, 82, 216, 294-301, 306-7, 309, 458, 461, 466-7, 476, 483

HIERARCHIE ACCENTUELLE : 9, 17, 27, 38-41, 43-6, 48, 54-6, 58, 60, 66-7, 69-70, 77, 80-1, 83-5, 91-2, 94, 96, 160, 313, 315, 524

HYPERCORRECTION : 411

I

INFORMATION ANCIENNE : 34, 120, 132, 159, 307, 336, 355, 380, 383, 387, 599

INFORMATION NOUVELLE : 34, 36, 159, 336, 380, 383

INTENSITE : 6, 16, 19, 21, 25, 36, 63, 94, 106, 113, 135, 139, 221, 226, 294, 296, 300-4, 306, 308-9, 315-8, 320, 322, 324-30, 357, 363-4, 448-9, 452-61, 465-9, 471-6, 478, 481-2, 509, 511, 518, 522, 526, 539, 557, 559-76, 598

INTONATION : 25-6, 34-5, 37, 46-51, 55, 67, 80, 87, 92, 97, 203, 211, 216, 225-6, 298, 300, 308, 321-2, 324, 326, 328, 331, 373, 378, 386, 452, 454, 458-9, 461-3, 467, 481, 483, 524, 570

ISOCHRONIE : 311

ISOMORPHISME : 56, 73

L

LEXICALISATION : 9-10, 12, 94, 96, 101, 107, 142, 166, 198-201, 204, 210-14, 218, 222-3, 226, 229-31, 233, 280, 284, 346, 355, 383, 401, 403, 408, 418, 420, 423, 425-6, 428, 440-2, 444, 446-7, 452, 457, 467, 475, 480, 481, 485, 487, 506, 508, 509, 524-6, 528, 551, 591, 593, 594

LEXICOGRAPHIE : 113, 598

LEXICOLOGIE : 13, 132, 200, 369, 370

LLIR (LAST LEXICAL ITEM RULE) : 35

LOCUTEUR NATIF : 8, 12, 30, 31, 79, 104, 119, 120, 128, 130, 143-5, 201-3, 205, 207, 228-9, 233, 284, 340, 346, 468, 470, 551 591

LOCUTEUR NON-NATIF : 8, 16, 30, 41, 113, 203, 205, 299, 593

LONGUEUR : 16, 19-21, 23-5, 63, 66, 127, 181, 191, 217, 221, 226, 232, 294, 309, 311, 343, 394, 451, 454, 464, 470, 591

M

MORPHE : 115, 144

MORPHEME LIBRE : 115, 117, 120-1, 123-7, 129, 131, 133-9, 142, 148, 151 253

MORPHEME LIL : 114-5, 117, 120-1, 125, 145

MOYEN ANGLAIS : 145, 147

MPW : 6, 164-6, 184, 235, 245, 248, 253

N

NÉOLOGISME : 11, 101-2, 108, 118, 139, 154, 156, 198, 201, 210, 213, 215, 219-20, 229-30, 333-4, 441, 484, 509, 510, 551

NOMENCLATURE : 17, 19, 26, 34, 39-40, 48, 50, 58, 70, 83, 91, 95-7, 100, 120, 126, 128, 214, 319, 345, 367, 369, 444

NOTATION : 10, 40, 41, 45-6, 50, 54-5, 57-8, 83-9, 91, 95-6, 160, 169, 177-81, 194, 197, 244, 319, 344, 345, 347, 351, 367-9, 360, 388, 449, 564, 595, 596, 597

NUCLEAR STRESS RULE : 6, 59-61, 64, 81, 111, 374, 381

NUMERISATION : 3, 160, 162-3, 177-8, 208-10

O

OPACIFICATION : 130, 133, 136, 145-8, 186, 212, 256, 391, 403, 413

OSCILLOGRAMME : 208, 211, 228, 319, 328-9, 448, 455-6, 472-3, 518, 564-5, 569

P

PÉDAGOGIE : 9, 41, 381, 463, 470

PERCEPTION : 19, 21-3, 27, 29-30, 44, 47, 60, 91, 121, 128, 216, 221-2, 227, 229, 294-6, 298-304, 306, 308, 317, 321, 423, 445, 450 452, 457, 463

PHONOLOGIE METRIQUE : 52-3, 65-6, 81, 311-2, 315, 386-7

PHRASE EDGE PROMINENCE : 82, 387, 475, 477, 480, 524, 566-7, 576

PIED : 29, 53, 68, 307, 311-6, 387, 451, 510, 520

PRAAT : 298, 301-2, 319-20, 322, 328, 331, 340, 347, 356, 358, 365, 558

PRAAT, TEXTGRID : 319, 328, 329, 564

PREFIXE : 32, 56, 73, 75-6, 81, 116-7, 120-8, 133-4, 137, 155, 177, 181-2, 236, 239, 252, 284, 427, 430

PREVISION : 9-10, 12, 198-9, 207, 211, 221, 229, 231, 284, 344-7, 392, 441, 445-7, 484-5, 506-7, 511, 513-4

PRODUCTION : 19, 21-3, 27, 29-31, 44, 47, 60, 128, 295, 563

PRODUCTIVITE : 102, 130, 355, 444

PROEMINENCE : 18-21, 25-7, 29-30, 39, 44, 48-9, 56-7, 67-69, 82, 84, 87, 90, 93-4, 294, 297, 299- 300, 308-309, 374, 377, 379-80, 386, 472, 481, 509 511, 513, 524-5, 566, 569-70, 575-6, 598

Q

QUALITE VOCALIQUE : 21, 24, 41, 46-7, 53-7, 59, 70, 93 171, 190, 192, 294, 305, 450

QUASI-LEXEME : 78, 118, 120-1, 127-9, 132, 138, 151, 153, 156, 182, 194-5, 212, 252, 284, 430

QUASI-MORPHEME : 79, 118

R

REDUCTION VOCALIQUE : 15-7, 24, 43, 53, 55, 62, 64, 70-3, 76-7, 92-3, 98, 109-10, 122, 128, 133, 150, 155, 170, 173-6, 184-5, 189, 191-3, 203, 239, 255-6, 305-6, 347,

379-80, 391, 403, 409-15, 454, 470-2, 539, 597

REGLE D'ACCENTUATION : 71, 372, 374, 381, 396, 483-4, 486, 506, 512, 523, 589, 599, 569

RELATION PRECONSTRUITE : 141, 384, 487, 491, 494

RELATION PRIMITIVE : 105, 384-5

REQUETE : 164-6, 168, 171-2, 176, 178-80, 183, 185, 195, 333, 336, 392-3, 397, 399, 415, 423, 432, 484, 495, 503, 507, 516-7, 530-42, 579, 581

[illegible]TION GERMANIQUE : 72-3, 196, 433, 435, 436

RYTHME : 15-7, 23-4, 29-31, 38-40, 47, 51-7, 67, 79, 81, 83, 89, 94-5, 117, 124, 181, 305, 307, 311-2, 314, 319, 368-9, 387, 450, 451, 473-5, 505, 511, 525, 596

S

SAMPA : 6, 160-2, 172, 347, 539

SEMANTIQUE : 31-2, 34-5, 90, 103-6, 113, 115-8, 122-6, 130, 138, 140, 147-8, 154-5, 175, 369-71, 373, 375-80, 382-3, 390-2, 404, 414-5, 428, 451, 481, 483-84, 486, 508, 514, 591, 598

SONORITE, THEORIE DE LA : 302, 309

SPECTROGRAMME : 298

STRESS-TIMED : 15-17, 23, 29, 75

SUFFIXE : 76, 116, 118, 121, 130, 155, 157, 178, 181-3, 194, 236-7, 379, 410, 415, 516

SURCOMPOSE : 61-62, 213, 397, 441

SYLLABE FAIBLE : 16, 43, 45, 47, 51, 75-6, 85, 87, 180, 189, 193, 304-6, 321, 411, 454, 459, 462, 471, 551, 598

SYLLABE FORTE : 23-5, 28, 30-1, 34, 37, 39, 43, 47, 50, 76, 85, 87, 171, 299, 301, 303-4, 310-1, 319, 321, 324, 334, 368-9, 452, 454, 457, 460, 482, 596

SYLLABLE-TIMED : 15

SYNTAGME : 33, 37, 59-60, 64-5, 82, 111-3, 140, 148, 196, 214, 224-6, 284, 314, 320, 325-6, 328, 371, 413, 477, 487, 493, 521, 525, 597

SYNTAXE : 12, 59-60, 69, 106, 141, 162-5 170, 233, 284, 335, 348, 370, 374, 376, 391-2, 403, 417-8, 420-1, 424, 429, 485, 488, 512, 549, 578, 590-1

T

TABLE : 222, 283, 331, 336-42, 344, 368, 388-9, 393-5, 398, 400, 403-4, 406, 408-10, 417-22, 425-28, 484, 495, 502-3, 516, 532-5, 539, 541, 543, 545-6, 549-50, 578, 581, 586, 595

TAUX D'ECHANTILLONNAGE : 208-10, 302

THEORIE GENERATIVE : 41, 58, 65-66, 76, 80, 93, 111, 140 190, 192, 370, 374-6, 378, 381

TMPP (THEORIE MOTRICE DE LA PERCEPTION DE LA PAROLE) : 21

TON CIRCONFLEXE : 322, 326, 462, 467, 577

TON CREUSE : 322, 326, 459

TON DESCENDANT : 26, 32, 34, 36, 40, 43, 50-1, 84, 89-90, 93-5, 97, 113, 225, 319, 321-3, 326, 368, 385, 424, 457-9, 462, 464-5, 468, 474, 566, 568, 570-1, 573, 576

TON MONTANT : 35, 225, 299, 322, 326, 459-62, 467, 483, 577

TRONCATION : 116, 138-9, 151, 154, 156, 186, 188, 212, 256, 257

V

VARIATION : 12, 45, 62, 73, 80, 82, 104, 123, 144, 151, 158, 162-3, 165, 188, 197-8, 206, 212, 221, 227-8, 231-2, 236, 244, 300, 307, 315, 333-4, 340, 342, 344, 346, 352, 387, 390-2, 397-8, 401-2, 404, 406, 408, 410, 414, 423, 434-6, 445-7, 468, 477, 479-81, 483, 497, 501, 510, 513, 517, 519, 536, 540, 546, 550, 568, 578-9, 590, 594, 595

VIEIL ANGLAIS : 145-147

INDEX DES AUTEURS

A

ABERCROMBIE, D. : 311, 312, 315

ADAMS, V : 106, 107, 109, 111, 115, 116, 122, 130, 132, 135, 136, 138, 139, 371, 372

ALLEN, G. : 307, 308, 450

B

BAUER, L. : 108, 109, 112, 115, 139, 144, 200, 214, 370, 372, 376, 387

BECKMAN, M. : 295, 297, 298, 299, 301, 302, 303, 304, 306

BENVENISTE, E. : 146

VAN DEN BERG, R. : 458

BLEVINS, J. : 302

BOISSON, C. : 8, 9, 16, 18, 23, 26, 34, 36, 41, 43, 46, 47, 49, 50, 51, 54, 63, 79, 84, 145, 146, 345, 373, 382, 383, 384, 386, 388, 464, 476, 486, 495, 594

BOLINGER, D. : 216, 378, 463

BOUCHER, P. : 32

BOUSCAREN, C. : 75

BROWN, W. & MCGLONE, R. : 296

BUCELLATO, P. : 306

C

CHOMSKY, N. : 6, 43, 51, 54, 58, 59, 60, 62, 63, 65, 141, 374, 375, 376

COLLIER, R. : 463

CORBIN, D. : 115

CREPIN, A. : 309

D

DESCHAMPS, A. . 35, 43, 51, 104, 107, 146, 381

DIANA, A. : 295, 299, 306

DOCHERTY, G. : 458

DOWNING, P. : 370

DUCHET, J.-L. : 8, 15, 35, 43, 45, 51, 73, 74, 75, 78, 104, 107, 146, 235, 380, 381, 386

DUCHIRON, E. : 177

DUGARDIN, S. : 295, 296, 297, 298, 317, 318, 322, 324, 326, 457

DURAND, J. : 18, 21, 22

E

EDWARDS, J. : 295, 304

F

FOURNIER, J.-M. : 31, 34, 35 43, 51, 55, 76, 78, 79,104, 107, 117, 121, 122, 146, 160, 161, 381

FOWLER, H.W. : 108, 132, 408

FRY, D. B : 294, 295, 301, 304, 306, 307, 308

FUDGE, E. : 31, 43, 45, 51, 52, 104, 321, 376, 379

G

GIEGERICH, H. J. : 18 23, 27, 28, 29, 37, 43, 51, 53, 55, 65, 309, 376, 386, 387

GIMSON, A. C. : 15, 16, 18, 23, 27, 34, 43, 45, 47, 48, 51, 381, 386

GINÉSY, M. : 43, 51, 381, 384, 385, 416, 444, 486, 495, 508, 516

GOLDSMITH, J. A. : 18, 24, 25, 47, 52, 66, 67, 68, 69, 82, 302, 309 311, 387, 475

GOYVAERTS, D. L. & PULLUM, G. K. : 51, 59, 63

GRAY, M. : 297, 298, 378, 385, 458, 459, 460, 462, 463

GROSS, G. : 106, 140

GUIERRE, L. : 6, 10, 15, 32, 43, 51, 52, 55, 69, 70, 71, 72, 73, 75, 76, 77, 90, 109, 117, 118, 119, 157, 160, 161, 162, 165, 169, 173, 174, 192, 193, 235, 343, 388, 396, 403, 404

GUSSENHOVEN, C. : 295, 307, 311, 458

H

HALLE, M. : 6, 43, 51, 54, 58, 59, 60, 62, 63, 64, 65, 141, 374, 375, 376

HAMMOND, M. : 18, 19, 23, 308

HUART, R. : 8, 18, 23, 25, 26, 27, 33, 41, 43, 49, 50, 54, 95,103, 104, 141, 204, 378, 379, 384, 385, 416, 463, 487, 488, 491, 494, 495, 498, 507, 509, 514, 516, 521

J

JACOBS, H. : 307, 311

JENSEN, J. : 66, 308, 355

JESPERSEN, O. 104, 113, 141, 372

JOHNSON, K. : 210, 302, 324

JONES, D. : 3, 6, 47, 78, 8[illegible] 88, 89 111, 345 :

K

KEYSER, S. J. : 60, 64

KINGDON, R. : 18, 23, 26, 37, 43, 49, 72, 74, 75, 81, 95, 103, 104, 107, 378, 382, 387, 595

KUIPER, K. : 371

L

LADD, D. R. : 140, 301, 385, 458

LADEFOGED, P. : 16, 18, 21, 23, 27, 43, 46, 47, 49, 51, 57, 96, 294, 304, 305, 306, 307, 308, 309, 310

LIBERMAN, M. & PRINCE A. : 307, 311, 370, 377, 387

LIEBERMAN, P. : 295, 304, 306

LILLY, R. : 43, 53, 54, 55, 56, 57, 77, 98, 597

LUTSTORF, H. : 140, 379, 387, 594, 598

LYONS, J. : 109, 140, 142, 378

M

MANIEZ, F. : 444, 593

MARCHAND, H. : 106, 112, 132, 144, 145, 146, 155, 167, 371, 372, 379, 414, 521

MONTREUIL J.-P. : 386

MOSSÉ, G. : 103, 104

N

NESFIELD, J.-C. : 145

NICAISE, A. : 378, 385, 458, 459, 460, 462, 463

NOLAN, F. : 309

O

O'NEIL, M. : 35, 43, 51, 104, 107, 146, 381

P

PAILLARD, M. : 109, 114, 118, 138, 213

PICOCHE, J. : 102, 104, 142, 214

PIERREHUMBERT J. : 299

POLDAUF, I. : 18, 23, 27, 28, 74

POUNTNEY, M. A. : 82, 98

R

RENOUF A. : 109, 200, 214

RIETVELD, A. : 295, 458

ROACH, P. : 15, 18, 19, 20, 23, 24, 29, 43, 45, 51, 78, 80, 88, 89,192, 367

ROEPER, T. & SIEGEL, M. : 370, 371, 387, 515, 593

S

SAMPSON, R. : 387

SAWUSCH, J. : 303

SELKIRK, E. : 67, 82, 378, 387, 463, 475, 476

SHATTUCK-HUFNAGEL, S. : 387

SLUIJTER, A. & VAN HEUVEN, V. : 221, 296, 301, 303, 306

T

TAYLOR, D. : 31

TERKEN, J. : 295, 296, 299, 300, 318

THUN, N. : 146, 517

TOURNIER, J. : 102, 107, 108, 111, 112, 113, 129, 132, 137, 138, 350, 372, 592

TREVIAN, I. : 118

TROUBETZKOY, N. S. : 37, 69, 110

TURK A. : 303, 309

V

VIEL, M. : 43, 53, 55, 56, 57, 76, 77, 98, 597

W

WELLS, J. C. : 6, 46, 55, 56, 57, 84, 88, 89, 93, 118, 122, 125, 128, 132, 160, 175, 178, 196, 310, 372, 381, 414, 416, 473, 497, 597

WOOD, S. : 153

Z

ZUMSTEIN, F. : 159, 228

TABLE DES MATIÈRES

TABLE DES MATIERES

REMERCIEMENTS 3

SOMMAIRE 5

ABREVIATIONS & SYMBOLES 6

INTRODUCTION 7

CHAPITRE 1 14

1 L'ACCENTUATION ET LES COMPOSES DANS LES TRAVAUX PHONOLOGIQUES RELATIFS A L'ANGLAIS 15

1.1 Définition de l'accent 17
1.1.1 L'accent et la proéminence : un ensemble de facteurs 18
1.1.2 Les facteurs à prendre en compte pour une définition de l'accent 21
1.1.2.1 Intensité et force (perception et production) 21
1.1.2.2 Longueur 23
1.1.2.3 Qualité vocalique 24
1.1.2.4 Hauteur mélodique 25
1.1.2.5 Force relative des syllabes 27
1.1.3 Une définition de l'accent 29

1.2 Accent et compréhension 30

1.3 Les différents types d'accent 35

1.4 La hiérarchie accentuelle 39
1.4.1 Critères de détermination de la hiérarchie accentuelle 44
1.4.1.1 Perception et production 44
1.4.1.2 Mélodie 47
1.4.1.3 Rythme 50
1.4.1.4 Qualité vocalique 53
1.4.1.5 Rythme et qualité vocalique : l'accent 3 en question 55

1.5 Statut de l'accentuation dans la théorie phonologique 58
1.5.1 L'approche de *The Sound Pattern of English* (SPE) (1968) 58
1.5.2 La phonologie métrique 65
1.5.3 L. Guierre et ses successeurs 69
1.5.3.1 Les règles générales de l'accentuation de L. Guierre 72
1.5.3.1.1 Règle 1 74
1.5.3.1.2 Règle 2 74
1.5.3.1.3 Règle 3 75
1.5.3.1.4 Règle 4 76
1.5.3.1.5 La règle de Fournier 78
1.5.3.2 Les règles qui concernent les composés 79

1.6 L'accent dans le discours continu 80

1.7 Nomenclature et notation adoptée 83
1.7.1 Nomenclature 83
1.7.2 Notation : la hiérarchie accentuelle dans les composés 84

1.7.2.1 Notation de la hiérarchie accentuelle dans les dictionnaires de prononciation 85
1.7.2.1.1 Tous les composés 85
1.7.2.1.2 Composés à accentuation simple 87
1.7.2.1.3 Composés à accentuation double 89
1.7.2.1.4 Notation du déplacement de l'accent sur les composés 90
1.7.2.2 Notation de la hiérarchie accentuelle retenue 91
1.7.2.2.1 Tous les composés 91
1.7.2.2.2 Composés à accentuation simple 92
1.7.2.2.2.1 Deux éléments monosyllabiques 92
1.7.2.2.2.2 Un élément plurisyllabique 93
1.7.2.2.3 Composés à accentuation double 94
1.7.2.2.3.1 Deux éléments monosyllabiques 94
1.7.2.2.3.2 Un élément plurisyllabique 95
1.7.2.2.4 Récapitulatif de la notation 95

CHAPITRE 2 99

2 LA CONSTITUTION DU CORPUS 100

2.1 Le contenu du corpus 102
2.1.1 Qu'est-ce qu'un composé ? 102
2.1.1.1 Les critères de définition d'un composé 103
2.1.1.1.1 Le critère sémantique 103
2.1.1.1.2 Le critère syntaxique 105
2.1.1.1.3 Le critère graphique 107
2.1.1.1.4 Le critère phonétique et phonologique 109
2.1.1.1.5 Le critère accentuel 111
2.1.1.2 Les critères de définition d'un composé : Quel critère appliquer ? 112
2.1.2 Qu'est-ce qu'un composé ? Les cas limites 114
2.1.2.1 La dérivation et la composition 114
2.1.2.1.1 Eléments à statut simple de morphème lié 121
2.1.2.1.1.1 Les préfixes 121
2.1.2.1.1.1.1 Les cas de *re-* et *de-* 122
2.1.2.1.1.1.2 Le cas de *inter-* 124
2.1.2.1.1.1.3 Le cas de *co-* 125
2.1.2.1.1.2 Les critères d'identification des préfixes 126
2.1.2.1.1.3 Autres éléments en début de mot : les quasi-lexèmes 127
2.1.2.1.2 Eléments à statut double : élément en construction et morpheme libre 129
2.1.2.1.2.1 Les éléments en construction placés en seconde position 129
2.1.2.1.2.1.1 Les cas de *-like*, *-most*, *-ship* et *-scape* 129
2.1.2.1.2.1.2 Les cas de *-way*, *-ways* et *-wise* 131
2.1.2.1.2.1.3 Le cas de *-meter* 132
2.1.2.1.2.1.4 Les cas de *-by* et *-out* 133
2.1.2.1.2.2 Eléments en construction en première position 133
2.1.2.1.2.2.1 Le cas de *in-* 133
2.1.2.1.2.2.2 Les cas de *pro-* et *anti-* 134
2.1.2.1.2.2.3 Le cas de *fore-* 135
2.1.2.1.2.2.4 Les cas de de *out-* et *under-* 135
2.1.2.1.2.2.5 Les cas de de *counter-* et *by-* 136
2.1.2.1.2.2.6 Les cas de *semi-* et *sub-* 137
2.1.2.1.2.3 Eléments de construction en première ou seconde position 138
2.1.2.1.2.3.1 Les quasi-lexèmes 138
2.1.2.1.2.3.2 Eléments qui se ramènent à des éléments tronqués 138
2.1.2.2 Les syntagmes et les composés 139
2.1.2.3 Les éléments non reconnus 142

2.1.2.3.1 Entrées inconnues 143
2.1.2.3.2 Entrées connues 144
2.1.2.3.3 Les composés historiques 145
2.1.3 Les critères de sélection des composés pour le corpus d'analyse 148
2.1.4 Quelle sera l'étendue du corpus d'analyse ? 149
2.1.4.1 Composés étrangers 149
2.1.4.2 Quasi-lexèmes 151
2.1.4.3 Noms propres 151
2.1.4.3.1 Noms propres étrangers 152
2.1.4.3.2 Noms propres indigènes 152
2.1.4.3.3 Noms propres tdmk 154
2.1.4.3.4 Acronymes 154
2.1.4.4 Formes contractées 155
2.1.4.5 Amalgames 156

2.2 Sélection des données : sources écrites 156
2.2.1 Choix des sources 157
2.2.1.1 Les informations 157
2.2.1.2 Les sources non retenues 159
2.2.1.3 Les sources retenues 160
2.2.1.3.1 *English Pronouncing Dictionary*, 12ème édition (EPD12) 160
2.2.1.3.2 *New Pronouncing Dictionary* de Guierre (NPD) 162
2.2.1.3.3 *Longman Pronouncing Dictionary*, édition 1990 (LPD90) 162
2.2.2 Comment extraire les données ? 163
2.2.2.1 MPW et l'extraction des données 164
2.2.2.2 Extraction des composés selon le critère syntaxique 164
2.2.2.3 Extraction des composés selon des critères graphiques 166
2.2.2.4 Extraction des composés selon les règles phonologiques et phonétiques 170
2.2.2.4.1 La non-réduction de la voyelle du deuxième élément 170
2.2.2.4.1.1 Méthode d'extraction 171
2.2.2.4.1.2 Remarques 172
2.2.2.4.2 Les séquences phonétiques indiquant une frontière morphologique 174
2.2.2.5 Extraction des composés selon un critère accentuel 177
2.2.2.5.1 Les composés à accentuation simple 177
2.2.2.5.1.1 La notation de l'accent /3 177
2.2.2.5.1.2 Autres requêtes 179
2.2.2.5.2 Les composés à accentuation double 180
2.2.2.6 Extraction des composés selon une recherche sur les éléments 181
2.2.2.7 Les exceptions 183
2.2.3 Résultats des extractions par traitement automatique 184
2.2.3.1 Conclusions 187
2.2.3.1.1 Le critère graphique 187
2.2.3.1.2 Le critère phonologique / phonétique 189
2.2.3.1.3 Le critère accentuel 194
2.2.3.1.3.1 L'accent /3/ 194
2.2.3.1.3.2 Les autres sous-fichiers 195
2.2.4 Saisie des données complémentaires : LPD version papier 196

2.3 Sélection des données : sources orales 198
2.3.1 Objectifs 198
2.3.2 Choix des sources 199
2.3.2.1 Source documentaire 202
2.3.2.1.1 Reporters 202
2.3.2.1.2 *Documentaires divers* 204
2.3.2.2 Source littéraire 205
2.3.3 La sélection des unités 207
2.3.3.1 Informations techniques 207

2.3.3.2 Méthode de sélection 210
2.3.3.2.1 Présélection des extraits sonores 210
2.3.3.2.1.1 Composés lexicalisés 210
2.3.3.2.1.1.1 Critères de sélection : sources documentaires 210
2.3.3.2.1.1.2 Critères de sélection : source littéraire 211
2.3.3.2.1.2 Composés non lexicalisés 212
2.3.3.2.1.2.1 Critères de sélection : les unités 213
2.3.3.2.1.2.2 Critères de sélection : le contexte du discours 215
2.3.3.2.1.2.2.1 L'accent de phrase 216
2.3.3.2.1.2.2.2 L'intonation 216
2.3.3.2.1.2.2.3 Le déplacement de l'accent 217
2.3.3.2.2 Présélection des unités: résultats 217
2.3.3.2.3 Critères de sélection pour le corpus oral d'analyse 219
2.3.3.2.3.1 Sélection des extraits où il y a un élément en commun 221
2.3.3.2.3.2 Sélection des extraits selon leur formation 222
2.3.3.2.3.2.1 's ou pas 223
2.3.3.2.3.2.2 Composés réduplicatifs 223
2.3.3.2.3.2.3 Constructions cumulatives 223
2.3.3.2.3.2.4 Composés et syntagmes 224
2.3.3.2.3.3 Sélection des extraits afin d'étudier l'influence du contexte de discours 225
2.3.3.2.3.3.1 L'intonation 225
2.3.3.2.3.3.2 L'accent de phrase 226
2.3.3.2.3.3.3 Le déplacement de l'accent 226
2.3.3.2.3.4 Sélection des extraits : la variation et la perception de l'accent 227
2.3.3.2.3.4.1 La variation 227
2.3.3.2.3.4.2 Perception et facteurs acoustiques 228
2.3.3.2.3.5 Sélection des extraits en vue de la prévision de l'accent 229
2.3.3.2.4 Sélection pour le corpus oral d'analyse : Résultats 230
2.3.4 Choix des données saisies 231

ANNEXES DU CHAPITRE 2

Annexe A: Extraction des sous-fichiers : programmation 236

Annexe B : Récapitulatif détaillé des fichiers LPD90 : vérification du statut de composé 257

Annexe C : Les reportages de la source documentaire 259

Annexe D : La source littéraire 278

Annexe E : Le corpus oral d'analyse 283

CHAPITRE 3 292

3 METHODES D'ANALYSE DU CORPUS 293

3.1 Analyse acoustique 293
3.1.1 Les corrélats acoustiques de l'accent 294
3.1.1.1 La fréquence fondamentale 295
3.1.1.1.1 La relation entre la fréquence fondamentale (F_0) et l'accent 295
3.1.1.1.2 Comment interpréter la frequence fondamentale 297
3.1.1.1.2.1 Paramètres physiologiques 297
3.1.1.1.2.2 Paramètres contextuels 298
3.1.1.1.2.3 Déclinaison de F_0 299
3.1.1.2 L'intensité 300
3.1.1.2.1 La relation entre l'intensité et l'accent 300

3.1.1.2.2 Comment interpréter l'intensité 301
3.1.1.3 La durée 304
3.1.1.3.1 La durée et l'accent 304
3.1.1.3.1 Comment interpréter la durée 304
3.1.1.4 La qualité vocalique 305
3.1.1.4.1 La qualité vocalique et l'accent 305
3.1.1.4.2 Comment interpréter la qualité vocalique 305
3.1.1.5 Un ensemble de paramètres 306
3.1.2 Les unités de comparaison 307
3.1.2.1 La syllabe 307
3.1.2.1.1 Application à l'analyse des composés 309
3.1.2.2 Le pied 311
3.1.2.2.1 Application à l'analyse des composés 312
3.1.2.3 La syllabe ou le pied : quelle unité pour l'analyse des composés ? 314
3.1.3 Echantillonnage des extraits ayant l'élément «dark» 316
3.1.3.1 La méthode 316
3.1.3.1.1 Le choix de l'échantillonnage 316
3.1.3.1.2 Le choix des mesures et des unités de référence 317
3.1.3.2 L'analyse 320
3.1.3.2.1 Les unités de référence 320
3.1.3.2.2 Les mesures 321
3.1.3.2.2.1 Le F_0 322
3.1.3.2.2.2 L'intensité 324
3.1.3.2.2.3 La durée 325
3.1.3.2.2.4 Le temps du F_0 et de l'intensité 326
3.1.3.3 Conclusions 327
3.1.4 Les mesures et les unités de référence retenues pour le corpus oral d'analyse 327

3.2 La base de données sous Access 331
3.2.1 Structure de la base de données 331
3.2.1.1 Les requêtes 332
3.2.1.1.1 Comparaison des données de sources différentes 332
3.2.1.1.2 Comparaison des données d'une même source 334
3.2.1.2 Les tables et les champs 336
3.2.1.2.1 Les tables de la base de données 336
3.2.1.2.2 Les champs de la base de données 337
3.2.1.2.2.1 L'identification de la source 341
3.2.1.2.2.2 La graphie 342
3.2.1.2.2.3 Le schéma accentuel 343
3.2.1.2.2.4 La prononciation des segments 346
3.2.1.2.2.5 Les informations complémentaires 347
3.2.2 L'indexation et la présentation des données 348
3.2.2.1 Accès et présentation des données dans LPD 348
3.2.2.2 Accès aux données de la base 350
3.2.2.3 Présentation des données de la base 351

3.3 Conclusion 354

ANNEXES DU CHAPITRE 3

Annexe A : Analyse acoustique : tableaux récapitulatifs des mesures pour l'échantillonnage des extraits sonores en *dark* 357

Annexe B : Codes utilisés par PRAAT pour la phonie 365

CHAPITRE 4 366

4 REGLES D'ACCENTUATION DES COMPOSES : ANALYSES ET RESULTATS 367

4.1 Rappel de la nomenclature 367

4.2 Règles pour l'accentuation des composés : principes généraux 369
4.2.1 La formation des composés 369
4.2.1.1 L'approche syntaxique 370
4.2.1.2 L'approche sémantique 370
4.2.2 L'accentuation des composés 372
4.2.2.1 L'approche syntaxique : le mécanisme de l'accent 374
4.2.2.2 L'approche sémantique : le rôle de l'accent 377
4.2.2.2.1 Désaccentuation d'un élément 378
4.2.2.2.1.1 Considérations accentuelles 378
4.2.2.2.1.2 Considérations phonétiques et phonologiques 379
4.2.2.2.2 Accentuation simple ou accentuation double 380
4.2.2.2.2.1 L'accentuation « normale » 381
4.2.2.2.2.2 L'accentuation « exceptionnelle » 381
4.2.2.2.2.3 L'accentuation : reflet d'une relation syntaxique entre les éléments 382
4.2.2.2.2.4 L'accentuation : reflet d'une relation sémantique entre les éléments 383
4.2.2.2.2.5 L'accentuation : reflet d'une relation de fonction / nature 384
4.2.2.2.2.6 L'accentuation : reflet d'une relation préconstruite (ou non) entre les éléments 384
4.2.3 Le déplacement de l'accent 386

4.3 Analyse de l'accent lexical dans la base de données 388
4.3.1 Vérification des différentes tables de la base de données : les doublons 389
4.3.2 Etude de la corrélation entre le schéma accentuel et les autres critères de définition d'un composé 391
4.3.2.1 Etude du schéma accentuel 392
4.3.2.1.1 La structure des tables 392
4.3.2.1.2 Conclusions 395
4.3.2.2 Etude de la graphie 396
4.3.2.2.1 Caractéristiques des tables 396
4.3.2.2.2 La structure des tables 397
4.3.2.2.3 Conclusions 402
4.3.2.2.3.1 Le contenu des tables : critères de sélection des composés 402
4.3.2.2.3.1.1 Entrées manquantes de LPD-Cp/N-COMPACT par rapport à NPD-Cp 403
4.3.2.2.3.1.2 Entrées manquantes de LPD-Cp par rapport à NPD-Cp 404
4.3.2.2.3.2 La graphie des tables : la variation 404
4.3.2.2.4 La corrélation entre le schéma accentuel et la graphie 405
4.3.2.3 Etude de la phonie 409
4.3.2.3.1 Conclusions 415
4.3.2.3.2 La corrélation entre le schéma accentuel et la phonie 415
4.3.2.4 Etude de la catégorie syntaxique 417
4.3.2.4.1 Structure des tables 417
4.3.2.4.2 Conclusion 420
4.3.2.4.3 Corrélation entre le schéma accentuel et la catégorie syntaxique 420
4.3.2.4.3.1 Conclusion 424
4.3.3 Etude de l'évolution du lexique 425
4.3.3.1 Le contenu 426
4.3.3.1.1 Les entrées manquantes dans LPD 426
4.3.3.1.2 Les entrées manquantes dans LPD12 428

4.3.3.2 La graphie et le schém : accentuel 431
4.3.3.2.1 La graphie 431
4.3.3.2.2 Le schéma accentuel 432
4.3.3.2.2.1 Accentuation simple vers accentuation double 433
4.3.3.2.2.1.1 La variation 434
4.3.3.2.2.1.2 Le déplacement de l'accent 434
4.3.3.2.2.2 Accentuation double vers accentuation simple 435
4.3.3.2.2.2.1 La variation 435
4.3.3.2.2.2.2 Le déplacement de l'accent 436
4.3.3.2.2.3 Conclusion 436
4.3.3.2.3 La corrélation entre le changement de schéma et la graphie 437
4.3.3.2.3.1 Accentuation simple vers accentuation double 438
4.3.3.2.3.2 Accentuation double vers accentuation simple 439
4.3.3.2.4 Conclusions 439

4.4 L'accent lexica. et le corpus oral 440
4.4.1 Le taux de lexicalisation des extraits sonores 440
4.4.2 Les prévisions de l'accent et le comportement des unités en discours 445

4.5 Analyse acoustique des extraits sonores de la base de données 448
4.5.1 Le schéma accentuel et les corrélats acoustiques 450
4.5.1.1 Le rythme 450
4.5.1.2 Les corrélats acoustiques de la perception de l'accent 452
4.5.1.2.1 L'accentuation simple 452
4.5.1.2.2 L'accentuation double 455
4.5.1.2.3 L'interprétation du contour intonatif 458
4.5.1.2.3.1 Les différents contours intonatifs 458
4.5.1.2.3.1.1 Intonation descendante : cas particuliers 458
4.5.1.2.3.1.2 Intonation montante 460
4.5.1.2.3.1.3 Intonation montante-descendante 462
4.5.1.2.3.2 Résumé des caractéristiques de l'accentuation simple et de l'accentuation double en fonction du contour intonatif 462
4.5.1.2.4 Les contextes non neutres 463
4.5.1.2.4.1 Focus 463
4.5.1.2.4.2 Le contraste 466
4.5.1.2.4.3 L'insistance 467
4.5.1.2.4.4 Les facteurs psychologiques 468
4.5.2 Le schéma accentuel et la phonie 470
4.5.3 Le schéma accentuel et la position syntaxique des composés dans le discours 473
4.5.3.1 Le déplacement de l'accent 473
4.5.3.2 *Phrase Edge Prominence* et le déplacement de l'accent 475
4.5.3.3 Les composés isolés et les composés dans une construction cumulative 477
4.5.4 Le schéma accentuel et la variation 479

4.6 L'analyse acoustique, l'analyse de l'accent lexical et les règles d'accentuation des composés 484
4.6.1 Etude des entrées ayant deux types d'accentuation 486
4.6.2 Etude des entrées ayant une majuscule 495
4.6.2.1 Vérification des statistiques de N-COMPACT 496
4.6.2.2 Etude des entrées 496
4.6.2.3 Etude des entrées noms propres de la base de données 502
4.6.3 Etude des entrées comportant un des éléments des extraits du corpus oral 506
4.6.3.1 Les différents types de composés 508
4.6.3.1.1 Composés endocentriques 508
4.6.3.1.2 Composés co-ordonnées 512
4.6.3.1.3 Composés exocentriques 513
4.6.3.1.4 Composés synthétiques 515

4.6.3.2 Les composés réduplicatifs 517
4.6.3.3 Les composés en 's 520
4.6.3.4 Les constructions cumulatives et les composés à plusieurs éléments 524

ANNEXES DU CHAPITRE 4

Annexe A : Requêtes **530**

Annexe B : La structure des tables : statistiques à partir des requêtes **543**

Annexe C : Analyse de l'accent lexical et le corpus oral **551**

Annexe D : Analyse acoustique : tableaux récapitulatifs des mesures **557**

Annexe E : Graphiques correspondant aux extraits sonores **564**

Annexe F : Listes d'entrées extraites de la base de données **578**

CONCLUSION **588**

GLOSSAIRE **600**

BIBLIOGRAPHIE THEMATIQUE **602**

REFERENCES BIBLIOGRAPHIQUES **616**

INDEX DES NOTIONS **628**

INDEX DES AUTEURS **633**

TABLE DES MATIERES **636**

TABLE DES FIGURES **645**

CONTENU DU CDROM **648**

TABLE DES FIGURES

TABLE DES FIGURES

Figure 1.1 : Récapitulatif des paramètres de l'accent 23
Figure 1.2 : Conventions de notation des dictionnaires de prononciation 40
Figure 1.3 : Synthèse des différentes hiérarchies accentuelles 43
Figure 1.4 : Récapitulatif des accents, d'après Gimson 48
Figure 1.5 : Grille métrique : *black board eraser* 67
Figure 1.6 : Arbre métrique *black board eraser* 68
Figure 1.7 : Grille métrique *fruit market* 76
Figure 1.8 : Dictionnaires de prononciation, notation des schémas accentuels dans les composés 86
Figure 1.9 : Récapitulatif de la nomenclature proposée 96
Figure 2.1 : Récapitulatif des résultats par critère d'extraction 185
Figure 2.2 : Récapitulatif des résultats : sous-fichiers des dissyllabes 190
Figure 2.3 : Nombre d'extraits sonores présélectionnés 218
Figure 2.4 : Nombre d'extraits sélectionnés pour le corpus oral d'analyse 231
Figure II.1 : Système de correspondance entre les voyelles de l'anglais et N° voyelle 238
Figure II.2 : Tableau récapitulatif : nombre d'entrées incluses / exclues du corpus d'analyse, par catégorie 257
Figure II.3 : « *Documentaires divers* » : Reportages retenus pour la présélection des extraits sonores 259
Figure II.4 : *Reporters* : reportages retenus pour la présélection des extraits sonores 261
Figure II.5 : Les répertoires des extraits de la source littéraire Harry Potter 278
Figure II.6 : Informations correspondant au tableau récapitulatif (Figure II.7) 283
Figure II.7 : Tableau récapitulatif des extraits sonores du corpus oral d'analyse, classés selon le répertoire et le fichier son 285
Figure 3.1 : Représentation de *flying-key room* : arbre métrique 312
Figure 3.2 : Représentation de *flying-key room* : grille métrique 313
Figure 3.3 : Representation de *trophy room*: arbre métrique 313
Figure 3.4 : Representation de *trophy room*: grille métrique 313
Figure 3.5 : dark\defdark2 : Oscillogramme et TextGrid avec segmentation par syllabe 329
Figure 3.6 : dark\defdark2 : Courbe de F_0 avec valeur erronée 329
Figure 3.7 : dark\defdark2 : Courbe de F_0 avec plafond à 350 Hz et I de 40 à 100 dB 330
Figure 3.8 : Les champs associés aux différentes tables 340
Figure 3.9 : Présentation des composés selon des tris différents 353
Figure III.1 : Echantillonnage des extraits sonores en *dark* : mesures du temps 359
Figure III.2 : Echantillonnage des extraits sonores en *dark* : mesures du F_0 361
Figure III.3 : Echantillonnage des extraits sonores en *dark* : mesures de l'intensité 363
Figure 4.1 : Tableau récapitulatif des « doublons » 390
Figure 4.2 : Tableau récapitulatif du type d'accentuation par table des entrées de la base de données (« s » accentuation simple / « d » accentuation double) 393
Figure 4.3 : Accentuation simple ou double des composés par table et par nombre de syllabes : « s » accentuation simple / « d » accentuation double 394
Figure 4.4 : Tableau récapitulatif de la structure graphique des tables 397
Figure 4.5 : Comparaison des entrées graphiques des tables, graphie d'origine 400
Figure 4.6 : Comparaison des entrées graphiques des tables, graphie neutralisée 401
Figure 4.7 : Structure des entrées graphiques / schémas accentuels, table par table 406
Figure 4.8 : Eléments susceptibles de réduire : entrées à accentuation double 416
Figure 4.9 : Nombre d'entrées par catégorie syntaxique et par table : EPD, NPD, LPD 418
Figure 4.10 : Nombre d'entrées par catégorie syntaxique et par type de schema, par table 421
Figure 4.11 : Entrées communes EPD/LPD, par type de graphie 432
Figure 4.12 : "Evolution" du schéma accentuel EPD12/LPD 433
Figure 4.13 : Graphie des entrées dont le schéma accentuel est différent EPD12 LPD 438
Figure 4.14 : L'Accent prévu / perçu : extraits sonores figurant dans le dictionnaire 446
Figure 4.15 : newyork\newyork1 : Oscillogramme et courbes de F_0 et I 448
Figure 4.16 : hole\keyhole1 et hole\portrh1 : Courbes de F_0 et I 453
Figure 4.17 : newyork\newyork3 : Oscillogramme et courbes de F_0 et I 455

Figure 4.18 : vodka\talkbot1 : Courbes de F_0 et I 459
Figure 4.19 : debt\distinv1 : Courbes de F_0 et I 460
Figure 4.20 : dragon\dragbite1 : Courbes de F_0 et I 461
Figure 4.21 : test\pillarb1 et test\razorbl1 : Courbes de F_0 et I 465
Figure 4.22 : test\anyone1 : Courbe de F_0 et I 467
Figure 4.23 : debt\vultfun1 (locuteur GB) et debt\vultfun3 (locuteur US) : Courbes de F_0 et I 469
Figure 4.24 : room\trophy1 & room\trophy2 : Courbes de F_0 et I 471
Figure 4.25 : shift\cigcomp1 : Oscillogramme et courbe de F_0 et I 473
Figure 4.26 : oilseed\farmtri2 : Courbe de F_0 et I 476
Figure 4.27 : shift\selfstir1 & shift\self2 : Courbes de F_0 et I 478
Figure 4.28 : vodka\armchai1 & vodka\HParmcha3 (accentuation double) : Courbes de F_0 et I 482
Figure 4.29 : vodka\HParmcha1 & vodka\HParmcha2 (accentuation simple) : Courbes de F_0 et I 482
Figure 4.30 : Règles d'accentuation des composés 495
Figure 4.31 : room\flykeyrm : Courbe de F_0 et I 511
Figure 4.32 : test\hocuspo1 : Oscillogramme et courbe de F_0 et I 518
Figure 4.33 : dragon\dragblood1 & dragon\draghead1 : Courbes de F_0 et I 522
Figure 4.34 : test\birdsof : Oscillogramme et courbe de F_0 et I 526
Figure IV.1 : Type d'accentuation par table et par nombre de syllabes 543
Figure IV.2 : Informations sur les variantes graphiques de LPD 546
Figure IV.3 : Schéma accentuel pour les entrées communes EPD/LPD/N-COMPACT 547
Figure IV.4 : Schéma accentuel et type de graphie, par syllabe : LPD-Cp 548
Figure IV.5 : Type de schéma par catégorie syntaxique et par syllabe : LPD-Cp, NPD-Cp, EPD-Cp 549
Figure IV.6 : Influence des facteurs du déplacement de l'accent et la variation 550
Figure IV.7 : Graphie comparée EPD15/LPD des entrées différentes EPD12 LPD 550
Figure IV.8 : Extraits sonores, accent perçu et accent lexical : indications des dictionnaires 552
Figure IV.9 : Répertoire\oilseed : mesures pour les extraits sonores, par syllabe 559
Figure IV.10 : Répertoire\oilseed : mesures pour les extraits sonores, par unité 560
Figure IV.11 : Répertoire \vodka : mesures pour les extraits sonores, par syllabe 561
Figure IV.12 : Répertoire \vodka : mesures pour les extraits sonores, par unité 563
Figure IV.13 : broom\schow11 : Oscillogramme et courbes de F_0 et I 565
Figure IV.14 : debt\baddebt1 : Courbes de F_0 et I 566
Figure IV.15 : debt\deadcom1 : Courbes de F_0 et I 567
Figure IV.16 : debt\vultfun4 : Courbes de F_0 et I 568
Figure IV.17 : dragon\dragheart1 : Oscillogramme et courbes de F_0 et I 569
Figure IV.18 : dragon\dragliver1 : Courbes de F_0 et I 570
Figure IV.19 : emiss\emiss1 : Courbes de F_0 et I 571
Figure IV.20 : emiss\emiss2 : Courbes de F_0 et I 572
Figure IV.21 : hole\portrh3 et hole\portrh4 : Courbes de F_0 et I 573
Figure IV.22 : newyork\newyork2 : Courbes de F_0 et I 574
Figure IV.23 : shift\hardear1 : Courbes de F_0 et I 574
Figure IV.24 : shift\publici : Courbes de F_0 et I 575
Figure IV.25 : shift\thirty1 : Courbes de F_0 et I 576
Figure IV.26 : vodka\kitchen1 : Courbes de F_0 et I 576
Figure IV.27 : Sélection d'entrées de N-COMPACT 579
Figure IV.28 : Entrées contenant l'élément *break* 582
Figure IV.29 : Entrées contenant l'élément *fly* 583
Figure IV.30 : Entrées contenant l'élément *green* 584
Figure IV.31 : Entrées contenant l'élément *work* 585

CONTENU DU CDROM

CONTENU DU CDROM

thèse

Ensemble des fichiers de la thèse : TH00.doc à TH09.doc.

corporal

Les extraits sonores du corpus oral d'analyse, selon les répertoires indiqués en Annexe E, chapitre 2.

base

Fichier COMPOSES00.mdb : la base de données comprenant les tables des entrées du corpus d'analyse et les requêtes.