

Formation doctorale : Langage et Parole
Discipline : phonétique anglaise

[illegible]

pour obtenir le grade de
DOCTEUR DE L'UNIVERSITE AIX-MARSEILLE I

présentée et soutenue publiquement
le 3 décembre 2004
par

Caroline Bouzon

Rythme et structuration prosodique en anglais britannique contemporain

Sous la direction de
Monsieur Daniel Hirst
Directeur de recherche, CNRS, UMR 6057
Laboratoire Parole et Langage – Université de Provence

JURY

Jean-Louis Duchet, Professeur à l'université de Poitiers (rapporteur)
Sophie Herment-Dujardin, Maître de Conférences à l'université de Provence
Daniel Hirst, Directeur de recherche CNRS à l'université de Provence
Gerry Knowles, Senior Lecturer à l'université de Lancaster
Michel Viel, Professeur à l'université Paris Sorbonne-Paris IV (rapporteur)

*A mes parents,
ma famille,
Tony*

REMERCIEMENTS

Je tiens à remercier de nombreuses personnes sans qui cette thèse ne serait pas. Je remercie donc tout particulièrement :

Daniel Hirst

qui me dirige depuis maintenant 5 ans et que je remercie grandement pour sa disponibilité, ses conseils avisés, son soutien et sa patience.

Jean-Louis Duchet, Gerry Knowles et Michel Viel

d'avoir accepté de faire partie de mon jury et de s'être déplacés.

Sophie Herment-Dujardin

d'avoir accepté d'être dans mon jury mais également pour ses encouragements, ses conseils et sa bonne humeur.

Le Laboratoire Parole et Langage

qui m'a donné tous les moyens pour réussir ma thèse.

Le Département d'Etudes du Monde Anglophone

qui m'a permis de finir ma thèse dans les meilleures conditions.

Cyril Auran

d'abord en tant que collègue... Ce fut un grand plaisir de travailler avec toi sur le projet Aix-MARSEC et plus globalement dans le cadre d'EPGA. Merci aussi pour les encouragements et tous les nombreux coups de main, notamment en terme de programmation. Continue à faire des phrases et à utiliser le terme *grosso modo* mais pense un peu plus au point ... !

Isabelle Marlien,

pour sa relecture inestimable (et donc sa patience !), ses conseils et sa gentillesse, et plus généralement pour tous les nombreux moments partagés dans la salle de lecture. Merci aussi pour ce moment très apprécié dans le monde du silence !

Robert Espesser

pour les nombreuses heures passées à m'aider sur les analyses statistiques et, évidemment, pour sa patience.

Cyril Deniaud et Louis Sémandi

pour m'avoir de nombreuses fois dépannée sur des problèmes techniques et pour m'avoir facilité la réalisation de mon expérience en production.

Valérie Kerfelec, Laetitia Leonarduzzi,

et plus généralement les collègues du département d'anglais enseignant la phonétique pour leurs conseils et leurs encouragements, mais aussi pour leur compréhension au moment des examens.

Les sujets anglophones

qui ont bien voulu se prêter à mon expérience en production.

Sandrine Clairet et Annelise Coquillon

Pour tous les moments de détente et de rigolade au bureau des doctorants, mais aussi pour leurs encouragements et leurs conseils.

Martine Faraco

Pour son soutien, ses conseils et par dessus tout sa gentillesse.

Les doctorants du Laboratoire Parole et Langage

notamment Manu, Malie, Irina, Cristel, Tristan, Katia, Julie, etc. qui contribuent à la bonne ambiance de la salle des doctorants.

J'ai également eu la chance de bénéficier d'un support inestimable de ma famille et de mes amis. Je tiens donc à adresser mes plus sincères remerciements à :

Mes parents

sans qui le parcours que j'ai eu ne serait pas. Ils m'ont toujours soutenue dans mes choix et mes décisions et m'ont toujours donné les moyens nécessaires à ma réussite. Malgré les moments difficiles, je leur dois tout. Il n'y a pas assez de mots pour que je puisse un jour les remercier suffisamment pour tout ce qu'ils m'ont apporté, je leur dis simplement « merci à tous les deux du fond du cœur ».

Ma famille

qui m'a toujours soutenue et encouragée et qui a toujours été présente quand j'en avais besoin.

Tony

pour son soutien, ses encouragements, sa présence, sa compréhension. Je te suis extrêmement reconnaissante à tous les points de vue, sans toi, je pense que je n'y serais pas arrivée.

Cyril Auran (cette fois en tant qu'ami !!)

ce travail de thèse aura été l'occasion de rencontrer un ami inestimable sur qui j'ai pu compter de nombreuses fois et avec qui on a beaucoup ri et avec qui on rira, j'en suis sûre, encore de nombreuses années.

Thi Hoarau

pour son amitié si précieuse, sa gentillesse, ses encouragements et tous les moments de bonne rigolade, sans oublier les « instants poltergeist » évidemment. Merci également de nous avoir supportés avec Cyril quand on ne pouvait pas s'empêcher de parler boulot... !

tous mes amis d'ici ou d'ailleurs

notamment Dada, Rémi, Fafa, Maria, John, Ellen, David, Joëlle, Kizzi et tous les autres que j'oublie.

et enfin, un clin d'œil à mes quadrupèdes !

TABLE DES MATIERES

TABLE DES ILLUSTRATIONS	19
INTRODUCTION	27
<hr/>	
PARTIE I. ASPECTS THEORIQUES	35
CHAPITRE 1. ORGANISATION TEMPORELLE ET MODELES DE PREDICTION	37
1.1. ORGANISATION TEMPORELLE	39
1.1.1. Caractéristiques générales	39
1.1.2. Problèmes soulevés par cette tâche prédictive	40
1.2. DEFINITION DE LA OU DES UNITE(S) TEMPORELLE(S)	42
1.3. TROIS UNITES, TROIS TYPES DE MODELES	45
1.3.1. "Modèles segmentaux"	45
1.3.1.1. Présentation	45
1.3.1.2. Arguments en faveur d'une approche segmentale	47
1.3.1.3. Limites de l'approche segmentale	50
1.3.2. "Modèles syllabiques"	54
1.3.2.1. Présentation	54
1.3.2.2. Problème : la syllabification en anglais	58
1.3.3. "Modèles supra-syllabiques"	59
1.3.3.1. Présentation générale	59
1.3.3.2. Problème : définition de l'unité rythmique	59
1.4. CONTROLE DU DEBIT DE PAROLE	60
1.5. RESUME ET CONCLUSION	64

CHAPITRE 2. PROBLEMATIQUE DE L'AMBISYLLABICITE.....	67
2.1. INTRODUCTION : ROLES DE LA SYLLABE	69
2.2. STRUCTURE INTERNE DE LA SYLLABE	73
2.2.1. Représentation linéaire de la syllabe	73
2.2.2. Représentation non-linéaire de la syllabe	74
2.2.2.1. Introduction de la structure hiérarchique	74
2.2.2.2. Phonologie autosegmentale et phonologie CV	74
2.2.2.3. Phonologie de Dépendance	75
2.2.2.4. Phonologie métrique	76
2.3. AMBISYLLABICITE : PROBLEMATIQUE DE LA SYLLABIFICATION	80
2.3.1. Syllabification et resyllabification : définitions.....	80
2.3.2. Principe d'attaque maximale et contrainte de rime branchante	82
2.3.3. L'ambisyllabité	85
2.3.3.1. Contraintes phonotactiques de syllabification	86
2.3.3.2. Variations allophoniques : aspiration, glottalisation et <i>flapping</i>	90
2.3.3.3. Syllabification et coarticulation	92
2.3.3.4. Travaux en psycholinguistique	96
2.3.3.5. Synthèse	97
2.3.4. Syllabe vs syllabité	99
2.4. RESUME ET CONCLUSION	100
 CHAPITRE 3. STRUCTURATION RYTHMIQUE : DEFINITION DES UNITES	 101
3.1. DEFINITION	104
3.2. DESCRIPTION DES UNITES RYTHMIQUES	107
3.2.1. Introduction.....	107
3.2.2. Modèle proposé par Jassem	109
3.2.2.1. Description.....	109
3.2.2.2. Approche similaire proposée par O'Connor	113

3.2.3.	Le pied accentuel.....	115
3.2.3.1.	Approche d'Abercrombie.....	115
3.2.3.2.	Approche de Halliday.....	117
3.2.4.	Comparaison des deux modèles	117
3.3.	COMPARAISON ET PERSPECTIVES DES DEUX MODELES	122
3.3.1.	Traitement différentiel des syllabes inaccentuées.....	122
3.3.1.1.	Syllabes inaccentuées post-accentuelles	123
3.3.1.2.	Syllabes inaccentuées pré-accentuelles.....	126
3.3.2.	Rôle et fonction du mot.....	127
3.3.3.	Confusion entre le niveau du pied et le niveau du mot	130
3.3.4.	Facteur syntaxique.....	132
3.3.5.	Limitation d'étendue des facteurs rythmiques	133
3.4.	PHENOMENE ACCENTUEL EN ANGLAIS.....	134
3.4.1.	Définition	135
3.4.2.	Corrélat acoustiques de l'accent	136
3.4.3.	Distinction entre accent (lexical) et proéminence.....	138
3.4.4.	Prédiction de l'accent et de la proéminence.....	140
3.5.	RESUME ET CONCLUSION.....	142

CHAPITRE 4. L'ISOCRONIE : « ILLUSION PERCEPTIVE » OU REALITE ?143

4.1.	DEFINITION	146
4.2.	PEUT-ON PARLER D'ISOCRONIE EN PRODUCTION ?	147
4.2.1.	Facteurs influençant la durée des intervalles accentuels.....	147
4.2.2.	Isochronie stricte ou isochronie faible ?.....	150
4.2.3.	Problème lié à l'isochronie : comment la mesurer ?	152
4.3.	L'ISOCRONIE EN TANT QUE PHENOMENE PERCEPTIF	154
4.4.	CATEGORISATION DES LANGUES	157
4.4.1.	Historique.....	157
4.4.2.	Échec de la catégorisation classique des langues en fonction du rythme .	160

4.4.3.	Approches alternatives du rythme de la parole.....	161
4.4.4.	Exemple de différence rythmique intra-langue	167
4.4.5.	Identification automatique des langues à partir des particularités rythmiques 167	
4.5.	RESUME ET CONCLUSION	171
<hr/>		
PARTIE II. BASE DE DONNEES ET RESULTATS EXPERIMENTAUX		175
CHAPITRE 5. CONSTITUTION ET EXPLOITATION D'UNE BASE DE DONNEES ORALES :		
PROJET AIX-MARSEC.....		177
5.1.	QUEL(S) CORPUS ?.....	179
5.1.1.	Types de corpus	180
5.1.2.	Qu'entend-on par parole "spontanée" ?	183
5.1.3.	Transcription et annotation de corpus oraux.....	185
5.1.4.	Disponibilité des corpus.....	187
5.2.	DESCRIPTION DU CORPUS.....	188
5.2.1.	Le corpus d'origine : du SEC au MARSEC	188
5.2.2.	Homogénéisation des données et pré-traitements.....	190
5.3.	TRANSCRIPTION PHONETIQUE DU CORPUS.....	190
5.3.1.	Phonétisation brute	191
5.3.1.1.	Méthodologie	191
5.3.1.2.	Traitements spécifiques	192
5.3.2.	Optimisation de la phonétisation	195
5.3.2.1.	Méthodologie	195
5.3.2.2.	Règles d'élision de phonèmes.....	197
5.3.2.3.	Evaluation de l'efficacité des règles d'élision	199
5.4.	ALIGNEMENT AUTOMATIQUE DES PHONEMES	200
5.4.1.	Méthodologie	200
5.4.2.	Evaluation et perspectives d'amélioration de l'alignement automatique .	201

5.5. AUTRES NIVEAUX D'ANALYSE ET D'ALIGNEMENT	202
5.5.1. Regroupement en syllabes et en constituants syllabiques	202
5.5.2. Regroupement des syllabes en unités rythmiques	204
5.5.3. Mots et unités intonatives	206
5.5.4. MOMEL & INTSINT	206
5.5.5. Aix-MARSEC : une base de données pluri-linéaire	207
5.6. MISE EN FORME DES DONNEES	209
5.6.1. Description des divers paramètres	209
5.6.2. Remarques sur les découpages syllabiques	213
5.7. RESUME ET CONCLUSION	214
 CHAPITRE 6. COMPARAISON ET EVALUATION DES DEUX MODELES RYTHMIQUES.....	215
6.1. HYPOTHESES ET OUTILS STATISTIQUES	217
6.1.1. Récapitulatif	217
6.1.2. Hypothèses	218
6.1.2.1. Syllabe	219
6.1.2.2. Pied accentuel	219
6.1.2.3. Unité rythmique et anacrouse	220
6.1.2.4. Structure prosodique et accentuation	220
6.1.3. Données et outils statistiques	220
6.1.3.1. Homogénéisation des données	221
6.1.3.2. Outils statistiques	223
6.2. RESULTATS ET INTERPRETATION	225
6.2.1. Isochronie stricte	225
6.2.1.1. Passage C01	225
6.2.1.2. Totalité des données	228
6.2.2. Isochronie faible	230
6.2.2.1. Passage C01	230
6.2.2.2. Totalité des données	238
6.2.3. Effet du paramètre accentuel	243

6.2.3.1. Passage C01	243
6.2.3.2. Totalité des données.....	251
6.3. RESUME ET CONCLUSION	257
CHAPITRE 7. SYLLABIFICATION EN ANGLAIS BRITANNIQUE	259
7.1. HYPOTHESES ET OUTILS STATISTIQUES	262
7.1.1. Hypothèses.....	262
7.1.2. Outil statistique	264
7.1.3. Expérience en production	264
7.2. TEST SUR LA BASE DE DONNEES AIX-MARSEC	265
7.2.1. Résultats dans le passage C01	265
7.2.2. Résultats dans la totalité des données Aix-MARSEC	267
7.2.3. Conclusion	267
7.3. EXPERIENCE EN PRODUCTION	269
7.3.1. Elaboration des consignes.....	270
7.3.1.1. Phase d'entraînement (sujets francophones).....	271
7.3.1.2. Phase d'entraînement (sujet anglophone)	272
7.3.1.3. Ajout d'une deuxième partie.....	273
7.3.2. Elaboration de l'expérience	274
7.3.2.1. Stimuli.....	274
7.3.2.2. Aspects techniques.....	275
7.3.2.3. Sujets.....	276
7.3.3. Mise en forme des données.....	276
7.3.4. Interprétation des résultats	281
7.3.4.1. Répartition des différents types de découpages syllabiques.....	281
7.3.4.2. Occurrences de consonnes ambisyllabiques	285
7.3.4.3. Regroupement en position codaïque.....	289
7.3.4.4. Regroupement selon le principe d'attaque maximale.....	292
7.3.4.5. Comparaison des découpages des Phases 3 et 4.....	294
7.3.4.6. Bilan de l'expérience	296

7.4. RESUME ET CONCLUSION.....	298
CONCLUSION.....	301
REFERENCES BIBLIOGRAPHIQUES.....	309
INDEX.....	327
ANNEXES	331
TABLE DES ANNEXES.....	333
1. ANNEXE 1. CORRESPONDANCE SYMBOLES SAMPA/ IPA.....	335
2. ANNEXES 2. ISOCHRONIE ET CONSTITUANCE.....	336
2.1. Isochronie stricte	336
2.1.1. Passage C01 : Tableau 7 (page 225)	336
2.1.2. Totalité des données : Tableau 8 (page 227).....	342
2.2. Isochronie faible.....	349
2.2.1. Passage C01.....	349
2.2.2. Totalité des données	366
2.3. Effet du paramètre accentuel.....	384
2.3.1. Passage C01.....	384
2.3.2. Totalité des données	399
3. ANNEXES 3. SYLLABIFICATION EN ANGLAIS BRITANNIQUE	414
3.1. Résultats du test t (passage C01).....	414
3.2. Résultats du test t (totalité des données)	417
3.3. Fonctionnement du test de production	420
3.4. Phrases de l'expérience en production et différentes phases de lecture.....	421
3.4.1. Première partie de l'expérience.....	421
3.4.2. Deuxième partie de l'expérience.....	422
3.5. Matrices des effectifs	423












3.5.1.	Matrices des effectifs pour les regroupements ambisyllabiques	423
3.5.2.	Matrices des effectifs pour les regroupements codaïques	424
3.6.	Résultats des χ^2	425
3.6.1.	Regroupements ambisyllabiques	425
3.6.2.	Regroupements codaïques	427
CONTENU DES CD-ROM		429
CD-ROM 1. LA BASE DE DONNEES AIX-MARSEC		429
	Audio	429
	Dictionnaires.....	429
	Sripts	429
	TextGrid.....	429
	données.zip	429
	thèse.pdf.....	429
CD-ROM 2. EXPERIENCE EN PRODUCTION		430
	Données	430
	Enregistrements	430
	Présentations	430
	get_pauses_v2.6.praat (script de détection des pauses).....	430
CD-ROM 3. EXPERIENCE EN PRODUCTION (SUITE).....		431
	Enregistrements	431

TABLE DES ILLUSTRATIONS

Figures

<i>Figure 1. Représentation des différents types d'allongement au sein de la syllabe (Campbell 1992a: 123)</i>	<i>56</i>
<i>Figure 2. Représentation de la syllabe dans le cadre de la phonologie autosegmentale</i>	<i>75</i>
<i>Figure 3. Représentation de la syllabe en phonologie métrique (citée par Meynadier 2001 :115)</i>	<i>77</i>
<i>Figure 4. Regroupement des phonèmes en syllabes dans le mot city, avec double affiliation de la consonne intervocalique /t/ (ambisyllabité).....</i>	<i>85</i>
<i>Figure 5. Découpage syllabique de l'énoncé these are old eggs (Giegerich 1992 : 280)</i>	<i>89</i>
<i>Figure 6. Représentation des énoncés summer dresses et some addresses selon le modèle de Jassem (1951)</i>	<i>112</i>
<i>Figure 7. Représentation de l'énoncé summer dresses selon les deux modèles rythmiques ('S' correspond à 'Syllable' et 'F' à 'Foot').....</i>	<i>118</i>
<i>Figure 8. Représentation de l'énoncé some addresses selon les deux modèles supra-syllabiques.....</i>	<i>118</i>
<i>Figure 9. Regroupements syllabique et rythmique des énoncés player (1) et play a (2) selon l'approche d'Abercrombie (partie haute de la figure) et selon l'approche de Jassem (partie basse de la figure).....</i>	<i>119</i>
<i>Figure 10. Représentation d'énoncés identiques en fonction du modèle de Jassem (colonne de gauche) et de celui d'Abercrombie (colonne de droite)</i>	<i>120</i>
<i>Figure 11. Exemples de grilles métriques pour un énoncé anglais et un énoncé italien (Selkirk 1984a :40-41)</i>	<i>158</i>

Figure 12. Regroupement des différentes langues selon leurs caractéristiques rythmiques (Ramus 1999 : 41).....	159
Figure 13. Courbe rythmique représentant les différences de tension dans le cadre de la syllabe (Wenk 1983: 157)	163
Figure 14. Axe représentant le degré d'importance de l'accent dans différentes langues (Dauer 1983: 60)	164
Figure 15. Représentation tridimensionnelle des divers styles de parole en fonction du degré d'intelligibilité, de la familiarité et de l'appartenance sociale (Eskénazi 1993: 503).....	183
Figure 16. Représentation des différents corpus en fonction du type de parole	184
Figure 17. Liste des types d'enregistrement dans le SEC.....	188
Figure 18. Liste des 14 symboles ASCII utilisés pour l'annotation prosodique du MARSEC (Roach 1994 : 94)	189
Figure 19. Formules permettant d'obtenir le facteur z (2) en fonction de la durée du mot (1) et de la moyenne et de l'écart type de chaque phonème	195
Figure 20. Attaques bi-consonantiques licites (Gimson réédité par Cruttenden 1997: 218).....	203
Figure 21. Attaques tri-consonantiques licites (Gimson réédité par Cruttenden 1997 : 218).....	203
Figure 22. Illustration d'un extrait du corpus Aix-MARSEC (dans Praat) avec les neuf niveaux d'alignement.....	208
Figure 23. Regroupement des phonèmes en syllabes en tenant compte du principe d'ambisyllabité pour les mots <i>degree</i> et <i>inquisition</i>	214
Figure 24. Répartition des durées segmentales brutes dans la base de données Aix-MARSEC.....	223
Figure 25. Box plot de la durée des syllabes en fonction du nombre de phonèmes qu'elles contiennent (C01)	228
Figure 26. Durée normalisée des phonèmes en fonction du nombre de phonèmes dans la syllabe (passage C01).....	234
Figure 27. Durée normalisée des phonèmes en fonction de la complexité de la NRU et du statut accentuel (passage C01).....	244

<i>Figure 28. Durée normalisée moyenne des phonèmes dans la NRU (dont on a enlevé les deux phonèmes en fin d'unité intonative) en fonction de la complexité en phonèmes et de l'accent</i>	<i>245</i>
<i>Figure 29. Représentation des durées normalisées des phonèmes en fonction du nombre de phonèmes dans le pied et du statut accentuel (C01)</i>	<i>246</i>
<i>Figure 30. Durée brute moyenne des syllabes en fonction du statut accentuel et de la complexité de la syllabe</i>	<i>248</i>
<i>Figure 31. Représentation des durées syllabiques brutes en fonction du paramètre accentuel et du nombre de syllabes dans la NRU (passage C01).....</i>	<i>249</i>
<i>Figure 32. Représentation des durées syllabiques brutes en fonction du paramètre accentuel et du nombre de syllabes dans le pied accentuel (passage C01).....</i>	<i>249</i>
<i>Figure 33. Durée normalisée des phonèmes en fonction de la complexité de la NRU et du paramètre accentuel (totalité des données)</i>	<i>252</i>
<i>Figure 34. Durées normalisées moyennes des phonèmes dans la NRU en fonction de l'accent et de la complexité en ayant au préalable supprimé les deux derniers phonèmes des NRU en position finale d'unité intonative</i>	<i>253</i>
<i>Figure 35. Représentation des durées normalisées des phonèmes en fonction de la complexité du pied et du paramètre accentuel (totalité des données)</i>	<i>254</i>
<i>Figure 36. Représentation des durées syllabiques en fonction du paramètre accentuel et de la complexité des syllabes (totalité des données)</i>	<i>255</i>
<i>Figure 37. Représentation des durées syllabiques dans la NRU en fonction du paramètre accentuel et de la complexité de la NRU (totalité des données).....</i>	<i>256</i>
<i>Figure 38. Représentation des durées syllabiques dans le pied accentuel en fonction du paramètre accentuel et de la complexité du pied (totalité des données)</i>	<i>256</i>
<i>Figure 39. Double codage des consonnes intervocaliques dans la base de données Aix-MARSEC selon le principe d'attaque maximale et celui d'ambisyllabité</i>	<i>263</i>
<i>Figure 40. Extrait de l'étiquetage de la Phase 3 du fichier Aphr13ph3.wav (the passenger picked up some lilies, daisies and daffodils) avec au premier niveau la transcription phonémique et l'alignement temporel au niveau de la syllabe et au deuxième niveau la transcription orthographique et l'alignement temporel au niveau du mot</i>	<i>277</i>
<i>Figure 41. Illustration de la répartition des types de découpage syllabique en fonction de la phase de lecture (Phases 2, 3 et 4).....</i>	<i>282</i>

<i>Figure 42. Différents types de découpage syllabique en fonction des locuteurs.....</i>	<i>284</i>
<i>Figure 43. Illustration dans Praat de la Phase 4 de la phrase 4 lue par le locuteur B</i>	<i>285</i>
<i>Figure 44. Nombre d'occurrences de cas d'ambisyllabité en fonction des locuteurs</i>	<i>286</i>
<i>Figure 45. Phase 4 de la phrase 1 lue par le locuteur A.....</i>	<i>290</i>
<i>Figure 46. Phase 3 de la phrase 16 lue par le locuteur A.....</i>	<i>293</i>

Tableaux

<i>Tableau 1. Occurrences des différentes séquences en fonction de leur position initiale ou finale de mot (O'Connor et Trim 1953 : 121)</i>	<i>83</i>
<i>Tableau 2. Présentation des onze principes d'élision de phonèmes</i>	<i>197</i>
<i>Tableau 3. Résultats de l'évaluation des élisions</i>	<i>199</i>
<i>Tableau 4. Résultats de l'évaluation de l'alignement automatique</i>	<i>202</i>
<i>Tableau 5. Énumération des différentes informations présentes dans le tableau de données.....</i>	<i>213</i>
<i>Tableau 6. Résumé des différentes prédictions selon les hypothèses de corrélation concernant l'isochronie.....</i>	<i>225</i>
<i>Tableau 7. Résultats des analyses de régression linéaire concernant l'hypothèse d'isochronie stricte pour le passage C01. Les lignes correspondent aux variables dépendantes (durée des constituants) et les colonnes aux variables indépendantes (complexité).....</i>	<i>227</i>
<i>Tableau 8. Résultats des analyses de régression linéaire pour l'hypothèse d'isochronie stricte sur la totalité des données avec comme variable dépendante la durée des constituants et comme variable indépendante leur complexité.....</i>	<i>229</i>
<i>Tableau 9. Résultats des analyses de régression linéaire en ce qui concerne l'hypothèse d'isochronie faible dans le passage C01. Les lignes représentent les variables indépendantes (complexité des différents constituants en terme phonèmes) et les colonnes les variables dépendantes (durée des phonèmes). 'zp' et 'dp' sont respectivement les durées normalisée et brute des phonèmes.</i>	<i>231</i>
<i>Tableau 10. Résultats des analyses de régression (C01) concernant la durée des phonèmes en fonction du nombre de syllabes dans les constituants supérieurs...</i>	<i>235</i>
<i>Tableau 11. Résultats des analyses de régression (C01) concernant la durée des phonèmes en fonction du nombre de pieds dans l'unité intonative</i>	<i>236</i>
<i>Tableau 12. Résultats des analyses de régression (C01) concernant la durée des syllabes en fonction de la complexité des constituants supérieurs en terme de nombre de syllabes.....</i>	<i>237</i>
<i>Tableau 13. Résultats des analyses de régression linéaire pour l'hypothèse d'isochronie faible : durée des phonèmes en fonction de la complexité des constituants en terme de phonèmes (totalité des données).....</i>	<i>238</i>

Tableau 14. Résultats des analyses de régression linéaire avec comme variable dépendante la durée des phonèmes et comme variable indépendante la complexité des constituants supérieurs en terme de syllabes (totalité des données).....	240
Tableau 15. Résultats des analyses de régression avec comme variable dépendante la durée des phonèmes et comme variable indépendante la complexité de l'unité intonative en terme de nombre de pieds (totalité des données).....	241
Tableau 16. Résultats de l'analyse de régression avec comme variable dépendante la durée des syllabes (brute et exprimée en log) et comme variable indépendante la complexité des constituants supérieurs en terme de nombre de syllabes qu'ils contiennent (totalité des données)	242
Tableau 17. Extrait du tableau global de données (tableau(global).xls, CD-ROM 2).	279
Tableau 18. Extrait du tableau2(ambi).xls (CD-ROM 2).....	281
Tableau 19. Extrait du Tableau3(coda).xls (CD-ROM 2).....	281
Tableau 20. Répartition en pourcentages des différents types de découpages syllabiques pour les 14 locuteurs.....	282
Tableau 21. Répartition exprimée en pourcentages des types de découpage (Attaque, Coda et Ambisyllabique) en fonction de la phase de lecture (2, 3 ou 4).....	283
Tableau 22. Matrice d'effectifs pour la catégorie de consonne intervocalique selon qu'il s'agit d'une sonorante ou une obstruante et selon qu'elle est ambisyllabique ou pas.....	287
Tableau 23. Résultats des χ^2 pour les quatre facteurs testés sur les cas de regroupement ambisyllabique.....	288
Tableau 24. Résultats des χ^2 pour les regroupements codaïques.....	291

INTRODUCTION

La communication se fait principalement par le biais de la parole. Dans notre vie quotidienne, nous planifions constamment des messages plus ou moins complexes. Ces messages sont construits en fonction d'un certain nombre de règles morpho-syntaxiques. Ils sont ensuite produits oralement dans le but de communiquer leur contenu sémantique ; cette phase de production ne découle pas d'une simple concaténation des mots les uns à la suite des autres car le message résultant d'une telle production serait absolument incompréhensible. La chaîne de formes de citation doit au contraire être transformée en une suite intelligible de syllabes (Levelt 1989). Cette transformation est très complexe et elle passe par de nombreuses phases.

C'est grâce à la prosodie que se fait le passage de la concaténation de formes isolées à la production d'un message compréhensible. Pour reprendre la terminologie de Levelt, c'est le « générateur prosodique » qui permet le passage de la forme de surface (résultant de l'encodage grammatical) à la forme phonétique prête à être produite par les articulateurs. La prosodie possède par conséquent une fonction intégrante d'organisation de la chaîne parlée :

« Prosody has an integrating function in the organisation and production of speech by embedding semantic information, syntactic and morphological structure and the segmental chain in a consistent address frame (...) » (Möbius et Van Santen 2000 : 161)

La prosodie permet le regroupement des informations sémantiques, de la structure morpho-syntaxique et de la chaîne segmentale, elle permet par conséquent la structuration des messages et la création du sens. Il est donc évident qu'elle joue un rôle fondamental en communication.

La structuration prosodique des énoncés est tridimensionnelle : elle se fait en fonction d'une dimension tonale, d'une dimension métrique et d'une dimension

temporelle (Di Cristo 2000). Cette caractéristique tridimensionnelle soulève la question de la définition d'un modèle de représentation de la structure prosodique d'une langue. Hirst (1998) précise en effet que pour l'anglais :

« It is not, however, clear whether it is possible to account for both rhythm and melody by a single model of prosodic structure. » (Hirst 1998: 58)

C'est dans cette optique que Jassem avait proposé en 1952 un modèle distinguant l'unité rythmique totale et l'unité tonale ; la première permet de rendre compte de la structure rythmique alors que la deuxième représente le domaine des variations mélodiques. Sans pour autant négliger l'importance de la dimension tonale dans la production langagière, ce travail se focalise sur les dimensions très proches que sont les aspects temporels et métriques.

La dimension temporelle, ou « organisation temporelle » (terme emprunté à Nooteboom 1991), est capitale en prosodie car elle véhicule tout un ensemble d'informations linguistiques. Elle joue un rôle fondamental dans la structuration de l'information dans le sens où elle permet de signaler les différentes frontières et ainsi de diviser la chaîne parlée en groupes syntaxiques et rythmiques. Par conséquent, un changement dans l'organisation temporelle signale un regroupement différent des blocs d'information et ainsi du sens global d'un énoncé (Campbell 1996 : 232). Elle contribue d'ailleurs selon Möbius et Van Santen à l'intégration des informations sémantiques, de la structure morpho-syntaxique et de la chaîne segmentale (suite de la citation donnée à la page précédente) :

« (...) Temporal properties of the prosodic components crucially contribute to this integration. » (Möbius et Van Santen 2000 : 161)

Il convient donc de rendre compte de manière efficace de l'organisation temporelle.

Etant donné cette importance, de nombreux travaux ont été consacrés à la prédiction de l'organisation temporelle dans de nombreuses langues, notamment dans l'optique de développer et d'améliorer les modèles de synthèse vocale. Il est d'ailleurs intéressant de noter que la dimension temporelle a été longtemps négligée au profit de la mélodie ; or, on s'est rendu compte que l'organisation temporelle jouait un rôle tout aussi important et qu'elle devait donc être prédite avec exactitude. Dans ce cadre là, l'anglais est l'une des langues les plus étudiées. Toutefois, si la littérature concernant la prédiction de

l'organisation temporelle de l'anglais est abondante, une question reste toutefois ouverte, celle de l'unité sur laquelle fonder un modèle de prédiction.

Les unités utilisées sont variées mais globalement on en retient trois : les phonèmes (ou segments), les syllabes et les unités rythmiques. Le choix de l'unité de prédiction induit des approches différentes que nous nommerons respectivement approches « segmentale », « syllabique » et « supra-syllabique ». La question de la prédiction de l'organisation temporelle en anglais et la problématique de l'unité sont détaillées dans le Chapitre 1 de ce travail. Ce premier chapitre est introductif dans le sens où, à partir des remarques concernant les différents modèles de prédiction, nous soulèverons deux problématiques que nous testerons dans la partie expérimentale (Chapitres 7 et 8).

Dans ce premier chapitre, nous décrivons également le fonctionnement des modèles segmentaux qui cherchent à prédire la durée des segments (ou phonèmes) en fonction d'un certain nombre de facteurs. On reproche souvent à cette approche de ne pas tenir compte des caractéristiques suprasegmentales de la parole, comme par exemple la structure rythmique. De plus, ces modèles se trouvent confrontés à une variation extrêmement importante des durées segmentales, rendant par conséquent leur prédiction délicate. C'est principalement en raison de ces difficultés que les approches syllabiques et supra-syllabiques ont été proposées.

Le deuxième type d'approche concerne les modèles syllabiques qui cherchent à prédire la durée des syllabes pour ensuite prédire la durée des phonèmes dans le cadre de la syllabe. Le problème posé directement par cette approche, ainsi que par toutes les approches *supra*-segmentales, concerne la question de la syllabification en anglais : pour prédire la durée des syllabes, il faut au préalable découper les énoncés en syllabes. Ce problème concerne principalement la syllabification des consonnes intervocaliques, qu'elles soient à l'intérieur d'un mot ou à la frontière entre deux mots. En cas de consonne intervocalique, plusieurs principes de syllabification entrent en concurrence. Cette problématique de syllabification fait l'objet du Chapitre 2 dans lequel nous rappelons dans un premier temps les théories fondamentales concernant la syllabe, puis nous décrivons les différents principes de syllabification en anglais, avec une attention particulière pour le principe d'ambisyllabité. Selon ce principe, une consonne intervocalique est rattachée aux deux syllabes environnantes en respectant les contraintes de cooccurrence propres à la langue en question. De nombreuses études ont fait référence à ce principe, en l'intégrant ou non à leur modèle, mais peu d'entre elles

cherchent à tester ce phénomène de manière expérimentale. La question est donc de savoir si l'ambisyllabité fait partie intégrante du système phonologique de l'anglais britannique. La question de la syllabification et du statut du phénomène d'ambisyllabité en anglais britannique représente par conséquent le premier centre d'intérêt de ce travail.

Le troisième type d'approche porte sur les unités supra-syllabiques. Les modèles supra-syllabiques cherchent à prédire la durée des constituants de taille supérieure à la syllabe pour ensuite "déduire" la durée de leur sous-constituants (phonèmes et/ ou syllabes) en fonction d'un ensemble de facteurs. Dans cette catégorie, on trouve notamment les modèles cherchant à prédire la durée des unités rythmiques dans le but d'obtenir une prédiction qui tienne compte des informations relatives à la structure rythmique. Ces modèles sont abordés dans le Chapitre 3 dans lequel nous focalisons cette description sur deux modèles, le modèle de Jassem (1952) et le modèle classique d'Abercrombie (1964)/ Halliday (1967). Ces deux modèles diffèrent fondamentalement dans le regroupement des syllabes en unités rythmiques et par voie de conséquence dans les prédictions temporelles qui en découlent. Tout un ensemble de travaux sur la structuration rythmique est ensuite cité et mis en parallèle avec les prédictions faites par les deux modèles supra-syllabiques auxquels nous nous intéressons ici.

Les modèles supra-syllabiques sont directement liés à la notion d'isochronie. En anglais, les unités rythmiques sont supposées avoir une durée quasi constante, quel que soit le nombre de sous-constituants qu'elles contiennent. Ce principe a donné lieu à une catégorisation des différentes langues du monde en langues à rythme accentuel, syllabique ou moraïque. Ces notions sont détaillées dans le Chapitre 4 dans lequel nous passons également en revue l'abondante littérature visant à prouver qu'il n'existe pas en anglais d'isochronie, notamment dans sa définition stricte, et par voie de conséquence que la distinction des langues sur la base de leurs caractéristiques rythmiques ne tient pas. Par conséquent, l'hypothèse d'isochronie, dans sa version stricte ou faible, ainsi que la validation d'un des deux modèles rythmiques représentent le deuxième centre d'intérêt de ce travail.

L'objectif de ce travail n'est pas d'élaborer un modèle de prédiction de l'organisation temporelle de l'anglais, mais de tester quelques pistes qui nous semblent fondamentales pour une prédiction potentiellement fiable. Les caractéristiques rythmiques doivent être prises en considération dans un modèle prédictif de l'anglais, la

question étant de savoir quelle unité choisir. De plus, la plupart des modèles se trouvent confrontés à un moment donné au découpage syllabique des énoncés, il convient donc d'explorer les stratégies de syllabification. En conséquence, les deux problématiques centrales de ce travail sont la comparaison des prédictions des modèles supra-syllabiques de Jassem et d'Abercrombie/ Halliday et le phénomène d'ambisyllabité en anglais britannique. Ces informations concernant le rythme et la structuration prosodique nous semblent fondamentales, notamment dans le cadre de la prédiction de l'organisation temporelle de l'anglais britannique.

Dans le but d'explorer ces deux problématiques, il est nécessaire d'avoir une base de données adéquate, suffisamment dense pour rendre compte de la diversité des phénomènes testés et naturelle pour pouvoir observer les caractéristiques rythmiques authentiques de la parole. Une partie des tests ici réalisés est effectuée dans la base de données Aix-MARSEC, décrite dans le Chapitre 5. Cette base de données correspond à un projet du même nom, commun avec C. Auran et D. Hirst, dont le but est d'obtenir à partir du corpus MARSEC une base de données avec un maximum d'informations sous forme d'annotations, le tout étant obtenu automatiquement. La base de données est transcrite phonémiquement et orthographiquement et alignée automatiquement aux niveaux du phonème, de la syllabe, des sous-constituants syllabiques, des unités rythmiques (selon les deux modèles que nous voulons tester), des mots (mots seuls et mots avec annotation prosodique sous forme de marques tonétiques), et des unités intonatives mineures et majeures. Les données sont ensuite extraites sous la forme de données tabulaires pour être exploitées dans le cadre des deux problématiques de ce travail.

Dans le Chapitre 6, après avoir présenté et décrit les outils statistiques que nous jugeons les plus adaptés, nous comparons et testons dans la base de données Aix-MARSEC les deux modèles supra-syllabiques décrits dans le Chapitre 3 au travers de trois hypothèses : l'isochronie stricte, l'isochronie faible et l'effet du paramètre accentuel. Le but de ce chapitre est notamment de voir s'il existe un quelconque phénomène de compression au sein des unités rythmiques propres à chaque modèle. S'il existe une certaine compression des unités rythmiques comme le prédisent les deux modèles, il est intéressant de voir si un des deux modèles possède un degré de compression plus élevé. De plus, les deux modèles font la prédiction que les syllabes dans les unités rythmiques se répartissent sa durée globale équitablement. Ceci suppose

que l'accent n'influence pas la durée des syllabes/ phonèmes accentués, contrairement à l'acceptation générale de l'effet très marqué de l'accent sur la durée des phonèmes et des syllabes (Klatt 1987). Nous testons donc l'effet du paramètre accentuel dans les unités rythmiques pour déterminer son effet sur la durée des sous-constituants. Ces divers tests statistiques ont pour but de valider un des deux modèles comme représentant mieux les caractéristiques rythmiques de l'anglais britannique.

L'hypothèse d'ambisyllabité, plus globalement de syllabification, en anglais britannique est abordée dans le Chapitre 7 au travers tout d'abord de divers tests statistiques dans la base de données Aix-MARSEC. Le but de cette première phase est de déterminer s'il existe une différence significative entre la durée des consonnes intervocaliques regroupées en attaque de syllabe suivante (principe d'attaque maximale) et des consonnes intervocaliques ambisyllabiques, le tout sur de la parole authentique. Dans un deuxième temps, nous testons le phénomène d'ambisyllabité au travers d'une expérience en production dans laquelle nous incitons implicitement 14 sujets anglophones britanniques à découper des énoncés en syllabes. Le but est d'observer les stratégies de syllabification des locuteurs et de déterminer si l'ambisyllabité fait partie intégrante du système phonologique de l'anglais britannique.

Ce travail s'articule par conséquent en deux grandes parties, une première portant sur les aspects théoriques relatifs aux approches prédictives de l'organisation temporelle de l'anglais (Chapitre 1); nous aborderons ensuite la problématique de l'ambisyllabité (Chapitre 2), puis celle de la structuration rythmique de l'anglais et de la définition des unités rythmiques (Chapitre 3) et enfin nous définirons la notion d'isochronie (Chapitre 4). La deuxième partie concerne dans un premier temps la description de la base de données Aix-MARSEC utilisée dans cette étude (Chapitre 5), puis nous consacrerons les deux derniers chapitres à la présentation des différents résultats expérimentaux concernant la question de l'isochronie et de la constituance (Chapitre 6) puis la problématique de la syllabification et de l'ambisyllabité en anglais britannique (Chapitre 7).

PARTIE I.
ASPECTS THEORIQUES

CHAPITRE 1.
ORGANISATION TEMPORELLE ET MODELES DE PREDICTION

1.1. ORGANISATION TEMPORELLE	39
1.1.1. Caractéristiques générales	39
1.1.2. Problèmes soulevés par cette tâche prédictive	40
1.2. DEFINITION DE LA OU DES UNITE(S) TEMPORELLE(S)	42
1.3. TROIS UNITES, TROIS TYPES DE MODELES	45
1.3.1. "Modèles segmentaux"	45
1.3.1.1. Présentation	45
1.3.1.2. Arguments en faveur d'une approche segmentale	47
1.3.1.3. Limites de l'approche segmentale	50
1.3.2. "Modèles syllabiques"	54
1.3.2.1. Présentation	54
1.3.2.2. Problème : la syllabification en anglais	58
1.3.3. "Modèles rythmiques"	59
1.3.3.1. Présentation générale	59
1.3.3.2. Problème : définition de l'unité rythmique	59
1.4. CONTROLE DU DEBIT DE PAROLE	60
1.5. RESUME ET CONCLUSION	64

Ce premier chapitre est consacré à la situation du cadre général dans lequel s'inscrit ce travail. La composante temporelle joue un rôle fondamental dans la structuration prosodique de l'anglais et de nombreux travaux sont consacrés à la prédiction de l'organisation temporelle de l'anglais. Il existe plusieurs types de modèles prédictifs, la différence résidant dans le choix de l'unité prédite, principalement le phonème, la syllabe ou l'unité rythmique. Nous présenterons donc dans ce chapitre les trois types de modèle ce qui nous permettra d'introduire les différentes problématiques auxquelles nous nous intéressons dans ce travail, à savoir la syllabification en anglais, plus précisément le phénomène d'ambisyllabité, et la définition de l'unité rythmique.

1.1. Organisation temporelle

1.1.1. Caractéristiques générales

La problématique de la prédiction de l'organisation temporelle a fait l'objet de nombreuses études, particulièrement en anglais. Zellner-Keller (1996) souligne à ce sujet que la synthèse vocale manque encore de naturel du fait même que le paramètre temporel ne soit pas correctement prédit. Ceci ressort notamment des paroles synthétiques pour lesquelles le contour mélodique est manipulé pour correspondre au maximum avec le contour d'origine mais qui manquent de naturel (Campbell 1996). L'organisation temporelle a en effet longtemps été reléguée au second plan par de nombreux chercheurs au profit de la mélodie ; or, la composante temporelle est tout aussi importante dans un système de synthèse vocale que, par exemple, la mélodie.

La structuration temporelle des énoncés est fondamentale au niveau PRODUCTIF et PERCEPTIF (Klatt 1976a ; Eefting 1991 ; Nooteboom 1991). En effet, lorsqu'un locuteur produit un énoncé, c'est dans le but de communiquer son contenu, il va donc adopter un

certain nombre de stratégies pour atteindre ce but. La structuration temporelle des énoncés véhicule une signification précise et celle-ci va être interprétée par le ou les interlocuteur(s). En conséquence, la structure temporelle est un ingrédient essentiel à la bonne compréhension des énoncés. Ainsi, même si les variations temporelles sont nombreuses et fréquentes, elles ne sont pas toutes aléatoires et permettent un accès au sens de l'énoncé en question. Les variations de la structure temporelle sont perçues très clairement par les sujets ; de même que des structures temporelles erronées seront nettement perçues comme étant non naturelles (Eefting 1991). Une bonne prédiction de l'organisation temporelle est donc essentielle dans le cadre des systèmes de production de parole synthétique (comme par exemple les modèles de type *text-to-speech*), mais également des systèmes de reconnaissance automatique.

Si l'importance de la composante temporelle est indéniable, elle reste un paramètre difficile à prédire notamment parce qu'elle résulte de l'interaction de nombreux facteurs de nature diverse.

1.1.2. Problèmes soulevés par cette tâche prédictive

La difficulté majeure à laquelle on se retrouve confronté dès lors que l'on travaille sur l'organisation temporelle de la parole est sa variabilité, ceci quelle que soit la nature des phénomènes temporels en jeu. Une même voyelle dans des contextes différents aura une durée qui pourra varier de 0 à plusieurs centaines de millisecondes (Nootboom 1991 : 228). Une telle variabilité peut en partie être prédite grâce aux effets d'un certain nombre de facteurs ainsi qu'à leurs interactions, mais le tout reste délicat.

Il reste toujours en effet une part d'imprévisible. La durée des segments peut être prédite, mais en partie seulement. La première difficulté, soulevée notamment par Nootboom, réside dans la diversité des facteurs influençant l'organisation temporelle. En effet, ceux-ci ne se situent pas tous au même niveau prosodique de représentation et sont caractérisés par une importante interaction entre eux. En conséquence, il est difficile de les répertorier et d'implémenter leur interaction. Cette question d'interaction de facteurs est en soi problématique puisque les divers facteurs influant sur l'organisation temporelle ne doivent pas être considérés de manière isolée mais bien en fonction des autres facteurs (Nootboom). De plus, certains facteurs peuvent s'avérer

peu influents lorsqu'ils sont considérés individuellement, mais devenir très influents lorsqu'ils sont combinés avec d'autres facteurs.

La conséquence directe de cette interaction de facteurs est la nécessité absolue de robustesse et de capacité des outils utilisés, notamment statistiques ; ceux-ci doivent être suffisamment puissants pour pouvoir prendre en compte l'interaction de plusieurs facteurs. De plus, se pose la question des données sur lesquelles on teste les hypothèses puisqu'elles doivent se prêter à ces observations dans le sens où elles doivent être suffisamment denses pour permettre l'observation de l'interaction des nombreux facteurs.

Le deuxième type de difficulté est celui des variations dues aux locuteurs (attitude, affect particulier, etc.), difficiles, voire impossibles, à contrôler (Carlson 1991). Un énoncé identique produit par des locuteurs différents aura des patrons temporels différents (Campbell 2000). Cette part de variabilité très importante s'exprime d'un locuteur à un autre pour des raisons très diverses telles que l'âge, le sexe, le dialecte, l'appartenance sociale, etc., mais elle peut également ressortir du discours d'un même locuteur selon la situation de communication ; ainsi, la structuration temporelle d'un même énoncé produit par un même locuteur varie d'une réalisation à une autre. La variabilité intra et inter-locuteur joue par conséquent un rôle important dans la structuration d'un énoncé.

D'autres facteurs extérieurs sont également à prendre en considération comme le style de parole (parole lue, parole naturelle, etc.) ; nous reviendrons dans le Chapitre 5 sur la question des divers styles de parole et de leurs enjeux dans le cadre de l'observation et de la prédiction de l'organisation temporelle. Le contexte de communication va également jouer un rôle important : en situation de bruits extérieurs, ou dans le cas d'une discussion entre un natif et un non-natif par exemple, le locuteur maximise ses messages en surarticulant et en parlant plus lentement. On va donc retrouver un style hyperarticulé lorsque les informations contextuelles sont minimales et que le locuteur veut être bien compris. Ces différents styles se reflètent directement sur l'organisation temporelle.

Il existe par conséquent une limite évidente à la prédiction de l'organisation temporelle à partir d'un texte car persiste toujours une part de variations imprévisibles, principalement liées à l'attitude du locuteur dans sa situation de communication. Un facteur supplémentaire posant problème est celui du débit, lui-même étroitement lié à

celui de l'attitude du locuteur. On suppose généralement que le débit de parole est constant dans un énoncé donné alors que s'opèrent constamment des variations de débit au cours des énoncés (Carlson 1991). Le débit n'est pas nécessairement identique pour tous les mots ou morphèmes d'un énoncé, de la même manière qu'il varie selon la position syntaxique ; par exemple, le débit est ralenti lorsqu'on s'approche de la fin d'un énoncé, conséquence directe du relâchement progressif des articulateurs, ce qui provoque l'allongement des phonèmes et des syllabes en position finale d'énoncé. Nous reviendrons sur la question du débit dans la section 1.4.

Il ressort de ces diverses remarques que la structuration temporelle est un phénomène très complexe dépendant de nombreux paramètres, eux-mêmes étant de nature très variée. Toutefois, malgré cette part importante de variabilité, certaines variations sont régulières et donc prévisibles.

La question de la prédiction de l'organisation temporelle nous amène à la problématique fondamentale de la définition de l'unité sur laquelle fonder cette prédiction.

1.2. Définition de la ou des unité(s) temporelle(s)

Lorsque l'on cherche à prédire l'organisation temporelle de la parole, et ceci quelle que soit la langue étudiée, la première question qui se pose est celle de l'unité que l'on veut utiliser pour cette prédiction, autrement dit, à quel niveau on décide d'observer les phénomènes de variation temporelle. Les différents travaux prédictifs se distinguent en fonction de l'unité choisie.

En ce qui concerne l'anglais, le premier type d'unité est le phonème : on cherche à prédire la durée des phonèmes en fonction d'un certain nombre de paramètres (Klatt 1987 ; Van Santen 1992a). Le deuxième type d'études prédictives se fonde sur des unités de niveau supérieur telles que la syllabe (Campbell 1992a), le mot (Nooteboom 1991 ; Bell-Berti 1991) ou le pied métrique (Jassem 1952 ; Abercrombie 1964 ; Halliday 1967). Certains ont également proposé le groupe de centres inter-perceptifs (*inter-perceptual-center group*) correspondant à l'intervalle délimité par les centres perceptifs, à savoir les attaques vocaliques (Barbosa et Bailly 1994 ; 1997). Quelle que soit l'unité de niveau supérieur au phonème, la prédiction se fait généralement en deux étapes : une première prédiction de la durée de l'unité de rang supérieur (syllabe, groupe

de centres inter-perceptifs, mot ou pied) puis une répartition de la durée globale en fonction de la composition segmentale de cette unité, ou syllabique puis segmentale.

Cutler et Isard (1980 : 246) résument ces différents types d'approche et les regroupent en deux grandes catégories (nos soulignements) :

« Some writers, e.g. Klatt (1975) and Cooper (this volume), take the position that segmental durations such as vowel lengths are determined directly on the basis of syntax and the intrinsic nature of the phoneme concerned. Others, notably Abercrombie (1965), Halliday (1967) and Lehiste (1977) prefer to divide the utterance into "feet", stretching from one stressed syllable to the next. According to this scheme, the feet are assigned target lengths, which are shared out among the syllables composing the feet and then among the phonemes composing the syllables. »

Cette citation illustre les deux grands types d'approche, le premier se fondant sur la prédiction de la durée segmentale et le deuxième cherchant tout d'abord à prédire la durée des pieds (ou autre unité de niveau supérieur) pour ensuite en déduire la durée des syllabes puis des segments qui les composent. La première approche peut être qualifiée d'ASCENDANTE (*bottom-up*) dans le sens où elle se fonde sur la plus petite unité temporelle (le segment ou phonème), à partir de laquelle on peut déterminer la durée des unités supérieures par regroupement. En revanche, la deuxième approche est plutôt DESCENDANTE (*top-down*) car elle prédit l'organisation temporelle au travers des contraintes rythmiques pour ensuite déduire l'organisation des unités temporelles de niveaux inférieurs.

La citation de Cutler et Isard rejoint la distinction faite par Port, Al-ani et Maeda (1980 : 235-6) entre le niveau de la MICROSTRUCTURE TEMPORELLE, représenté par les segments phonologiques, et le niveau de la MACROSTRUCTURE TEMPORELLE, représentée par les unités plus larges que le segment phonétique. Dans cette dernière, on retrouve par exemple le concept d'isochronie ainsi que les oppositions entre langues à rythme syllabique et langues à rythme accentuel (nous reviendrons sur ces concepts dans le Chapitre 4).

Cependant, cette opposition stricte entre les niveaux de la micro et de la macrostructure temporelle peut être restrictive. En effet, il est difficile d'imaginer un modèle absolument restreint à l'un ou l'autre de ces deux niveaux de représentation. Il est plus raisonnable de considérer que les effets localisés à différents niveaux de la hiérarchie prosodique vont interagir et se combiner pour déterminer les différents

patrons de durée. L'organisation hiérarchique classique de l'anglais s'organise autour du phonème, de la syllabe, du pied accentuel et de l'unité intonative (Hirst 1983). Chacun de ces niveaux prosodiques de regroupement possède des caractéristiques temporelles spécifiques, sous forme notamment d'effets de frontière se manifestant par des effets d'allongements et de raccourcissements. Par conséquent, la structuration temporelle de la parole est un phénomène qui ne se confine pas à un seul niveau de représentation mais intervient simultanément aux niveaux des phonèmes, des syllabes, des pieds et des unités intonatives.

Le problème de la définition de l'unité temporelle reste donc intact ; la citation de Klatt (1987 : 760) ci-dessous témoigne de cette problématique (nos soulignements) :

« In fact, one of the unsolved problems in the development of rule systems for speech timing is the size of the unit (segment, onset/rhyme, syllable, word) best employed to capture various timing phenomena. »

Remarquons que cette citation est tirée de l'ouvrage récapitulatif de ses travaux (1987), ce qui montre que la question de l'unité à adopter reste toujours en suspens.

Si le problème de la définition de cette unité n'est pas résolu, c'est peut-être parce qu'une seule unité ne peut rendre compte des phénomènes temporels en jeu dans la parole. Des caractéristiques temporelles apparaissent à chacun des niveaux que sont le phonème, l'attaque et la rime, la syllabe et le mot (selon Klatt). On s'éloigne alors de la distinction binaire entre micro et macrostructure temporelles pour se rapprocher d'une interdépendance entre les différents niveaux de structuration prosodique.

En fonction des citations de Cutler et Isard et de Klatt, nous avons regroupé les différents types d'approches prédictives en trois catégories : les modèles segmentaux, les modèles syllabiques et les modèles supra-syllabiques. Les trois catégories tiennent par conséquent leur nom de l'unité adoptée : le segment, la syllabe et les unités supra-syllabiques (mot ou unité rythmique).

1.3. Trois unités, trois types de modèles

1.3.1. "Modèles segmentaux"

1.3.1.1. Présentation

Les modèles segmentaux sont nombreux et ont comme but de prédire les durées segmentales en fonction d'un ensemble de facteurs : on a par exemple Klatt, Van Santen et Riley (1992) pour l'anglais américain, Emerard, Mortamet et Cozannet (1992) pour le français et Kaiki, Takeda et Sagisaka (1992) pour le japonais. Le modèle de Klatt et de son synthétiseur fut le point de départ de nombreuses recherches (entre autres Nooteboom 1991 ; Campbell 1992a ; Van Santen 1992). Il se fonde sur l'application de onze règles : tous les facteurs influençant la durée des segments sont représentés sous la forme de règles de raccourcissement ou d'allongement.

Deux notions primordiales ressortent de cette approche : celle de durée inhérente (initialement proposée par Peterson et Lehiste 1960, puis reprise par Lehiste 1970) et celle de durée minimum. La durée intrinsèque des noyaux vocaliques correspond à la durée moyenne d'un noyau donné mesurée dans des paires minimales différant de par le voisement de la consonne post-vocalique. La DUREE INHERENTE représente la durée à laquelle on pourrait s'attendre en contexte non marqué, avec un minimum de contraintes externes (Klatt 1973b), elle est le point de l'application des différentes règles. La DUREE MINIMALE découle du concept d'incompressibilité des segments introduit par Klatt (1973b) : une voyelle raccourcie par une règle est moins compressible face à des raccourcissements additionnels, les pourcentages de raccourcissement ne peuvent donc pas se cumuler. Cette incompressibilité est relative à la durée inhérente d'un segment phonétique et reflète un temps minimum d'exécution du programme articulatoire requis pour la production du phonème en question. Par conséquent, la durée minimale est une mesure de la compressibilité des phonèmes.

Le modèle de Klatt (1975) se décompose en deux étapes :

- Détermination de la durée inhérente de chaque segment
- Application d'un certain nombre de règles qui modifient cette durée inhérente en fonction des interactions segmentales et de l'influence des unités linguistiques de plus haut niveau. Les onze règles (1969 ; 1987 : 761) se présentent sous la forme d'un pourcentage d'allongement ou de raccourcissement, en partant du principe que la durée

des segments ne peut être inférieure à leur durée minimum (principe d'incompressibilité).

Globalement, le modèle de Klatt convertit une durée segmentale D_i en entrée en une durée segmentale D_o en sortie selon la formule suivante (1976 : 1216) :

$$D_o = k (D_i - D_{min}) + D_{min}$$

k est un terme constant qui se situe entre 0 et 1 ($0 < k < 1$) dans le cas des règles de raccourcissement et qui est supérieur à 1 ($k > 1$) pour les règles d'allongement. L'originalité de ce modèle est qu'il est à la fois additif (au niveau des paramètres spécifiques aux segments) et multiplicatif (paramètres contextuels représentés par le facteur k).

Ce modèle rend compte de 84% de la variation des durées segmentales du corpus de Klatt, avec un écart type de 17ms. Le naturel des énoncés synthétisés à partir des durées segmentales prédites est jugé similaire à celui d'un texte synthétisé à partir des durées naturelles.

Certains modèles segmentaux (Pitrelli et Zue 1989 ; Riley 1992) sont fondés sur des arbres de classification et régression (CART). Le but est de prédire la durée des segments en fonction d'un certain nombre de facteurs tels que les informations inhérentes aux segments, les effets sémantiques, psychologiques et physiques (Pitrelli et Zue) ou bien la position lexicale, le statut accentuel et le contexte phonétique (Riley). Une telle technique permet de prédire des variables dépendantes continues en fonction de variables catégoriques et ainsi de visualiser sous la forme d'un arbre de régression la hiérarchie de l'influence de chacun des facteurs en entrée, ainsi qu'une estimation de la durée des segments en fonction de l'interaction des différents facteurs. L'avantage de cette approche est qu'elle permet de modéliser la complexité des interactions des divers facteurs et de modéliser ces interactions en simultané.

Van Santen (1992a et b ; 1997 ; 1998) présente le modèle de sommes et produits qui découle du fait que les effets des divers facteurs et leurs interactions affectant la durée des segments peuvent être additifs et/ ou multiplicatifs. On parle d'interaction additive lorsque les effets d'un facteur mesurés en millisecondes sont modifiés par l'influence d'un autre facteur. Mais cette interaction peut être additive *et* multiplicative : une voyelle a tendance à être plus longue quand elle est suivie d'une consonne voisée, ce

phénomène se combine à celui de position finale de l'énoncé, l'allongement est donc amplifié. L'interaction est alors additive et multiplicative (1998 : 125).

La notion principale sur laquelle se fonde le modèle de Van Santen (1992a) est celle d'"indépendance de facteur unique" (*single-factor independence*) : Van Santen propose de considérer la direction des effets de chacun des facteurs influençant la durée des phonèmes. En effet, l'influence des facteurs est généralement homogène ; par exemple, l'effet de l'accent est toujours allongeant, la direction de son effet ne varie pas selon le contexte, seule la magnitude peut être modifiée. Van Santen adopte cette approche car, devant le caractère extrêmement variable des phénomènes temporels, il est important de récupérer les invariabilités fondamentales. Dans le cas d'une interaction irrégulière, l'espace de traits doit alors être divisé en catégories homogènes, elles-mêmes régulières pour être "capturées" par le modèle de sommes de produits. Ce concept d'"indépendance de facteur unique" est étendu à celui d'"indépendance jointe" (*joint independence*) qui signifie que le principe peut s'appliquer à deux facteurs : les effets de ces deux facteurs sont indépendants des autres facteurs.

En prenant en compte sept facteurs influençant la durée des phonèmes, le modèle de Van Santen permet de prédire 86% de la variabilité de la durée des voyelles américaines à partir du texte. Le but sous-jacent de ce modèle est d'améliorer la qualité des systèmes de synthèse vocale de type *text-to-speech*, mais certaines informations ne peuvent être prédites à partir du texte seulement. Notons que la même approche a été appliquée à l'allemand, permettant de prédire environ 80% de la variabilité des durées observées (Möbius et Van Santen 2000).

1.3.1.2. Arguments en faveur d'une approche segmentale

Les "modèles segmentaux" sont nombreux, notamment en anglais. Klatt (1975) émet l'hypothèse de l'influence du rythme sur l'organisation temporelle (comme c'est le cas dans les modèles "supra-syllabiques") pour aussitôt s'en détourner, ceci pour deux raisons :

- Il n'a jamais pu observer une telle influence dans ses données.
- Il explique qu'il serait difficile de maintenir de larges différences de durée inhérente entre les types de voyelle s'il existait vraiment une compensation rythmique significative.

Klatt souligne également que les durées segmentales sont influencées par l'intonation, mais reste à déterminer si c'est la réalisation des modulations de la fréquence fondamentale qui impose des allongements ou si au contraire, c'est l'organisation temporelle qui conditionne les patrons mélodiques.

Lors d'un congrès, Lehiste (1979) critique l'approche de Klatt en raison de l'absence d'influence de tout facteur rythmique dans son modèle prédictif. Face à cette critique, Klatt répond qu'aucun facteur explicitement rythmique n'est intégré dans son modèle, mais les règles de raccourcissement des groupes consonantiques et des mots polysyllabiques permettent de retrouver une organisation qui tend vers une certaine régularité rythmique ; Campbell (1989) ajoute à ces facteurs le raccourcissement des syllabes inaccentuées, l'allongement en fin de syntagme et la désaccentuation des verbes fréquents, tous trois contribuant à la représentation d'une certaine régularité. Carlson (1991) confirme que le modèle de Klatt intègre des paramètres liés au rythme et précise que la prédiction du modèle de Klatt est meilleure qu'une prédiction entièrement fondée sur l'isochronie qui correspond en anglais à la récurrence régulière des intervalles accentuels ; nous reviendrons sur cette notion dans le Chapitre 4.

Van Santen (1997) se fonde sur le modèle prédictif de Klatt et précise que la description conventionnelle de l'organisation temporelle doit se faire en terme de durée des segments phonétiques qui composent les énoncés. Toutefois, il explique que la prédiction au niveau du phonème pose le problème de l'interaction des nombreux facteurs qui influencent la durée des unités temporelles (quelles qu'elles soient). Il cite l'exemple des onze facteurs définis par Klatt, chacun d'eux possédant un ensemble de paramètres comme par exemple, le facteur d'accentuation syllabique avec deux paramètres : accentué ou inaccentué (1992a). Ainsi, le nombre de paramètres augmente linéairement en fonction du nombre de facteurs. En revanche, le nombre de cellules augmente exponentiellement ; par cellules, Van Santen fait référence au faisceau de données qui découle du croisement des différents paramètres de chacun des onze facteurs. Möbius et Van Santen (2000) attribuent à chaque phonème un vecteur de traits représentant les informations au niveau segmental (identité du phonème en question et des phonèmes adjacents), les informations prosodiques (accent syllabique, accent mélodique et structure du syntagme prosodique), la structure syntaxique ainsi que d'autres paramètres de position. Ces différents vecteurs, ajoutés aux différents

paramètres de ces vecteurs (comme les deux paramètres de l'accentuation) représentent un espace de traits énorme.

La conséquence directe est le problème de la rareté des données ou plutôt de la rareté de la combinaison des différents facteurs. La majorité des vecteurs observés ont une fréquence d'occurrence très faible. Möbius et Van Santen (2000) les décrivent comme appartenant à la classe des LNRE (*Large Number of Rare Events*), nombre important de cas rares, et ils estiment qu'au moins un vecteur unique apparaît dans chaque énoncé. Plus généralement, Van Santen estime que la probabilité qu'un énoncé de cinquante phonèmes pris au hasard contienne un vecteur qui s'applique à moins d'un segment sur un million est supérieure à 95%. Le rare devient donc fréquent.

Il convient par conséquent d'élaborer un système qui ne laisse pas de côté ces cas rares, un modèle de prédiction de la durée qui sache prédire la durée des vecteurs qui ne sont pas ou peu représentés dans les données d'apprentissage. Van Santen reproche d'ailleurs aux outils tels que les analyses de type CART (*Classification and Regression Tree*), utilisées notamment par Riley (1992), de ne pas rendre compte des cas rares.

Malgré cette difficulté, Van Santen critique et s'oppose à l'approche syllabique de l'organisation temporelle, cette critique se tournant principalement vers les études de Campbell (1992a et b) et de Barbosa et Bailly (1994). Le modèle de Campbell prédit la durée des syllabes en fonction de leur complexité segmentale (nombre de segments dans chaque syllabe), de plusieurs facteurs prosodiques et de la nature du noyau. Barbosa et Bailly ont développé un système similaire dont l'unité est le groupe de centres inter-perceptifs (plutôt que la syllabe) correspondant aux intervalles délimités par les attaques vocaliques. Selon Van Santen, ces deux modèles se fondent sur les deux suppositions suivantes :

- L'indépendance des segments : la durée syllabique est quasi indépendante des durées segmentales.
- La médiation syllabique : la durée segmentale dépend principalement de la durée syllabique et de l'identité segmentale.

Van Santen teste ces deux suppositions et montre dans un premier temps que la durée syllabique dépend considérablement des durées segmentales intrinsèques ; par conséquent, une prédiction précise de la durée syllabique suppose que l'on prenne en compte la composition segmentale totale de la syllabe et pas simplement le nombre de segments et la nature du noyau comme le fait par exemple Campbell. Deuxièmement,

Van Santen montre qu'il est impossible de prédire la durée segmentale à partir de la durée syllabique car les effets prosodiques n'influencent pas les différentes parties de la syllabe de manière uniforme. Par exemple, l'accent influence plutôt l'attaque et le noyau syllabique, alors que la position initiale dans l'unité influence l'attaque et la position finale influence la rime (Klatt 1976). Or, ne pas tenir compte de ces allongements spécifiques à certains constituants syllabiques peut engendrer des durées segmentales erronées. Il est toutefois important de préciser que ces allongements différentiels au niveau sub-syllabique sont intégrés par Campbell et Isard (1991) puis par Campbell (1992a).

Notons que même si le modèle de Van Santen cherche à prédire la durée des segments, il n'exclut en aucun cas l'influence des niveaux supérieurs. Les critères de position de la syllabe dans le mot et de position du mot dans l'énoncé sont intégrés au modèle, témoignant de la prise en compte des facteurs de niveaux supérieurs (Van Santen 1997 : 236¹). La prédiction de la durée segmentale se fait également par rapport au cadre plus large de la syllabe et du mot ; toutefois, selon Van Santen, si la syllabe joue un rôle important en production, il n'est pas souhaitable de l'utiliser comme unité de prédiction de l'organisation temporelle. Pour Van Santen, plus l'unité prédite est petite, mieux c'est (*the smaller, the better*). L'organisation temporelle ne se planifie pas à un niveau haut de représentation, mais à un niveau bas car la durée des segments est grandement influencée par l'organisation temporelle des unités supérieures. Toutefois, une prédiction correcte doit porter sur les segments.

1.3.1.3. Limites de l'approche segmentale

Malgré le point de vue de Van Santen, il est nécessaire de remarquer que travailler au niveau du phonème pose certains problèmes indéniables.

Premier problème, la durée des phonèmes est extrêmement variable, et donc difficile à prédire, notamment par rapport à la durée syllabique (Campbell 1992a). Klatt (1976a) reconnaît à ce sujet que la durée segmentale est très variable, il arrive même que certains segments durent jusqu'à huit fois plus que les mêmes segments dans des contextes différents. Ainsi, la tâche prédictive est rendue extrêmement complexe et,

¹ Van Santen (1997 : 236): « *there is agreement that for prediction of any temporal unit, varying phonological entities are needed (eg. phonemes, syllables, words)* »

dans le cadre de la synthèse vocale, une prédiction aussi fine que possible des durées segmentales n'est pas le critère primordial d'une parole synthétique naturelle (Campbell 1992b).

Si la durée des phonèmes est variable, c'est parce qu'elle résulte de l'interaction de nombreux facteurs. A partir du résumé de ces différents facteurs effectué par Campbell (1992a : 39), on peut citer neuf facteurs :

- Différences de durée intrinsèque entre les phonèmes, les voyelles étant généralement plus longues que les consonnes (Crystal et House 1988a), avec une distinction entre voyelles courtes et voyelles longues (Peterson et Lehiste 1960 ; House 1975 ; Van Santen 1992b), voire entre voyelles simples et voyelles complexes (Gimson, réédité par Cruttenden 1997). En ce qui concerne les consonnes, les fortis sont globalement plus longues que les lenis (Peterson et Lehiste 1960 ; Lehiste 1970), leur durée étant influencée par leur position dans le mot, notamment les plosives (Crystal et House 1988b).
- Influence du contexte phonémique : les voyelles sont généralement allongées lorsqu'elles sont suivies d'une consonne voisée plutôt que d'une consonne non-voisée (Peterson et Lehiste 1960 ; Chen 1970 ; House 1976 ; Luce et Charles-Luce 1985 ; Klatt 1987 ; Van Santen 1992b). On observe également un raccourcissement des consonnes dans les groupes consonantiques (Klatt 1973b ; Crystal et House 1988a ; Carlson et Granström 1989). De plus, on observe un allongement pour l'aspiration des plosives, sa durée dépendant du point d'articulation, avec un effet sur la durée de la voyelle qui suit (Allen et Miller 1999).
- Allongement dû à l'accent lexical (Lehiste 1970 ; Klatt 1974, 1987 ; Crystal et House 1988b, 1990 ; Fant, Kruckenberg et Nord 1991b ; Van Santen 1992b), sans différence significative entre accent primaire et accent secondaire (Van Santen 1992b).
- Informations inhérentes à la syllabe : la position (finale vs non-finale) de la syllabe dans les constituants de niveau supérieur (Van Santen 1992b) ainsi que la position des segments dans les différents constituants syllabiques influencent la durée des segments de manière significative (Campbell 1992a).
- Facteurs au niveau du mot : allongement des consonnes en position initiale de mot (Klatt 1987 ; Crystal et House 1988a ; Carlson et Granström 1989), allongement de la rime en fin de mot (Klatt 1976 ; Campbell 1992a, 200), raccourcissement

polysyllabique (Umeda 1975 ; Klatt 1987) et position du mot dans l'énoncé (Van Santen 1992b).

- Facteurs syntaxiques : allongement prépausal (Oller 1973 ; Klatt 1973a ; Umeda 1975 ; Cooper et Paccia-Cooper 1980 ; Cooper et Danly 1981 ; Crystal et House 1988a et b ; Carlson 1991 ; Cummins 1999) en proportion inverse avec l'éloignement de la frontière (Fant, Kruckenberg et Nord 1991b), allongement en fin de syntagme (entre autres Klatt 1975 ; Umeda 1975 ; Cooper et Paccia-Cooper 1980 ; Cooper et Danly 1981 ; Nakatani, O'Connor et Aston 1981), raccourcissement du premier élément d'un mot composé et allongement des mots lexicaux par rapport aux mots fonctionnels (Umeda 1975 ; Van Santen 1992b ; Campbell 2000).
- Facteur sémantique : allongement des segments sous l'effet de l'accent contrastif ou emphatique (Klatt 1987 ; Van Santen 1992b).
- Débit de parole : influence considérable du débit sur les durées segmentales (Lehiste 1970 ; Crystal et House 1982, 1988b, 1990). Nous reviendrons sur la question du débit dans la section 1.4 de ce chapitre.
- Locuteur : les caractéristiques inhérentes aux locuteurs représentent un paramètre important dans l'organisation temporelle de la parole (Carlson 1991), ces caractéristiques étant souvent liées au débit.

Il ressort de manière évidente que les facteurs influençant la durée des phonèmes sont nombreux et de nature très variée. La conséquence directe est la difficulté de rendre compte de l'influence de tous ces paramètres et de leurs interactions.

La prédiction au niveau du phonème pose un problème supplémentaire ; les critères de segmentation du signal en phonèmes ne sont pas clairement définis (Peterson et Lehiste 1960 ; Campbell 2000) et dépendent en grande partie de la personne qui effectue l'alignement. La segmentation permet l'alignement temporel entre la transcription phonétique et le contenu du signal sonore, elle est nécessaire pour recueillir les durées segmentales d'un signal donné. Les critères diffèrent d'une personne à l'autre et l'intuition de chacun influence grandement son alignement ; de plus, une même personne n'alignera pas un même signal de manière absolument identique à deux reprises. En résulte une très grande variation et la conséquence directe de cette hétérogénéité est la difficulté de comparaison des différentes études au niveau segmental. C'est en partie devant ce problème de segmentation que Campbell (1996) propose d'utiliser la syllabe plutôt que le phonème ; même si la définition de la syllabe

fait l'objet de nombreux débats (cf. chapitre suivant), ses frontières sont plus faciles à déterminer dans le sens où elles concernent la plupart du temps des consonnes aisément distinguables dans le signal sonore (1996 : 215).

Collier (1992) soulève deux problèmes supplémentaires liés aux approches segmentales. Premièrement, il remarque qu'il est surprenant et contradictoire de parler de "suprasegmentalité" lorsqu'on s'intéresse à la prosodie d'une langue tout en considérant le niveau segmental. Il conviendrait de considérer la temporalité à un niveau supérieur à celui du segment, à un niveau proprement suprasegmental comme par exemple celui de la syllabe, du mot ou du pied notamment en raison de l'organisation des gestes articulatoires². Deuxièmement, prédire l'organisation temporelle au niveau du segment présente le risque d'accumulation d'erreurs de prédiction des durées segmentales³. Ces erreurs, aussi petites soient-elles, s'accumulent dans l'énoncé (ou quelle que soit l'unité supérieure) et perturbent l'organisation temporelle de l'unité en question et, par voie de conséquence, son rythme. L'accumulation d'erreurs de l'ordre de quelques millisecondes va bouleverser l'organisation temporelle de l'unité supérieure, et si on se place dans l'optique de la synthèse vocale, le résultat synthétique perdra son naturel.

Selon Collier, prédire la durée de la syllabe ou du mot permet d'éviter ce risque : s'il existe une erreur de prédiction de la durée segmentale, elle se confinerà à la syllabe à laquelle le segment appartient, mais ne "contaminera" pas l'organisation temporelle de la totalité de l'unité de niveau supérieur. Collier remarque enfin qu'une fois la structure temporelle de l'unité supérieure correctement prédite, les erreurs au niveau segmental se compensent entre elles mais ne s'accumulent pas. Grâce à divers algorithmes comme celui de Campbell (1992a), une fois la durée syllabique déterminée, on calcule la durée approximative des segments selon les principes d'ajustement et d'élasticité (que nous

² Collier (1992 : 206) : « *Indeed, it is not without reason that prosody is referred to as 'supra-segmental' level. (...) Yet it seems intuitively more plausible that the speaker organises his articulatory behaviour in syllable-, word-, or foot-sized chunks and programs his articulatory timing at the supra-segmental level. This alternative view implies that global temporal structures determine to a large extent the duration of the individual segments.* »

³ Collier (1992 : 207) : « *[Campbell's] approach potentially has a major advantage, namely that it reduces the risk of an accumulation of prediction errors. Indeed, as long as the temporal structure of the higher order unit is predicted correctly, any errors at the lower segmental level will compensate for each other rather than accumulate: an overshoot in the timing of one phoneme will be cancelled by an undershoot of another phoneme within the same rhythmical unit.* »

définirons dans la section 1.3.2.1). Si la durée de ces unités est correctement prédite, les durées segmentales s'adapteront.

Un autre problème lié aux approches segmentales est celui du coût de l'alignement manuel ; on estime qu'il faut environ seize heures pour transcrire phonétiquement et aligner soigneusement un signal de parole d'une minute avec sa composition segmentale (Di Cristo (P.) et Hirst 1997). Il est donc très coûteux de faire l'alignement manuel d'un corpus de plusieurs heures de parole. Devant l'importance de la tâche, de nombreux systèmes d'alignement automatique se sont développés, facilitant ainsi l'alignement de la transcription phonétique avec le signal sonore. Cependant, à ce jour, aucun aligneur automatique efficace entre le signal sonore et les segments réalisés n'est aisément disponible et directement utilisable par un phonéticien sans formation informatique. Les approches segmentales soulèvent donc des problèmes techniques supplémentaires.

Notons enfin les modèles non-segmentaux développés d'une part par Coleman (1992) et Dirksen et Coleman (1997) et d'autre part par Local et Ogden (1997), dont la structure hiérarchique est constituée des niveaux suivants : les traits (comme par exemple le poids, la hauteur, etc.), les entités sub-syllabiques (attaque, noyau, coda et rime), la syllabe et le pied. Il est intéressant de voir que les phonèmes sont absents de cette hiérarchie. Möbius et Van Santen (2000) rapportent l'efficacité du premier de ces systèmes, tout particulièrement dans la représentation des effets coarticulatoires tels que les assimilations, les suppressions, les insertions, les substitutions et les réductions.

1.3.2. "Modèles syllabiques"

1.3.2.1. Présentation

Devant les difficultés que posent les approches segmentales, certaines approches ont pour objet d'étude une unité plus large que le phonème, comme par exemple la syllabe. On cherche en effet à prédire les phénomènes temporels à un niveau supérieur à celui du segment afin de mieux saisir l'organisation suprasegmentale.

Le principal représentant de ce type de modèle est celui élaboré par Campbell et Isard (1991) et Campbell (1992a et b ; 2000). Celui-ci fonctionne sur plusieurs niveaux, il est d'ailleurs décrit comme un modèle temporel à niveaux multiples (*multi-level timing model*). L'originalité de ce modèle est qu'il permet de calculer dans un premier

temps la durée des syllabes, pour ensuite prédire la durée des segments contenus dans ces syllabes. Il fonctionne donc avec un cadre de plus haut niveau et ne cherche pas à prédire directement la durée des segments. Campbell (2000) explique que cette approche permet de rendre compte de l'interaction entre les niveaux temporels de haut et bas niveau, la syllabe servant en fait de "médiateur" entre ces niveaux.

Le même auteur (1989) explique que le modèle de Klatt est tout à fait acceptable mais, alors qu'il permet la prédiction de 84% de la variabilité des durées segmentales en anglais américain, une fois appliqué à de l'anglais britannique, Campbell obtient une prédiction de seulement 71% de la variabilité des durées segmentales. Campbell propose alors de réécrire les règles de Klatt mais en les appliquant au niveau de la syllabe plutôt que du segment ; plus exactement, il localise les effets allongeants ou raccourcissants au sein des constituants syllabiques.

Le niveau premier de prédiction n'est donc plus le segment mais la syllabe, ce niveau permettant la prise en compte des facteurs d'accentuation, de rythme et de position. Campbell (1992a) insiste sur l'importance du facteur rythmique : même si les modifications engendrées sont parfois de faible importance, elles contribuent fortement au caractère naturel de la parole synthétique résultante. En anglais, la différence entre la durée des syllabes accentuées et des syllabes inaccentuées est très marquée (Klatt 1979) et des effets notables de compensation s'opèrent au niveau du pied. Il est difficile d'intégrer ces facteurs rythmiques dans un modèle segmental ; en revanche, ils peuvent être appliqués au niveau syllabique.

Dans le modèle de Campbell (2000 : 317), la durée des syllabes est tout d'abord prédite à partir d'informations au niveau segmental (complexité de la syllabe et nature du noyau vocalique) et à un niveau de représentation supérieure (position, nature du pied et statut accentuel, nature du mot). Le modèle permet de rendre compte de 85% de la variabilité des durées syllabiques (1992b).

Une fois les durées syllabiques calculées, les durées segmentales sont déterminées en fonction du principe d'ajustement (*accommodation*) dans le cadre de la syllabe grâce au principe d'élasticité. Les notions d'élasticité et d'ajustement sont centrales au modèle de Campbell ; la première fait référence à un degré de compression ou d'extension des

segments⁴ et la deuxième à la modification de ce degré pour s'ajuster à la durée du cadre syllabique⁵.

L'hypothèse d'élasticité suppose que dans une syllabe donnée, tous les segments sont compressés ou allongés uniformément. Cette définition correspond à une forme extrême du principe car, dans la réalité, Campbell observe que les allongements et compressions diffèrent selon la position de la syllabe en question dans l'énoncé (position finale ou non-finale), selon la position du segment dans la structure syllabique (attaque, noyau ou coda) et selon le contexte phonétique (effet du voisement de la consonne sur la voyelle précédente).

Ces trois effets sont testés par Campbell et Isard (1991) et Campbell (1992a). Campbell observe que l'allongement final porte sur la rime mais pas sur l'attaque et que l'allongement est de plus en plus marqué plus on s'approche de la fin de l'énoncé (effet de frontière). En effet, l'allongement ou la compression n'affecte pas les mêmes constituants syllabiques selon la position dans l'énoncé et la longueur moyenne de la syllabe en question. De plus, l'allongement d'une voyelle sera différent selon le caractère voisé ou non-voisé de la consonne qui suit : sa durée s'ajuste en fonction du contexte consonantique.

La Figure 1 ci-après représente l'illustration que donne Campbell des différents types d'allongements ou de raccourcissements au sein de la syllabe et des constituants syllabiques. La forme triangulaire représente le cadre syllabique et les différentes étendues d'allongement.

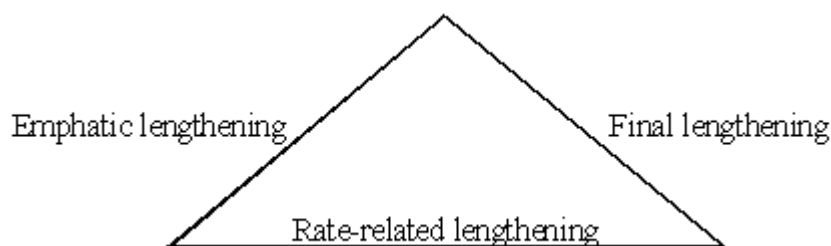


Figure 1. Représentation des différents types d'allongement au sein de la syllabe (Campbell 1992a: 123)

⁴ Campbell (1992a : 80): « a notional 'stretching' or 'squashing' of segment duration ».

⁵ Campbell (1992a : 80): « the process of 'stretching' and 'squashing' in order to fit a syllable-sized frame ».

Les trois types d'allongement ne portent pas sur les mêmes constituants syllabiques : l'allongement accentuel s'opère au niveau de l'attaque et du noyau, l'allongement final porte sur la rime et l'allongement lié au débit influe sur la totalité de la syllabe. Ces allongements différentiels permettent à Campbell (1993) de mettre au point un algorithme de détection automatique des frontières prosodiques qui marquent les bords de syntagmes intonatifs. Ainsi, l'allongement final porte sur la rime uniquement, contrairement à l'allongement accentuel qui porte sur l'attaque et le noyau, tout allongement sur la rime est par conséquent interprété comme un indice de frontière.

Parmi les modèles syllabiques, on peut également citer les travaux de Barbosa et Bailly (1994 ; 1997) sur le français, bien qu'ils utilisent les intervalles délimités par les attaques vocaliques plutôt que les syllabes. Le mode de fonctionnement du modèle de synthèse de Barbosa et Bailly est identique à celui du modèle de Campbell dans le sens où la prédiction s'effectue en deux étapes : prédiction de la durée des intervalles inter-centres perceptifs en fonction d'informations d'ordre phonotactique et syntaxique, puis calcul par répartition de la durée des segments compris dans ces intervalles (1994). Le modèle est ensuite enrichi (1997) par le fait qu'il rend compte de l'apparition de pauses lorsque le débit ralentit.

Comme nous l'avons souligné plus tôt, l'avantage de cette approche qualifiée de syllabique est de tenter de contourner la difficulté que représente la prédiction de la durée des segments et d'éviter les accumulations d'erreurs de prédiction de la durée segmentale (Collier 1992). Dans la mesure où les durées syllabiques (ou durées d'intervalles inter-attaques vocaliques) sont correctement prédites, les erreurs de durées segmentales se compensent au lieu de s'accumuler.

Toutefois, si Campbell fonde sa prédiction sur la syllabe, il reconnaît qu'il est nécessaire de valider ce niveau et de voir s'il est suffisant pour rendre compte de toutes les interactions mises en jeu dans la détermination de l'organisation temporelle⁶. Il soulève également la possibilité que la syllabe permette une représentation limitée de la variabilité des phénomènes temporels (1996 : 232)⁷.

⁶ Campbell (1996: 232): « *It still remains to be shown that the syllable is a sufficient unit to measure, and that all durational rules are capable of being expressed in terms of the syllable with its content and context* ».

⁷ Campbell (1996: 232): « *the syllable too may be a limitation when it comes to accounting for all the variance observed in the durations* ».

Campbell (1996) remarque que la syllabe est également caractérisée par une durée variable ; cette variabilité peut être due à la différence de complexité qui existe entre les syllabes. En effet, il existe des syllabes mono-phonémiques (comme l'article indéfini *a*) mais également des syllabes composées de bien plus de phonèmes (comme dans le mot monosyllabique *trapped* cité par Campbell). Outre la différence de complexité, la durée des syllabes varie en fonction de facteurs de débit, d'accentuation et de position dans le mot, le pied et/ou le syntagme (1996 : 221). Toutefois, cette variabilité relève de l'influence de facteurs concrets alors qu'il n'en va pas nécessairement de même dans le cas de la durée des phonèmes.

1.3.2.2. Problème : la syllabification en anglais

Si l'approche syllabique permet de contourner les principaux problèmes posés par une approche segmentale, elle présente également des difficultés majeures.

Il convient de souligner que travailler au niveau de la syllabe pose le problème de segmentation de la chaîne parlée en syllabes. La syllabe joue un rôle incontestable dans la structuration prosodique, mais les critères de segmentation de la chaîne segmentale en syllabes font toujours l'objet de nombreux débats. Nous avons précédemment soulevé le problème de segmentation de la chaîne parlée en phonèmes au niveau acoustique. Il convient ici de soulever le problème de la définition de la syllabe et donc de ses frontières.

En effet, la syllabe se trouve au cœur de nombreux débats et si on prend l'exemple d'un mot polysyllabique avec des groupes consonantiques intervocaliques plus ou moins complexes, il ne sera pas toujours évident de dire où commencent et où s'arrêtent les syllabes qui composent ce mot. La question des frontières syllabiques revient à poser la question du regroupement des consonnes intervocaliques.

Nous consacrerons le deuxième chapitre de cette première partie consacrée aux aspects théoriques à la problématique de syllabification en anglais qui nous semble être au cœur de la structuration prosodique en anglais.

1.3.3. "Modèles supra-syllabiques"

1.3.3.1. Présentation générale

Nous entendons par "modèle supra-syllabique" les approches qui cherchent à prédire l'organisation temporelle en fonction de la structure rythmique de l'énoncé dont il est question (Jassem 1952 ; Abercrombie 1967 ; Halliday 1967).

Le but premier est de prédire la durée des unités rythmiques pour ensuite déterminer la durée des syllabes et des phonèmes qui les composent. La durée de ces unités rythmiques est régie par le principe d'isochronie, ce qui signifie que quel que soit le nombre de syllabes qui les composent, leur durée est quasi similaire. Cette notion d'isochronie implique des facteurs de compression ou d'allongement au niveau des syllabes pour qu'elles s'adaptent à la durée quasi régulière de ces unités rythmiques.

En somme, l'organisation temporelle de la parole se situe à plusieurs niveaux : une première approximation des répartitions s'effectue au niveau de l'unité rythmique, on détermine alors la durée de ces unités accentuelles en fonction d'un certain nombre de facteurs. Une fois les durées d'unités déterminées, on cherche à prédire la durée des syllabes qui les composent : les syllabes sont alors compressées ou allongées pour correspondre au cadre de l'unité rythmique.

Cette approche correspond typiquement à ce que nous avons qualifié d'approche descendante (*top-down*) ; on prédit tout d'abord la durée des unités supérieures (au niveau de l'organisation rythmique) pour ensuite déduire la durée des syllabes qui les composent. L'avantage de ce type d'approche est la prise en compte des paramètres rythmiques dans la prédiction de l'organisation temporelle, chose plus délicate à réaliser lorsque la prédiction est ascendante.

1.3.3.2. Problème : définition de l'unité rythmique

L'approche supra-syllabique présente des avantages évidents puisqu'elle permet de prendre en compte divers facteurs purement rythmiques, contrairement à l'approche segmentale par exemple. Toutefois, la question qui se pose ici est la définition de l'unité rythmique.

En effet, l'unité rythmique classique généralement adoptée pour l'anglais est le pied accentuel (Abercrombie 1967 ; Halliday 1967) qui consiste en une syllabe accentuée et toutes les syllabes inaccentuées suivantes. La durée des pieds est supposée être

relativement stable quel que soit le nombre de syllabes (isochronie faible). Cette unité rythmique a depuis été largement utilisée sans pour autant être réellement remise en cause. Or, plus de dix ans auparavant, Jassem (1952) avait proposé un autre modèle supra-syllabique pour l'anglais, passé assez inaperçu.

Le modèle de Jassem consiste en une unité rythmique étroite qui débute avec une syllabe accentuée et englobe toutes les syllabes inaccentuées jusqu'à la fin du mot en cours ; les autres syllabes inaccentuées appartiennent à l'anacrouse. Il se démarque ici de la définition généralement attribuée à l'anacrouse, à savoir les syllabes inaccentuées en début d'unité intonative. La durée des unités rythmiques étroites a tendance à être constante quel que soit le nombre de syllabes qu'elles contiennent ; en revanche, la durée de l'anacrouse est directement proportionnelle au nombre de syllabes, celles-ci étant les plus courtes possible.

En conséquence, on se retrouve avec deux modèles supra-syllabiques de l'anglais fondés sur deux unités rythmiques distinctes. Il est évident que les prédictions de l'organisation temporelle découlant de ces deux modèles donneraient des résultats très différents. Cette problématique sera traitée au cours des Chapitres 3 et 4 de cette première partie. Le Chapitre 3 est consacré à la description détaillée des deux modèles supra-syllabiques et à l'évaluation de leur efficacité relative vis-à-vis de ce qui a été présenté dans la littérature à ce sujet. Dans le Chapitre 4, nous abordons la notion d'isochronie et les concepts et travaux qui en découlent.

1.4. Contrôle du débit de parole

Il est difficile de traiter la question de la prédiction de l'organisation temporelle d'une langue sans faire référence à l'influence du débit de parole. En effet, le débit de parole des locuteurs est un paramètre très important dans la prédiction de la durée des segments (ainsi que de toute autre unité), mais également le plus difficile à contrôler ; les différents effets cités dans la littérature ne concordent pas toujours du fait même de la complexité du phénomène.

Lorsqu'on fait référence au débit, on peut distinguer d'une part la vitesse d'élocution d'un énoncé unique et d'autre part celle de l'ensemble des énoncés d'un tour de parole (Laver 1994 : 539). La première constitue la vitesse d'articulation (*articulation rate*) et la deuxième la vitesse de parole (*speaking rate*), les deux n'étant

pas nécessairement équivalentes notamment en anglais. Plus précisément, la vitesse de parole est définie comme étant le nombre de syllabes divisé par le temps de locution (incluant les pauses) et la vitesse d'articulation le nombre de syllabes divisé par le temps d'élocution sans les pauses (Zellner 1998 : 29). Pfitzinger (1996 : 421) opère la même distinction entre le débit local et le débit global ; le premier concerne les différents constituants caractérisés par des variations de débit et le deuxième concerne le débit au niveau de l'énoncé entier.

Le débit de parole est certainement le paramètre le plus difficile à contrôler dans le cadre d'un modèle de prédiction de l'organisation temporelle. Or, il est fondamental de tenir compte du débit dans un tel modèle car il influe grandement sur l'organisation temporelle des énoncés. De plus, un système de synthèse se doit d'intégrer un traitement spécifique du débit pour aboutir à un résultat naturel. En effet, comme l'explique Eefting (1991), le débit doit être considéré à la fois sur le plan de la production et de la perception. Les variations du débit sont nombreuses et fréquentes mais non aléatoires si les locuteurs veulent être compris rapidement et correctement par leurs interlocuteurs (principe d'adaptation).

Les raisons qui expliquent la difficulté à contrôler le paramètre du débit sont multiples. Tout d'abord, les variations de débit entraînent des effets non-linéaires. En effet, ces variations n'affectent pas la totalité des segments et/ ou des syllabes de manière égale. Une augmentation du débit affecte plus les segments longs que les segments courts (Crystal et House 1988a), elle affecte plus une voyelle qu'une consonne et elle affecte plus une syllabe accentuée qu'une syllabe inaccentuée (Lehiste 1970). De plus, la durée relative des consonnes diminue lorsque le débit accélère, contrairement à l'hypothèse selon laquelle, lorsque le débit augmente, les voyelles sont élidées parce que le cumul des durées minimales des consonnes occupe toute la durée de la syllabe à laquelle elles appartiennent (Wood 1973). Les voyelles sont globalement plus longues avec un débit lent mais pas les consonnes (Nooteboom 1991 ; Carlson 1991). Par conséquent, les facteurs d'allongement et de raccourcissement varient d'un élément à un autre selon ses caractéristiques intrinsèques et son statut accentuel.

Les variations de débit sont également qualifiées de non-linéaires car elles sont brusques et nombreuses : les accélérations et décélérations sont imprévisibles dans le sens où elles peuvent porter sur n'importe quel élément et intervenir à n'importe quel moment de l'énoncé. Elles peuvent être fréquentes ou rares, elles peuvent être plus ou

moins brusques. Elles varient d'un locuteur à un autre, d'une langue à une autre, et même lorsqu'il s'agit d'une même langue, d'un même locuteur, elles sont totalement aléatoires. Toutefois, les variations de débit ne semblent pas être totalement libres si les locuteurs veulent être compris (Eefting 1991).

Les variations de débit ne sont pas unidimensionnelles et engendrent des modifications à la fois quantitatives ET qualitatives et elles interviennent à différents niveaux de structuration de la parole. Zellner (1998 : 31-33) explique que les effets au niveau de la macrostructure temporelle sont faibles puisque les regroupements ne sont pas affectés par les variations de débit ; en revanche, elles affectent les procédés de syllabation puisqu'un débit rapide engendre la suppression de certains phonèmes et donc la réorganisation des syllabes en phonèmes. Elles affectent également les constituants syllabiques car les variations de débit n'influencent pas la durée des constituants syllabiques de manière absolument homogène. Enfin, une accélération du débit affecte les traits acoustiques des consonnes ainsi que les zones de transition entre segments. Ces variations se reflètent par conséquent à tous les niveaux de structuration, d'où la difficulté de les implémenter.

On trouve fréquemment dans la littérature que c'est l'accélération du débit qui engendre des élisions de phonèmes : le temps d'exécution diminue, entraînant la réduction, voire la suppression de certains phonèmes. Toutefois, dans un article de 1973, Wood explique que les élisions et les contractions sont habituelles et régulières ; la preuve en est qu'on les retrouve également lorsque le débit est lent. Ainsi, elles ne sont pas causées par l'accélération du débit, c'est même plutôt leurs occurrences qui permettent d'accélérer le débit puisque le temps d'exécution des gestes articulatoires est diminué.

Les variations de débit influencent également les occurrences de pauses ; les pauses sont plus longues et fréquentes en débit lent, ceci étant lié au degré d'importance linguistique de chacune des pauses en question (Grosjean et Collins 1979 ; Crystal et House 1982). En cas de débit rapide, les pauses de respiration sont plus courtes et moins fréquentes.

La difficulté que pose le contrôle des variations de débit dans la parole relève également du paramètre extralinguistique qui le caractérise. Ces variations peuvent par exemple être dues à un changement d'humeur de la part du locuteur. Dans son édition des travaux de Gimson, Ramsaran (1989) attribue à un débit rapide une idée d'irritation

et d'insistance, alors qu'en cas de débit lent, il s'agit plutôt d'hésitation, de doute ou d'ennui avec les formes affirmatives et de compassion et d'encouragement avec les formes interrogatives et impératives. Il est évident que de telles variations au niveau extralinguistique sont (quasi) impossibles à contrôler et donc à implémenter.

Zellner (1998 : 94) observe qu'en français les effets de la variation du débit sont de trois ordres : ajout/suppression de pauses (8% des cas), recomposition de la chaîne phonémique sous la forme d'ajout ou de suppression de phonème (9% des cas) et variations de la durée des unités sonores avec un effet plus marqué pour les syllabes (83% des cas). On voit donc ici que les variations du débit ont principalement des répercussions sur la structuration temporelle des énoncés produits.

Dans les différentes études portant sur le débit, celui-ci se mesure souvent en terme de nombre de syllabes par seconde, voire de mots par seconde selon l'orientation de l'étude. Toutefois, cette mesure a été largement critiquée dans le sens où la complexité syllabique peut considérablement différer d'une syllabe à l'autre, notamment en anglais (*a* vs *strapped*). Par exemple, Crystal et House (1990) et Eefting (1991) proposent de plutôt utiliser l'inverse du nombre de syllabes par seconde c'est à dire la durée syllabique moyenne à l'intérieur d'une séquence excluant toute pause (*ASD*, *Average Syllable Duration*) en tant que mesure de la vitesse d'élocution. On pourrait également suggérer de représenter le débit en terme de nombre de segments par seconde, mais sa durée étant extrêmement variable, il est difficile d'appliquer une telle méthode. De plus, elle pose exactement le même problème qu'avec le nombre de syllabes par seconde puisqu'elle nous amène à comparer des segments dont les caractéristiques temporelles intrinsèques sont fondamentalement différentes. Pour revenir sur la remarque de Zellner (1998), ces différentes mesures ne permettent pas de rendre compte des variations qualitatives ET quantitatives engendrées par les modifications de débit.

Pour tenter de neutraliser l'effet des variations de débit, plusieurs solutions ont été adoptées. Pfitzinger (1996) teste les mesures du débit à partir du nombre de phonèmes ou de syllabes par seconde et conclut que ces deux mesures isolées ne sont pas satisfaisantes ; en revanche, il est plus probable que le débit soit une combinaison des deux. Roach (1998) cite à ce sujet les travaux de den Os (1988)⁸ dans lesquels aucune différence significative ne ressort entre le néerlandais et l'italien en terme de nombre de

⁸ den Os, O.A. (1988). *Rhythm and Tempo of Dutch and Italian*. Utrecht: Drukkerij Elinkwijk.

syllabes par seconde, alors que l'italien semble être plus lent en terme de phonèmes par seconde. En conséquence, le modèle de Pfitzinger (1999) estime le débit en fonction de la combinaison linéaire du débit en terme de syllabes ET de segments.

Citons enfin l'article de Roach (1998) dans lequel l'auteur explique que, contrairement à ce qui est généralement avancé dans la littérature, il n'existe pas nécessairement de différence de débit entre les langues. De la même manière qu'une langue est perçue comme étant plutôt à rythme accentuel ou plutôt à rythme syllabique (cf. Chapitre 4), certaines langues sont perçues comme étant plus rapides que d'autres sans qu'il y ait pour autant de différences physiques mesurables. Les variations de débit sont liées au contexte, à l'état émotionnel et aux caractéristiques individuelles, plutôt qu'aux caractéristiques intrinsèques des langues.

1.5. Résumé et conclusion

Ce premier chapitre nous a permis d'introduire un certain nombre de notions. Nous avons abordé la question de la prédiction de l'organisation temporelle de l'anglais et des problèmes soulevés par une telle tâche. Nous avons ensuite montré que subsiste toujours la problématique du choix de l'unité temporelle à utiliser pour une telle prédiction. On trouve trois types de modèles prédictifs, la distinction se fondant sur l'unité choisie.

Le but des modèles segmentaux est de prédire la durée des segments en fonction d'un ensemble de paramètres segmentaux, contextuels et syntaxiques (Klatt ; Van Santen). Ces modèles posent le problème majeur de la variabilité de la durée des segments du fait qu'ils sont influencés par l'interaction de nombreux facteurs de nature différente. De plus, le désavantage majeur d'une telle approche est la variabilité importante de la durée des segments et le risque d'accumuler des erreurs de prédiction dans le cadre d'un énoncé entier.

Le deuxième type de modèle est le modèle syllabique, le but est alors de prédire la durée des syllabes puis celle des segments qui les composent (Campbell). Cette méthode permet d'éviter les problèmes de variabilité des durées segmentales et d'accumulation d'erreurs de prédiction, mais pose le problème de la syllabification en anglais. En effet, prédire la durée des syllabes suppose un découpage préalable en syllabes ; or, la syllabification en anglais est au cœur de nombreux débats et la question qu'elle soulève est celle du regroupement des consonnes intervocaliques ou plus

précisément celle de l'ambisyllabité. Nous consacrerons le Chapitre 2 à cette problématique.

Le troisième type de modèle est de nature rythmique. Dans cette approche, le but est de prédire la durée des unités rythmiques pour ensuite déterminer la durée des syllabes (et segments) qui les composent. Plusieurs modèles ont été proposés pour l'anglais, mais se pose alors la question de la définition de l'unité rythmique. En effet, le modèle classique (Abercrombie/ Halliday) et le modèle proposé par Jassem ne reposent pas sur la même unité. Il convient par conséquent de déterminer quelle unité permet de mieux rendre compte de la structuration rythmique de l'anglais. Notons que même dans le cadre d'une approche segmentale ou syllabique, il est important de déterminer quelles sont les unités supérieures qui influent sur la durée segmentale ou syllabique. Nous consacrerons les Chapitres 3 et 4 à cette problématique.

Ce premier chapitre nous a donc permis d'introduire les différentes problématiques que nous allons aborder et tester dans la deuxième partie de ce travail. Il convient de préciser que le but de ce travail n'est pas d'obtenir un système de prédiction de la durée des phonèmes, des syllabes ou des unités rythmiques, mais d'observer l'influence de certains phénomènes sur l'organisation temporelle de l'anglais britannique et d'essayer d'en évaluer leur importance relative.

CHAPITRE 2.
PROBLEMATIQUE DE L'AMBISYLLABICITE

2.1. INTRODUCTION : ROLES DE LA SYLLABE	69
2.2. STRUCTURE INTERNE DE LA SYLLABE.....	73
2.2.1. Représentation linéaire de la syllabe.....	73
2.2.2. Représentation non-linéaire de la syllabe.....	74
2.2.2.1. Introduction de la structure hiérarchique	74
2.2.2.2. Phonologie autosegmentale et phonologie CV	74
2.2.2.3. Phonologie de Dépendance	75
2.2.2.4. Phonologie métrique	76
2.3. AMBISYLLABICITE : PROBLEMATIQUE DE LA SYLLABIFICATION.....	80
2.3.1. Syllabification et resyllabification : définitions	80
2.3.2. Principe d'attaque maximale et contrainte de rime branchante	82
2.3.3. L'ambisyllabité.....	85
2.3.3.1. Contraintes phonotactiques de syllabification.....	86
2.3.3.2. Variations allophoniques : aspiration, glottalisation et <i>flapping</i>	90
2.3.3.3. Syllabification et coarticulation	92
2.3.3.4. Travaux en psycholinguistique.....	96
2.3.3.5. Synthèse	97
2.3.4. Syllabe vs syllabité.....	99
2.4. RESUME ET CONCLUSION.....	100

Dans le premier chapitre, nous avons introduit la problématique de la définition de l'unité à adopter dans un modèle de prédiction de l'organisation temporelle de l'anglais. Nous avons brièvement décrit les modèles syllabiques, principalement celui de Campbell (1992a), qui fonctionnent en deux étapes : prédiction des durées syllabiques puis des durées des phonèmes qui les composent. Un tel modèle suppose un découpage syllabique de la chaîne parlée ; or, cette question pose beaucoup de problèmes en anglais et fait l'objet de nombreux débats. Cette méthodologie soulève par conséquent la question de la syllabification.

Dans ce deuxième chapitre, nous allons nous intéresser à la problématique de la syllabification en anglais ou plus précisément à celle de l'ambisyllabité. La première partie de ce chapitre concerne la définition et le statut phonologique de la syllabe car, si faire référence à la syllabe semble commun, en donner une définition pose des problèmes et aucun consensus ne ressort clairement de la littérature. Un excellent état de l'art sur la syllabe est présenté dans l'article de Meynadier (2001), dans lequel l'auteur décrit les différentes théories de la syllabe ; nous présentons ici uniquement les grandes lignes de ces différentes théories. La deuxième partie de ce chapitre est ensuite consacrée à la définition de l'ambisyllabité et à la présentation des divers arguments en faveur et défaveur de ce phénomène.

2.1. Introduction : rôles de la syllabe

L'évolution de la syllabe et de sa conception témoignent de l'évolution de la phonologie en général. Dans *The Sound Pattern of English* de Chomsky et Halle (1968), la syllabe n'est tout d'abord pas reconnue, puis elle est réintégrée sous la forme du trait [syllabique]. Hirst (1984) remarque que les auteurs ne font pas explicitement référence à

la syllabe alors qu'en contrepartie ils décrivent les séquences C_0VC_0 , à savoir la structure syllabique que l'on attribue typiquement à l'anglais. Bon nombre de recherches se posent comme problématique de définir le statut linguistique de la syllabe. La syllabe est qualifiée d'universel phonologique (Fudge 1969), d'unité phonologique de base (Hooper 1972) ou encore d'unité significative (Selkirk 1982). Une théorie linguistique se doit donc d'incorporer la syllabe dans son modèle. La syllabe est une unité linguistique formée de phonèmes agencés selon certaines règles phonotactiques ; tout comme le phonème, elle ne véhicule aucun sens, elle est plutôt employée pour former le sens d'un signe lorsqu'elle est regroupée avec un ensemble de syllabes (Pulgram 1970). Elle est descriptible en terme de forme et de frontières et elle n'a aucune fonction sinon celle de segmentation syllabique.

La syllabe permet également de rendre compte de l'organisation tonale et accentuelle d'une langue : elle a donc une fonction prosodique. Quelques divergences d'opinion subsistent, comme celle d'Angoujard (1997) qui réfute l'existence de la syllabe ; il constate que la syllabe est utilisée comme unité de base, mais aucune définition précise de la syllabe n'existe dans la littérature. Il en conclut que s'il n'existe aucune définition satisfaisante, c'est que peut-être la syllabe n'existe tout simplement pas (ou du moins qu'elle ne joue aucun rôle fondamental).

La syllabe est toutefois reconnue comme étant une unité de base : si on demande à des sujets de segmenter un énoncé, ils le feront de manière relativement systématique en syllabes. Les sujets ont la capacité de compter les syllabes facilement, même si le découpage syllabique est, lui, loin d'être aussi facile. Le dénombrement syllabique est aisé, tout autant dès le plus jeune âge (les enfants sont capables de compter les syllabes dans les comptines) que chez les adultes. Cette intuition linguistique est également présente lors de jeux de langue, de verlan en français, de "*rhyming slang*" en anglais, de rimes poétiques, etc. Le dénombrement métrique en français est fondé sur le dénombrement syllabique, la syllabe est ainsi une unité de compte. Cette conscience de la syllabe se retrouve également lors de manifestations de troubles pathologiques ; certains phonologues ont même montré que l'unité syllabique était la première à apparaître et la dernière à disparaître, que ce soit au niveau de l'acquisition ou des troubles pathologiques du langage.

Pour Kahn (1976), il existe une relation terme à terme entre les segments [+syllabiques] et les syllabes ; le dénombrement syllabique correspond donc à un

dénombrement de « sommets syllabiques » plutôt que d'unités syllabiques strictement définies. Ces sommets sont représentés par des noyaux vocaliques, le plus souvent des voyelles, quelquefois des consonnes vocaliques. Ainsi, la syllabe est décrite en fonction du degré d'ouverture de la cavité orale, le son le plus ouvert correspondant à l'élément vocalique (Knowles 1987), ou en terme de relation de sonorité entre les composants syllabiques. Meynadier explique que l'approche de sonorité remonte à Sievers (1881)⁹, pour qui la sonorité est liée à l'impression de proéminence d'un son par rapport aux autres sons environnants. Sievers, ensuite suivi par d'autres adhérents à cette approche, établit une échelle sonore dans laquelle tous les phonèmes sont rangés selon leur degré de sonorité.

Cependant, même si un consensus général existe au sujet de la pertinence de cette échelle de sonorité, il n'existe pas d'échelle unique reconnue par tous. On peut toutefois résumer ces différents points de vue grâce à une échelle de sonorité plus ou moins commune, établie en fonction des catégories articulatoires :

Occlusives < fricatives < nasales < liquides < glides < voyelles

Ainsi, la syllabe est formée en terme de relations de sonorité suivant le patron [croissant - décroissant] avec un pic correspondant à l'élément le plus sonore de la syllabe. Les ruptures de ce patron correspondent à des frontières syllabiques. La syllabe correspond par conséquent à un pic de sonorité, optionnellement précédé ou suivi par d'autres sons moins sonores et qui sont liés au pic en terme de coarticulation (Couper-Kuhlen 1986).

Dans la même perspective, Selkirk (1982) suggère une échelle de sonorité pour quantifier les relations de sonorité grâce à un indice allant de 1 à 10. Blevins (1995) résume cette approche en proposant le principe de Généralisation de Séquences de Sonorité (*Sonority Sequencing Generalisation* ou *SSG*) qui dit que dans une syllabe donnée, il doit y avoir une montée de la sonorité jusqu'au noyau de la syllabe, avec à ce point un plateau de la sonorité. Le but de Blevins serait de généraliser ce principe à plusieurs langues, mais la variété des langues est telle que des violations de ce principe subsistent.

⁹ Sievers, E. (1881). *Grundzüge der Phonetik*. Leipzig: Breitkopf & Harter.

La syllabe est également une unité perceptive essentielle. Les expériences de Dogil et Braun (1988) montrent que si on supprime les transitions CV d'un énoncé, le message n'est plus perçu, voire même certains sujets ne reconnaissent plus leur propre langue. En revanche, si on garde uniquement les attaques vocaliques et si on remplace le reste de l'information par du bruit, le message est tout à fait compris par les auditeurs. Ces transitions CV sont par conséquent les plus faciles à saisir et à décoder, Dogil et Braun les appellent les pivots. Notons que sur cette base, ils élaborent ce qu'ils appellent le *pivot parser*, un modèle de reconnaissance de la parole qui détecte et stocke les transitions CV et se contente d'observer la structure CV quelle que soit la complexité de la syllabe.

La syllabe est présentée comme ayant un rôle important dans l'accès au lexique. Mehler, Segui et Frauenfelder (1981) montrent que la syllabe est une unité de perception, du moins pour les auditeurs français : la reconnaissance d'une séquence phonémique est nettement plus rapide lorsqu'elle correspond à une syllabe du mot. Par exemple, le mot *balcon* est plus rapidement reconnu lorsque la séquence *bal* est présentée (plutôt que *ba*) et inversement *balance* est plus rapidement reconnu si la cible *ba* est présentée (plutôt que *bal*). Notons toutefois que les auteurs ne retrouvent pas les mêmes résultats en anglais (Cutler, Mehler, Norris et Segui 1986) ; la syllabe semble donc être une unité de perception en français, mais le traitement du signal de parole semble être différent en anglais.

La syllabe est également considérée comme une unité de longueur (Fox 2000). La longueur syllabique dépend de la structure et du poids de la syllabe, la longueur d'une syllabe n'est pas une simple somme arithmétique de ses composants segmentaux. En effet, une syllabe légère est généralement caractérisée par une voyelle courte suivie ou non d'une seule consonne alors qu'une syllabe lourde est composée d'une voyelle longue (monophthongue ou diphtongue) ou d'une voyelle courte suivie de plus d'une consonne. Une voyelle longue suivie d'au moins deux consonnes forme également une syllabe lourde, elle est hypercaractérisée. Partant de cette description classique, Fox en conclut que la longueur est une propriété de la syllabe plutôt que du segment : la longueur des différents segments et la structure syllabique se combinent pour déterminer le poids des syllabes. Toutefois, l'inverse peut être défendu : il est également possible que la longueur des segments soit déterminée par la structure et le poids syllabique. Pour Fox, la longueur des segments est ajustée pour maintenir le poids de la

syllabe à laquelle ils appartiennent, ils dépendent donc du poids syllabique plutôt que l'inverse. Ainsi, la structure de la syllabe détermine la quantité des segments.

Nous venons de voir divers rôles de la syllabe présents dans la littérature, la section suivante est consacrée à la structure interne de la syllabe et aux différents types de représentation de la syllabe.

2.2. Structure interne de la syllabe

Deux grandes conceptions de la syllabe se dégagent dans la littérature : la représentation linéaire et la représentation non-linéaire (ou hiérarchique). La représentation linéaire considère la chaîne parlée comme une série de segments positionnés les uns à la suite des autres, alors que la représentation non-linéaire considère que la syllabe est composée de constituants hiérarchisés.

2.2.1. Représentation linéaire de la syllabe

Dans cette conception de la syllabe, les segments sont considérés linéairement, en tant que chaîne délimitée par des frontières mais non organisée en une structure hiérarchique. Autrement dit, les représentations phonologiques sont uniquement composées de chaînes de segments sans aucune référence à des unités structurelles de plus haut niveau, telles que la syllabe, le mot ou le pied (Hooper 1972).

La syllabe est constituée d'une chaîne segmentale linéaire composée de trois segments (l'attaque, le noyau et la coda, le noyau étant l'élément minimal) ayant un poids et un statut égaux dans la syllabe.

Le découpage des différentes séquences syllabiques s'effectue en fonction de la composition segmentale et des combinaisons consonantiques. Le cadre syllabique est donc délimité grâce à un certain nombre de règles phonotactiques qui déterminent l'emplacement des frontières syllabiques : elles dépendent par conséquent des bornes des séquences maximales [attaque – noyau – coda] (Hooper). La syllabe est donc définie comme une chaîne maximale de segments entre deux frontières syllabiques (Venneman 1978).

2.2.2. Représentation non-linéaire de la syllabe

2.2.2.1. Introduction de la structure hiérarchique

Contrairement à la conception linéaire de la syllabe, on considère ici que les phonèmes sont regroupés et reliés à un niveau supérieur de représentation : la syllabe (Fudge 1969). L'approche suprasegmentale de Kahn (1976) s'inscrit dans cette approche non-linéaire : les séquences segmentales sont regroupées et reliées au nœud S, le nœud SYLLABE. Les segments sont ainsi dominés par le nœud S et reliés par des lignes d'association selon un certain nombre de règles. Kahn offre une preuve convaincante des avantages théoriques qu'implique la reconnaissance de la syllabe comme une unité hiérarchique. Les approches les plus représentatives de ce courant non-linéaire sont la phonologie autosegmentale et la phonologie métrique.

2.2.2.2. Phonologie autosegmentale et phonologie CV

La phonologie autosegmentale fut d'abord proposée pour rendre compte des phénomènes tonals, les tons agissant indépendamment des segments avec lesquels ils sont associés (Goldsmith 1990). Les représentations autosegmentales sont constituées de deux ou plusieurs lignes (*tiers*) et pas seulement de la chaîne segmentale linéaire. Chaque ligne constitue une unité indépendante correspondant à un type de séquences de gestes.

La phonologie autosegmentale repose principalement sur l'ajout d'une ligne squelette (*skeletal tier*) qui joue un rôle fondamental dans l'organisation de la structure phonologique ; elle est composée de positions (*slots*) vocaliques et consonantiques, on les appelle également autosegments. En somme, les segments vocaliques et consonantiques sont respectivement associés aux positions C et V du squelette : elles sont par conséquent intermédiaires entre la ligne segmentale et la ligne syllabique. Les segments sont reliés aux différentes positions sur le squelette grâce à des lignes d'association tout en respectant un certain nombre de règles visant à respecter les conditions de bonne formation (*Wellformedness Condition*). Ces conditions sont les suivantes :

- Toutes les voyelles sont associées avec au moins un ton
- Tous les tons sont associés avec au moins une voyelle
- Les lignes d'association ne peuvent se croiser

Cette représentation permet de rendre compte de divers phénomènes plus facilement, comme par exemple les segments géminés, les affriquées et les segments flottants (Goldsmith 1990), mais également la longueur et le poids syllabique.

Les éléments C et V ne sont donc pas uniquement des emplacements structuraux mais ce sont également des unités temporelles (*timing slots*). La longueur des différents segments résulte de leur association avec un nombre différent d'éléments. Cette ligne définit par conséquent les unités primitives temporelles au niveau sub-syllabique de la représentation phonologique puisque la ligne CV distingue les segments simples des segments complexes (ou géminés).

La phonologie CV est née de la phonologie autosegmentale. Clements et Keyser (1983) présentent une nouvelle approche de la syllabe dont le point de départ est l'étude de Kahn. Clements et Keyser introduisent la ligne CV (*CV-tier*) qui intervient entre la ligne segmentale et la ligne syllabique, donnant la structure présentée dans la Figure 2 (à partir d'un exemple de Kahn).

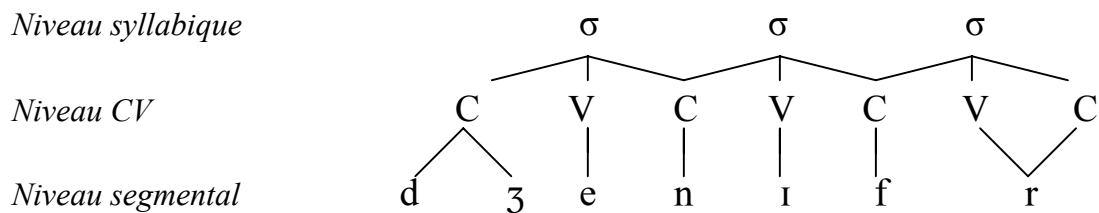


Figure 2. Représentation de la syllabe dans le cadre de la phonologie autosegmentale

Comme en phonologie autosegmentale, cette ligne est composée d'unités abstraites C et V qui qualifient la chaîne segmentale : les segments ne sont pas directement associés à la syllabe mais à la ligne CV. Grâce à cette approche, on distingue les syllabes pics et les syllabes non-pics, ceci revient en fait à remplacer le trait [syllabique] par les spécifications de la ligne CV.

2.2.2.3. Phonologie de Dépendance

La phonologie de dépendance se fonde sur l'idée que la langue repose sur une hiérarchisation de dépendance de la chaîne segmentale. Elle est basée sur les concepts de précedence linéaire et de dépendance, respectivement notés à l'aide des signes < et >

et des flèches \leftarrow et \rightarrow (Durand et Anderson 1986). Un constituant est donc défini en terme de dépendance, il est composé d'un segment et de tous les segments qui lui sont subordonnés : ce segment (le plus souvent le "syllabique") est la TETE du constituant et les segments qui dépendent de lui sont des MODIFICATEURS (*modifiers*). Ces structures sont représentées par des arbres non étiquetés dans lesquels chaque segment est associé avec un nœud, lui-même relié à d'autres nœuds par des arcs de dépendance. Un seul nœud ne dépend d'aucun autre (degré de dépendance 0), c'est la racine (Anderson 1986).

2.2.2.4. Phonologie métrique

L'approche métrique considère la parole comme l'agencement de chacun des constituants syllabiques en fonction de son degré de proéminence et du degré de proéminence des constituants syllabiques adjacents (Meynadier 2001). Ainsi, les différents nœuds d'un arbre métrique possèdent les étiquettes '*strong*' ou '*weak*' et les syllabes sont formées en fonction des diverses relations de proéminence. Les voyelles sont les plus proéminentes et les consonnes s'organisent autour de ces sommets vocaliques.

Il existe un large consensus sur la structure interne de la syllabe en phonologie métrique : la syllabe est divisée en unités intermédiaires d'ATTAQUE et de RIME, la rime étant elle-même divisée en NOYAU et CODA. La structure syllabique ainsi obtenue est illustrée dans la Figure 3 ci-après. L'attaque et la coda y sont facultatives. Cette structure semble être largement reconnue en phonologie et la syllabe se présente par conséquent comme une structure hiérarchique faite de sous-constituants. Elle est considérée comme un modèle sous-jacent, ou gabarit (*template*) chez Selkirk (1982), qui pré-définit la structure correcte de la syllabe. Selkirk ajoute à ce gabarit des contraintes phonotactiques régissant la structure hiérarchique : le gabarit reflète la relation plus proche qui lie le noyau à la coda, d'où le niveau sub-syllabique 1 qui regroupe le noyau et la coda.

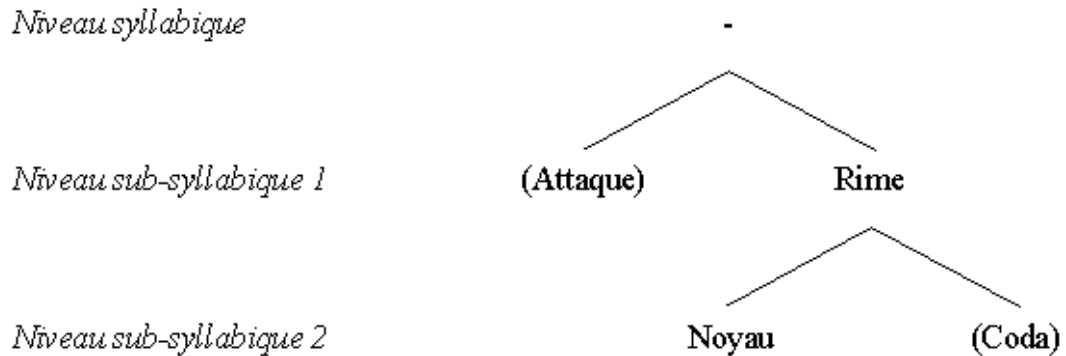


Figure 3. Représentation de la syllabe en phonologie métrique
(citée par Meynadier 2001 :115)

Plus généralement, il existe des relations de force sonore entre les constituants syllabiques (*weak* et *strong*), ceci donnant la hiérarchie de force sonore qui est la suivante (Hooper 1976¹⁰, cité par Selkirk) :

$$\text{noyau} > \text{attaque} > \text{coda}$$

Cette hiérarchie traduit en quelque sorte la représentation arborescente de la structure interne de la syllabe donnée plus haut. Ce système de "relations" de cooccurrence définit un cadre correspondant aux syllabes que l'on peut qualifier de bien formées.

Selkirk (1982 ; 1984b) utilise cette structure syllabique en tant que gabarit afin de définir l'acceptabilité des syllabes anglaises. Ce gabarit général repose sur un principe de composition fondamentale de la syllabe (*Basic Syllable Composition* ou *BSC*), regroupant un certain nombre de règles de bonne formation ainsi que des conditions de cooccurrence, tout ceci dans l'esprit de Fudge (1969) qui parle de « restrictions de cooccurrence » (*collocational restrictions*). On a par conséquent le gabarit général, accompagné du gabarit auxiliaire détaillant toutes les restrictions de cooccurrence. Ces restrictions de cooccurrence des segments affinent et complètent le gabarit général afin d'atteindre le stade de bonne formation de la syllabe.

En ce qui concerne les restrictions de cooccurrence, Goldsmith (1990) précise qu'il ne s'agit pas tant d'une question de restriction de la syllabe ou des segments, mais plutôt d'une question de restriction de lieux d'articulation : ces contraintes sont liées au

¹⁰ Hooper, J.B. (1976). *An Introduction to Natural Generative Phonology*. New York, Academic Press.

lieu d'articulation des segments en position d'attaque et de coda. On peut en général avoir au maximum un contraste de lieu d'articulation de deux segments dans l'attaque, voire un contraste supplémentaire, mais dans ce cas-là, il ne peut qu'apparaître dans la coda.

La sous-section suivante décrit certains arguments en faveur du niveau sub-syllabique 1 ([attaque + rime]), puis ceux en faveur du niveau sub-syllabique 2 ([attaque + noyau + coda]).

- **Syllabe : attaque et rime**

Cette subdivision est principalement liée à des contraintes de cooccurrence. Kurylowicz (1948) divise la syllabe en [attaque + rime], de même que le gabarit proposé par Selkirk (1982) suit également cette structure sub-syllabique.

Le but de cette sous-section n'est pas de détailler les restrictions de cooccurrence entre les différents constituants (cf. O'Connor et Trim 1953 ou Gimson 1989 pour une description détaillée des attaques (résumées ici dans les Figure 21 et Figure 22 du Chapitre 5) et codas licites en anglais), mais de noter que la relation entre noyau et coda est beaucoup plus contrainte que la relation entre attaque et noyau et que l'inventaire des consonnes codaïques est beaucoup plus limité que celui en attaque, limitée de par la position qu'occupent les consonnes codaïques et de par les caractéristiques du noyau vocalique adjacent. Ces limites sont une preuve de la pertinence du premier niveau de pré-découpage en attaque et rime.

Cette relation privilégiée entre le noyau et la coda se retrouve également au niveau sonore puisque les segments s'organisent autour du pic de sonorité (le noyau) en fonction du degré de sonorité des composants syllabiques (Selkirk 1984b). Durand et Anderson (1986) et Anderson (1986) reconnaissent également la nécessité de la rime dans le cadre de la phonologie de dépendance. En effet, Anderson (1986) remarque que dans la rime, plus généralement dans la syllabe, il existe une correspondance entre degré de sonorité et degré de dépendance : les segments les moins sonores dépendent des segments les plus sonores.

Clements et Keyser (1983) refusent l'idée selon laquelle il existe des contraintes de cooccurrence d'une part entre l'attaque et le noyau, et d'autre part entre le noyau et la coda. Ils représentent l'organisation des séquences de segments en syllabes, mais ne distinguent que les segments pics et non-pics. Le segment pic (noyau) forme une unité

de la syllabe mais aucune unité équivalente à la rime n'existe, de même que les segments non-pics ne sont pas agencés en groupes consonantiques au sein de la syllabe. Or, l'article de Fudge (1987a) vise à contrecarrer ces arguments et à prouver la validité de la rime. Fudge montre que la division de la syllabe en [attaque + rime] ne doit pas être remise en question, ses arguments se situant autant au niveau de la cooccurrence entre les segments, qu'à celui des erreurs et jeux de langue et du *rhyming slang* en anglais.

Durand et Anderson (1986) ainsi que Fudge (1987a), plus tard confirmés par Blevins (1995), montrent que cette division de la syllabe en [attaque + rime] permet également de mieux rendre compte des oppositions de poids de syllabe. Le traitement de la distinction lourde/légère de la syllabe est rendu possible par ce découpage spécifique : une syllabe légère correspond à une rime simple de type CV, alors qu'une syllabe lourde correspond à une rime lourde dite branchante. La quantité syllabique dépend donc de la rime, pas de l'attaque. Il est évident que le poids d'une syllabe est déterminé par les caractéristiques de la voyelle ainsi que par les consonnes codaïques, pas par la ou les consonnes initiales.

Treiman (1989) puis Treiman et Kessler (1995) montrent la validité de la rime à travers plusieurs travaux ; Treiman (1989) cite une série d'expériences (Treiman 1983¹¹) sur des jeux et erreurs de langue et des erreurs de mémoire en anglais : non seulement les sujets découpent les syllabes de type CCVCC selon la structure [attaque - rime] (CC/VCC), mais cette structure est jugée comme la plus naturelle par les sujets et la plus facile à appliquer lors de jeux de langue. Treiman et Kessler (1995) organisent d'autres tests sur les jeux de langue dont la tâche est d'insérer une série de deux phonèmes dans des mots ou non-mots. Les auteurs observent que les phonèmes sont majoritairement insérés entre l'attaque et la rime. La frontière entre attaque et rime est par conséquent plus saillante psychologiquement.

On peut également citer l'utilisation de la rime dans certains effets plus ou moins stylistiques comme par exemple le *rhyming slang* ; les expressions se fondent sur un parallélisme portant sur la rime. Knowles (1987 : 85) cite les exemples de *china plate* pour *mate* (la rime /eit/ est reprise dans le mot *mate*) et de *plates* pour *feet* car *plates* provient de *plates of meat* et *meat* et *feet* possèdent la même rime /i:t/. Citons enfin les

effets stylistiques de « rime » et d'« allitération » (Kowles 1987) qui correspondent respectivement au parallélisme portant généralement sur la rime (*embrace/ trace*) et l'attaque (*great/ grow*). Ces procédés stylistiques reflètent la cohérence qui existe dans la division de la syllabe en attaque et rime.

- **Rime : noyau et coda**

Au niveau sub-syllabique 2 (cf. Figure 3), la rime se divise en noyau et en coda. Cette ultime subdivision interne à la syllabe permet de distinguer les sommets syllabiques (le plus souvent des voyelles) des consonnes codaïques.

Treiman (1989) montre que le découpage à l'intérieur de la rime dépend de la nature de la consonne post-vocalique. Il existe un degré de cohésion entre le noyau et la coda plus ou moins marqué selon la nature des consonnes : les liquides ont un degré de cohésion très fort avec la voyelle, les obstruantes sont au contraire toujours rattachées à la coda, les nasales sont intermédiaires. On obtient une distinction triple en liquides, nasales et obstruantes codaïques en terme de sonorité : plus la consonne est sonore, plus elle aura tendance à être regroupée avec la voyelle, et inversement. Cette distinction en terme de cohésion ne se retrouve pas avec les consonnes en position d'attaque, confirmant ainsi le découpage syllabique en [attaque + rime].

Cette première partie introductive nous a permis de situer la syllabe en fonction de son statut phonologique en anglais. La deuxième partie de ce chapitre est consacrée à la première problématique de ce travail : l'ambisyllabité.

2.3. Ambisyllabité : problématique de la syllabification

2.3.1. Syllabification et resyllabification : définitions

La SYLLABIFICATION est l'agencement des segments à l'intérieur des constituants syllabiques selon la structure de la syllabe propre à chaque langue, cet agencement se faisant dans des mots sortis de leur contexte. En revanche, on parle de RESYLLABIFICATION lorsque la structure syllabique initiale est modifiée quand le mot est intégré dans un énoncé. On a donc une syllabification initiale qui va être ou non

¹¹ Treiman, E. (1983). The structure of spoken syllables: evidence from novel word games. *Cognition*,

modifiée lorsque le mot est considéré dans un énoncé. Citons en illustration l'exemple français donné par Meynadier (2002) : lorsque le mot *parc* [paʁk] est intégré dans *parc à jouets*, il est resyllabifié [paʁ.ka.ʒwe], les points marquant les frontières syllabiques. On voit dans cet exemple que la syllabification initiale du mot *parc* est modifiée lorsque celui-ci est intégré dans l'énoncé *parc à jouets*, même s'il s'agit ici d'un mot monosyllabique.

La question de syllabification, mais surtout de resyllabification, pose le problème fondamental du découpage syllabique et donc du regroupement des consonnes intervocaliques, ou « *interludes* » selon Troubetzkoy (1939) et Hockett (1965). Le rôle linguistique de la syllabe est désormais largement reconnu, mais on ne peut pas en dire autant de la définition de ses frontières.

Prenons le mot anglais *city*, la syllabification de ce mot peut être [sɪt.ɪ], [sɪ.tɪ] ou [sɪt.tɪ]. Le même type de choix se présente avec plus d'une consonne intervocalique comme dans le mot *Boston* (exemple cité par Kahn 1976), dont la syllabification peut être [bɒ.stən], [bɒs.tən] ou [bɒst.ən], mais également [bɒs.stən] ou [bɒst.tən]. Autrement dit, en présence d'une consonne intervocalique, trois possibilités de syllabification existent :

- La consonne appartient à la syllabe suivante : elle est en position d'attaque. Ceci correspond initialement pour Hooper (1972) au principe d'attaque obligatoire (*Obligatory Onset Principle*), ensuite devenu le principe de syllabité ouverte maximale pour Pulgram (1970) ou encore le principe d'attaque maximale pour Kahn (*Maximal Onset Principle* ou *MOP*). Un maximum de consonnes intervocaliques doit être placé en position d'attaque de la syllabe suivante, tout en respectant les conditions phonotactiques de bonne formation.
- La consonne appartient à la syllabe précédente : elle est en position de coda. Ceci correspond à la contrainte de rime branchante (*Branching Rhyme Constraint*) chez Selkirk (1982) et Schiller, Meyer et Levelt (1997).
- La consonne appartient aux deux syllabes adjacentes : elle est ambisyllabique.

En conséquence, trois types de syllabification se distinguent et ce sont ces trois approches que nous allons voir dans les sections suivantes.

2.3.2. Principe d'attaque maximale et contrainte de rime branchante

Le premier de ces deux principes relève en partie d'une certaine universalité tendant vers la syllabation ouverte, à savoir un découpage du type CV.CV. Par conséquent, une consonne intervocalique est normalement rattachée à la syllabe suivante, même si cette consonne représente une coda licite de la première syllabe (cf. par exemple Fudge 1984 ; Knowles 1987). Ce principe est également avancé en réponse aux propositions de syllabification ambisyllabique (section 2.3.3) notamment pour des raisons de hiérarchie prosodique car les éléments appartenant à un niveau prosodique sont inclus dans une construction unique (Selkirk 1982).

Le principe de regroupement en position codaïque de première syllabe découle en anglais de la contrainte de rime branchante selon laquelle une syllabe doit être lourde pour porter un accent. En effet, une syllabe ouverte avec une voyelle courte ne peut être accentuée : la consonne intervocalique doit être rattachée à la syllabe précédente pour former une syllabe lourde accentuable, tout en respectant les contraintes phonotactiques de bonne formation. C'est l'argument avancé notamment par Selkirk (1982), Wells (1990b) et Schiller, Meyer et Levelt (1997).

L'approche alternative de Pulgram (1970) consiste en une série ordonnée de principes de syllabation. Le premier est celui de syllabité ouverte maximale (*Principle of Maximal Open Syllabicity*) : lorsque cela est possible, une syllabe sera ouverte, ceci excluant les voyelles /æɒʊʌɪ/ car elles doivent être suivies par une consonne codaïque, celle-ci étant alors la plus courte possible. Autrement dit, la coda doit être minimale et l'attaque maximale (*Principle of Minimal Coda and Maximal Onset*). Par conséquent, lorsque la voyelle d'une syllabe accentuée est courte, la consonne suivante est rattachée à cette syllabe accentuée et devient alors coda.

Dans un mot comme *filling*, le découpage syllabique est /^hfɪl.ɪŋ/ (puisque la syllabe */^hfɪ/ n'est pas correcte) mais la réalisation phonétique est [fɪɫ.ɪŋ] avec un arc entre le /ɫ/ sombre et le /l/ clair. En fait, il ne s'agit pas là d'une succession de deux allophones mais plutôt d'un changement phonétique affectant le phonème en cours de réalisation. En cas de groupes consonantiques irréguliers (ne correspondant à aucune règle phonotactique), l'irrégularité portera de préférence sur la coda (*Principle of irregular Coda*) plutôt que sur l'attaque. Notons également qu'en raison de sa position finale de

syllabe, elle est nettement plus prédictible que l'attaque et donc plus sujette à la réduction et à l'effacement.

Pulgram cite les travaux de O'Connor et Trim (1953) qui font un traitement statistique des consonnes intervocaliques. Les auteurs relèvent dans un corpus le nombre d'occurrences des séquences CV, VC, CC, VV et V en position initiale et finale de mot en partant du principe que les tendances aux frontières syllabiques sont identiques à celles aux frontières de mots. En fonction de ces occurrences, ils calculent ensuite un indice de probabilité pour qu'une séquence VCV soit découpée V-CV ou VC-V. La liste des groupes ainsi que leurs occurrences sont présentées dans le Tableau 1 ci-dessous.

Séquences	Position initiale de mot	Position finale de mot
CV	421	276
VC	209	277
CC	26	59
VV	10	22
V	12	

Tableau 1. Occurrences des différentes séquences en fonction de leur position initiale ou finale de mot (O'Connor et Trim 1953 : 121)

Par exemple, pour le mot *city* pouvant être découpé VC-V et V-CV, on compare les occurrences de chacune des séquences et le nombre le plus grand l'emporte. Pour *city*, on a tout d'abord 277 (VC en finale) + 12 (V en initiale) = 289 et 12 (V en initiale) + 421 (CV en initiale) = 433 ; c'est donc le découpage V-CV qui l'emporte, à savoir *ci-ty*. Un tel découpage est incorrect selon l'approche de Pulgram (cf. également Wells 1990b) puisque la voyelle courte /ɪ/ ne peut se retrouver en contexte de syllabe ouverte : la consonne intervocalique doit être regroupée avec la première syllabe, on a donc *cit-y*.

Le problème majeur avec l'approche de O'Connor et Trim, selon Pulgram, est qu'elle peut uniquement être employée avec des consonnes intervocaliques simples. En effet, dans le cas de groupes consonantiques avec deux, trois voire quatre consonnes, le tableau de données utilisé serait bien plus complexe (les séquences seraient multipliées) et donc rapidement plus difficile à exploiter. De plus, Pulgram reproche également aux

auteurs de ne pas tenir compte des contraintes liées au type de voyelles environnantes (comme par exemple l'impossibilité d'avoir une voyelle courte dans une syllabe ouverte).

Concernant la question de syllabification et de resyllabification, Basbøll (1988) remarque que la syllabification dépend de la langue et qu'il est donc difficile d'établir des principes universaux ; de plus, la syllabification dépend surtout de la force des voyelles adjacentes : les voyelles accentuées ont tendance à attirer les consonnes médiales. De plus, la voyelle située après la ou les consonne(s) intervocalique(s) a(ont) plus tendance à attirer les consonnes que la voyelle précédente, ceci correspondant au principe d'attaque maximale.

Pour Wells (1990b), la syllabification en anglais se fait selon cinq principes :

1. LES CONSONNES INTERVOCALIQUES SONT RATTACHEES A LA SYLLABE AVEC LE PLUS FORT DEGRÉ D'ACCENTUATION

Par conséquent, *rising* est syllabifié /'raɪz.ɪŋ/, *enter* /'ent.ə/ et *attestation* /,æt.e.'steɪʃ.n/. Cette syllabification permet notamment de rendre compte de l'influence de la consonne intervocalique sur la durée des phonèmes précédents, comme par exemple le raccourcissement (*clipping*) de la voyelle accentuée et de la nasale dans un mot tel que *enter* sous l'effet de la plosive. Elle permet également la fermeture de la syllabe accentuée comportant une voyelle courte, syllabe phonotactiquement impossible en anglais. De plus, elle permet de rendre compte des variations allophoniques des consonnes intervocaliques, comme par exemple l'aspiration des plosives non-voisées, le *tapping* et la glottalisation du phonème /t/.

2. LORSQUE DEUX SYLLABES ADJACENTES ONT UN DEGRÉ D'ACCENTUATION IDENTIQUE, LA CONSONNE INTERVOCALIQUE EST SYLLABIFIÉE AVEC LA SYLLABE PRÉCÉDENTE

Par conséquent, *carpeting* est syllabifié /'kɑ:p.ɪt.ɪŋ/.

3. DANS LES MOTS POLYMORPHÉMIQUES, LE RATTACHEMENT SE FAIT EN FONCTION DE LA STRUCTURE DU MORPHEME

Par conséquent, *highness* (le titre) est syllabifié /'haɪn.əs/ mais *high#ness* (le fait d'être haut) est syllabifié /'haɪ.nəs/.

4. LES CONTRAINTES PHONOTACTIQUES NE DOIVENT PAS ETRE TRANSGRESSEES

Les contraintes de bonne formation de l'attaque et de la coda doivent être respectées, mais il doit en être de même pour la contrainte qui interdit l'occurrence de voyelles courtes en fin de syllabe accentuée.

5. LES AFFRIQUEES NE PEUVENT ETRE DIVISEES EN DEUX PARTIES PAR UNE FRONTIERE SYLLABIQUE

Wells précise que l'ambisyllabité représente un concept inutile et qu'il est plus juste de regrouper un maximum de consonnes intervocaliques avec la syllabe précédente.

2.3.3. *L'ambisyllabité*

L'ambisyllabité correspond au regroupement d'une consonne médiale avec la syllabe précédente ET la syllabe suivante. Dans un mot comme *city*, la consonne /t/ est rattachée aux deux syllabes adjacentes S₁ et S₂, elle est donc ambisyllabique. La représentation est donnée dans la Figure 4 ci-dessous.

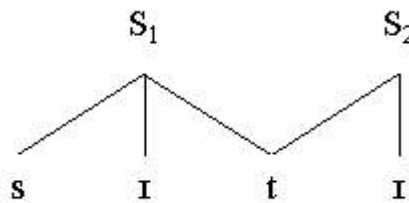


Figure 4. Regroupement des phonèmes en syllabes dans le mot *city*, avec double affiliation de la consonne intervocalique /t/ (ambisyllabité)

Dans l'introduction sur la syllabe, nous avons mentionné les conclusions de Mehler, Segui et Frauenfelder (1981) selon lesquelles la syllabe ressort comme unité fondamentale en perception ; les auteurs obtiennent avec des sujets francophones des temps de réaction plus courts lorsque les cibles visuelles présentées à des sujets correspondent à la première syllabe des mots entendus en parallèle. Ainsi, en français, le temps de réaction est plus court lorsque la cible *ba* est présentée pour le mot *balance* (plutôt que *bal*) ou *bal* pour *balcon* (plutôt que *ba*) ; en revanche, les résultats diffèrent

en anglais laissant penser que certaines consonnes en anglais sont ambisyllabiques. Notons toutefois que ces résultats n'ont jamais été retrouvés depuis par les auteurs.

Commence alors le débat sur la validité du phénomène d'ambisyllabité en anglais. On trouve dans la littérature un ensemble d'arguments en faveur ou défaveur de l'ambisyllabité mais nous tenons d'abord à préciser que nous avons éprouvé quelques difficultés dans nos recherches sur ce sujet. En effet, nous avons trouvé peu de références portant sur cette problématique, ou du moins peu de travaux expérimentaux ou d'ouvrages portant spécifiquement sur ce sujet, alors que l'enjeu de l'ambisyllabité nous semble être très important puisqu'il est au cœur même de la problématique de découpage syllabique.

Cette section détaille les différents arguments relatifs à l'ambisyllabité en anglais. Le but de cette présentation est d'essayer de déterminer comment s'effectue le regroupement des consonnes intervocaliques en anglais pour évaluer à terme son enjeu dans le cadre de l'organisation temporelle de l'anglais et de sa structuration prosodique.

2.3.3.1. Contraintes phonotactiques de syllabification

Le premier argument en faveur du principe d'ambisyllabité concerne l'impossibilité en anglais d'avoir une syllabe ouverte accentuée avec une voyelle courte. Lorsque ce cas de figure se présente dans un mot plurisyllabique, la consonne intervocalique (ou une des consonnes intervocaliques) est alors regroupée avec la syllabe précédente et la syllabe suivante. Ce double rattachement permet d'éviter les syllabes accentuées ouvertes avec une voyelle courte. Cette même situation se retrouve en hollandais (Schiller, Meyer et Levelt 1997) car, pour qu'une voyelle courte porte l'accent, il faut que la consonne intervocalique suivante soit ambisyllabique afin que la syllabe soit fermée.

Kurylowicz (1948) intègre le principe d'ambisyllabité à son modèle. Les groupes consonantiques intervocaliques avec deux ou plus de deux consonnes sont divisés en deux groupes : un groupe codaïque (composé de phonèmes formant une suite codaïque licite) et une attaque (attaque licite) avec un chevauchement ambisyllabique lorsque les deux conditions sont respectées (coda et attaque licites).

Kahn (1976) adopte également l'ambisyllabité dans son modèle de description de l'anglais américain. Plus exactement, la syllabification se fait en premier lieu en

fonction du principe d'attaque maximale ; dans un deuxième temps, la syllabe accentuée attire également la consonne initiale de la syllabe inaccentuée suivante pour permettre sa fermeture, la consonne intervocalique devient alors ambisyllabique.

Kahn définit une série de règles de bonne formation des syllabes en parole lente et appliquée, complétées par la suite par d'autres règles lorsque le débit s'accélère. Une première condition à l'ambisyllabité est l'ACCENTUATION puisque pour une séquence de type V_1CV_2 , la consonne est ambisyllabique si et seulement si V_1 est accentuée et V_2 inaccentuée. L'exception à cette règle est le phonème /ŋ/ qui ne pourra en aucun cas être en position initiale de syllabe : il ne peut donc jamais être ambisyllabique. Kahn précise que l'ambisyllabité est provoquée en partie par l'accélération du DEBIT de parole. Il cite l'exemple du mot *pony* qui sera syllabifié [pəʊ.nɪ] en parole lente et appliquée, mais le /n/ deviendra ambisyllabique avec une vitesse d'élocution supérieure.

Fallows (1981) teste l'ambisyllabité en anglais américain avec des sujets natifs âgés de 4 à 10 ans, ces derniers devant répéter une des deux syllabes des stimuli dissyllabiques de type *sham.shampoo* pour *shampoo*. Les résultats révèlent que l'ambisyllabité apparaît dans seulement 22% des cas ; l'auteur suppose que l'ambisyllabité est une stratégie utilisée lorsque deux règles de syllabification sont en conflit. Par exemple, la consonne /v/ dans *seven* est ambisyllabique pour respecter le principe d'attaque maximale ET pour fermer la syllabe accentuée avec voyelle courte. Ce phénomène est accru lorsqu'il s'agit de consonnes liquides et nasales, il concerne alors 47% des consonnes.

Les différentes approches développées dans le cadre de la phonologie de dépendance intègrent l'ambisyllabité (Anderson et Jones 1974 ; Jones 1976 ; Ewen 1982 ; Durand et Anderson 1986 ; Anderson 1986). Pour Anderson (1986), abandonner l'ambisyllabité compliquerait les règles de structure syllabique. Ainsi, il adopte la syllabification maximaliste générale (*General Maximalist*) qui dit qu'une consonne intervocalique appartient à la fois à la syllabe précédente et à la syllabe suivante, sauf aux frontières de pieds ; en effet, dans le mot *pedigree* par exemple, la séquence /gr/ appartient au deuxième pied uniquement.

Le deuxième argument en faveur de l'ambisyllabité donné par Anderson concerne la longueur vocalique : l'ambisyllabité est conservée après les voyelles courtes accentuées, mais pas après les voyelles longues accentuées. Ainsi, dans *epoch*, la

consonne /p/ est initiale (donc aspirée, non glottalisée), tandis que dans *petrol*, le /t/ est ambisyllabique, (donc aspiré et glottalisé). Le traitement des voyelles courtes accentuées nécessite l'ambisyllabité pour éviter les syllabes ouvertes et accentuées.

Gussenhoven (1986), fortement influencé par l'approche de Kahn, intègre également le principe d'ambisyllabité pour la même raison. Il cite l'exemple de *alley* dans lequel la voyelle /æ/ ne peut être accentuée si la syllabe est ouverte : le /l/ est alors en position de coda. Cependant, la consonne latérale est claire ce qui signifie qu'elle est en position d'attaque. Ces deux caractéristiques contradictoires mènent Gussenhoven à conclure que le /l/ est ambisyllabique, tout comme Fallows (1981) suppose que l'ambisyllabité apparaît lorsque deux types de syllabification sont en conflit.

Fudge (1987b) critique l'intégration de l'ambisyllabité dans la phonologie de dépendance et cite un problème majeur soulevé par cette approche. Il reprend l'exemple de *filling* (Pulgram 1970) dans lequel le /l/ devrait être un /l/ sombre si on accepte l'ambisyllabité, chose qui est absolument fausse (cf. également Knowles 1987). Il est intéressant de voir que, contrairement à Fudge, cette spécificité du /l/ tient lieu d'argument en faveur de l'ambisyllabité pour Gussenhoven.

De plus, Fudge soulève un argument supplémentaire en faveur de l'ambisyllabité : celui du raccourcissement des voyelles sous l'effet de la consonne intervocalique (ce que Wells (1990b) appelle « *clipping* »). En effet, Fudge explique que l'ambisyllabité est nécessaire dans des cas comme *writing* pour rendre compte du raccourcissement de la diphtongue sous l'effet de l'occlusive non voisée. Par conséquent, plutôt que d'intégrer l'ambisyllabité au niveau sous-jacent et ajouter des règles pour éviter le problème de *filling*, Fudge préfère la supprimer et ajouter des règles pour expliquer les cas comme *writing*.

Giegerich (1992) explique que la syllabification se fait en fonction de la Règle de Frontières Syllabiques (*Syllable-Boundary Rule*) qui dit qu'à l'intérieur des mots, les frontières syllabiques sont situées de sorte que les attaques soient maximales (tout en respectant les contraintes phonotactiques de la langue concernée). Ainsi, *enigma* est syllabifié *e.nig.ma*. Cette règle de syllabification fonctionne mais elle pose le problème des mots tels que *apple* puisque d'après la règle de frontières syllabiques, ce mot devrait être syllabifié *a.pple* ; or, ceci est impossible puisque pour que la première syllabe soit accentuée, il faut qu'elle possède une rime branchante, ce qui n'est pas le cas ici. La

syllabe accentuée doit donc être *ap*, mais la consonne /p/ doit toujours appartenir à la deuxième syllabe pour répondre à la règle de frontières syllabiques. La consonne est par conséquent ambisyllabique.

Dans le but de tester l'ambisyllabité, Giegerich teste cette hypothèse en demandant à des sujets natifs de répéter chaque syllabe d'un mot deux fois ; on obtient ainsi *ap.ap.ple.ple*. Ce procédé de répétition est un moyen évident pour observer comment les sujets syllabifient les énoncés. La consonne /p/ de *apple* est donc bien considérée comme étant ambisyllabique. Ainsi une consonne est ambisyllabique si elle appartient à une attaque qui répond aux contraintes phonotactiques et si elle suit immédiatement une voyelle courte accentuée.

Giegerich remarque que la règle d'ambisyllabité s'étend au-delà des frontières de mot, notamment en parole continue, on parle alors de liaison. Giegerich cite l'exemple *these are old eggs*, dans lequel la syllabification en parole continue sera /ði:.zɑ:.rəʊl.degz/, mais les consonnes originellement codaïques ne sont pas totalement dissociées des syllabes précédentes : elles sont ambisyllabiques (voir Figure 5 ci-dessous).

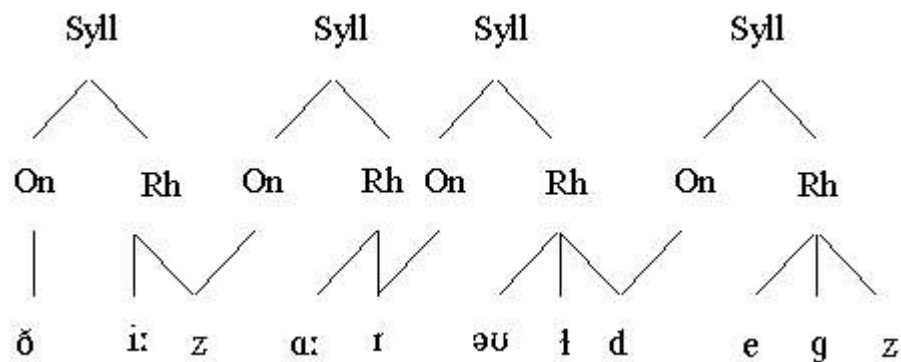


Figure 5. Découpage syllabique de l'énoncé *these are old eggs*
(Giegerich 1992 : 280)

Dans le même esprit, Deschamps, Duchet, Fournier et O'Neil (2004) citent l'exemple *red ink* ; la deuxième syllabe possède une attaque vide et le /d/ final de la première syllabe est alors rattaché à la deuxième syllabe. Cette liaison peut être interprétée comme étant le rôle ambisyllabique de la consonne finale puisque celle-ci est alors à la fois coda et attaque. Ce double rattachement peut s'expliquer

historiquement par l'existence de la forme phonologique double de l'article indéfini : *a* devant un mot commençant par une consonne et *an* devant un mot commençant par une voyelle (2004 : 25). Ce double rattachement est d'autant plus nécessaire lorsque la voyelle de la première syllabe est relâchée comme c'est le cas avec la voyelle /e/ dans l'exemple. Cependant, quand la première syllabe est composée d'une diphtongue ou d'une voyelle tendue et d'une coda vide, il peut quand même y avoir réalisation de l'élément de liaison, par exemple -r- (*higher ark*) ou /j/ (*high ark*), et il arrive également qu'il y ait introduction d'une consonne de liaison non présente dans la graphie comme le *intrusive r* : l'énoncé *law and order* peut être réalisé /'lɔ:rən'ɔ:də/ (2004 : 26).

2.3.3.2. Variations allophoniques : aspiration, glottalisation et *flapping*

Les variations allophoniques de certaines consonnes représentent un autre terrain de discussion concernant l'ambisyllabité.

L'ambisyllabité permet à Kahn (1976) d'expliquer les phénomènes d'aspiration, de glottalisation et de *flapping*. Premièrement, l'ambisyllabité empêche l'aspiration des occlusives non voisées : une consonne ambisyllabique ne peut être aspirée car elle est à la fois finale et initiale de syllabe ; or, une consonne finale ne peut être aspirée. Deuxièmement, la glottalisation est conditionnée de manière similaire puisqu'une occlusive non voisée initiale de syllabe n'est pas glottalisée, contrairement à cette même consonne en position finale de syllabe et précédée d'un segment non consonantique. Un segment ambisyllabique ne peut donc pas être glottalisé. Troisièmement, l'ambisyllabité est une condition positive au *flapping* des consonnes /t/ et /d/ (caractéristique de l'anglais américain) : pour que /t/ et /d/ soient transformés en /ɾ/, il faut que le segment en question soit à la fois final ET initial de syllabe, autrement dit ambisyllabique. Par conséquent, l'ambisyllabité empêche la glottalisation et l'aspiration et permet le *flapping*. Ces phénomènes s'appliquent à l'intérieur de mots, mais également à l'extérieur du cadre du mot.

Selkirk (1982) refuse l'ambisyllabité et montre que Kahn n'aborde pas ces trois questions d'une manière satisfaisante. Selkirk relie les consonnes médiales à la coda de la syllabe précédente, tout en respectant le gabarit syllabique (*template*) ainsi qu'un certain nombre de restrictions de cooccurrence. De ce fait, Selkirk n'intègre pas le principe d'ambisyllabité et son procédé de resyllabification entraîne la violation de la

MOP¹². Selkirk reprend les exemples d'aspiration, de glottalisation et de *flapping* cités par Kahn afin de montrer que ces phénomènes peuvent être expliqués différemment.

En ce qui concerne l'aspiration, Selkirk stipule d'abord les conditions d'aspiration, à savoir qu'une occlusive sourde en position initiale de syllabe, devant une voyelle accentuée, ou dans le contexte 'plosive + sonorante + voyelle inaccentuée' ou 'plosive + sonorante + voyelle accentuée', est aspirée sauf si elle est précédée du phonème /s/. Kahn donne comme exemple le mot *wacky* et explique que la consonne médiale n'est pas aspirée parce qu'elle est ambisyllabique. Selkirk, elle, explique que la resyllabification fait que la consonne médiale est rattachée à la coda de la première syllabe, de ce fait, elle n'est pas initiale de syllabe et donc non aspirée. Pour 's + obstruante', Selkirk explique que l'obstruante n'est pas aspirée pour deux raisons possibles : soit il n'y a pas eu de resyllabification, soit le groupe consonantique entier a été syllabifié et apparaît donc en position finale de la première syllabe.

En ce qui concerne la glottalisation et le *flapping*, Selkirk fait intervenir le trait de relâchement : une consonne est glottalisée quand elle n'est pas relâchée (*atkins*), et une telle consonne sera remplacée par le segment /ɾ/ (*flapped*) si elle est relâchée. Selkirk introduit donc le trait [release] (relâchement) à son approche, mais ne recourt pas à l'ambisyllabité. Ainsi, le trait [release] concerne uniquement les occlusives finales, le trait [-release] est une condition suffisante à la glottalisation et le trait [+release] au *flapping*. Par conséquent, Kahn et Selkirk proposent de rendre compte des mêmes phénomènes mais avec des syllabifications différentes.

Selkirk ajoute en conclusion que certains phonéticiens ont soutenu que l'ambisyllabité existait en tant qu'"*intuition*", intuition correspondant à une impression de frontière à l'intérieur de la consonne médiale (Kenyon 1950¹³), d'une appartenance phonétique de la consonne médiale aux deux syllabes adjacentes (Smalley 1968¹⁴) ou d'une fonction double de la consonne médiale dans la syllabe (Pike 1967¹⁵). Selkirk explique que ces intuitions découlent d'une sorte de compromis entre une perception de cette consonne comme finale (coda) et une tendance universelle consistant à considérer une consonne comme initiale de syllabe (attaque), autrement dit cette impression

¹² MOP : *Maximal Onset Principle*

¹³ Kenyon (1950). Ouvrage cité par Selkirk.

¹⁴ Smalley (1968). Ouvrage cité par Selkirk.

¹⁵ Pike, K. (1967). *Language in Relation to a Unified Theory of the Structure of Human Behavior*. The Hague: Mouton. Ouvrage cité par Selkirk.

correspond à un effet combiné de la perception des consonnes comme finales de syllabe en représentation de surface et du savoir intuitif que ces consonnes sont initiales au niveau sous-jacent. C'est dans cet esprit que Wells (1990b : 86) remarque que l'ambisyllabité est présente pour satisfaire une certaine préférence universelle pour la syllabation ouverte CV.CV et les arguments allophoniques (« écrasants ») en faveur d'un découpage en CVC.V.

2.3.3.3. Syllabification et coarticulation

La question de la syllabification des consonnes intervocaliques a fait l'objet de travaux en phonétique articulatoire. La méthode souvent utilisée, notamment par Krakow (1989) et Turk (1994), est de comparer les caractéristiques articulatoires des consonnes syllabifiées sans ambiguïté et celles des consonnes dont la syllabification pose problème, typiquement les consonnes intervocaliques. Krakow (1989) observe que les consonnes nasales intervocaliques ont tendance à être regroupées avec la syllabe portant l'accent primaire. L'auteur conclut que, malgré une certaine variabilité, cette syllabification fondée sur le placement de l'accent fonctionne globalement bien.

Turk (1994) observe l'affiliation des plosives bilabiales intervocaliques en anglais américain (à l'intérieur des mots) en utilisant la méthode des micro-rayons X. Cette méthode permet d'observer de manière dynamique la progression des articulateurs dans une production donnée grâce à des coupes sagittales de la cavité buccale, on obtient alors une description précise de la géométrie des divers articulateurs.

En accord avec les travaux de Treiman et Danis (1988), les résultats de Turk montrent que cette affiliation dépend en premier lieu du caractère accentuel de la syllabe précédant/ suivant la consonne intervocalique en question :

- Dans les mots dissyllabiques accentués sur la deuxième voyelle, les caractéristiques kinématiques des plosives bilabiales intervocaliques utilisées correspondent à celles des mêmes consonnes en position initiale de syllabe (*repair*).
- Dans les mots dissyllabiques accentués sur la première voyelle (comme *rabble* et *leper*), l'affiliation de la consonne intervocalique est plus ambiguë mais ses caractéristiques kinématiques ont tendance à ressembler à celles des mêmes consonnes en position finale de syllabe (coda).

- Lorsque la consonne intervocalique se situe entre deux syllabes inaccentuées, l'auteur observe certaines caractéristiques kinématiques de consonne en position initiale de syllabe, mais les résultats ne suffisent pas à tirer de conclusion fiable.

Le premier type de regroupement selon le caractère accentué ou inaccentué de la voyelle suivante est bien établi et possède des preuves psycholinguistiques et phonologiques. La question de regroupement des consonnes intervocaliques porte donc sur les mots dont la consonne intervocalique est suivie d'une voyelle inaccentuée et sur les contextes où la voyelle précédente ainsi que la voyelle suivante sont inaccentuées.

Suite aux travaux et observation de Turk (1994), Nolan (1994) explore plus en détail la problématique de syllabification au travers de mesures acoustiques et électropalatographiques (EPG) de stimuli de type /-'V_kV-/ et /-V'_kV-/ , la voyelle pouvant être /Λ/, /ɪ/ ou /ə/. La méthode EPG consiste à observer les contacts de la langue au cours de la phase de production grâce à un palais artificiel porté par les sujets, elle permet donc d'obtenir une représentation spatio-temporelle des mouvements de la langue. Contrairement aux micro-rayons X, les mesures se font de manière ponctuelle et non en continue.

Les résultats de ces deux mesures sont contradictoires dans le sens où leurs interprétations mènent à des conclusions différentes concernant le regroupement de la consonne intervocalique, mais il ressort clairement que la place de l'accent influence le lieu d'articulation de la consonne intervocalique. Nolan soulève alors le problème d'inadéquation théorique entre les structures phonologiques et les données phonétiques. Au vu des résultats, à savoir que dans les dissyllabes avec des voyelles variées, la plosive vélaire reflète une transition continue de la première voyelle à la deuxième, Nolan se demande si la syllabification des consonnes intervocaliques ne relève pas d'une tentative de mettre des frontières dans un événement physique continu (1994 : 165). Se pose alors la question de l'unicité des syllabifications phonologiques et peut-être existe-t-il des étapes successives dans la syllabification donnant lieu à des syllabifications alternatives des éléments phonétiques. La structure phonologique est alors plus complexe qu'une chaîne avec une syllabification unique.

Gick (2003) observe les caractéristiques articulatoires des consonnes /l-j-w/ en position initiale (précédées d'un mot finissant par une voyelle) et finale (suivies d'un mot commençant par une voyelle) de mot en anglais américain. Les mesures articulatoires sont la magnitude et le timing des gestes, le but étant d'observer les

corrélats phonétiques des différentes positions syllabiques. Les mesures sont effectuées en utilisant le système EMMA (*ElectroMagnetic Midsagittal Articulometer*) qui consiste en une série de petits récepteurs placés sur la langue, les lèvres, les mandibules, les maxillaires et le nez des sujets. Les résultats révèlent des caractéristiques opposées pour les consonnes en position initiale et finale de syllabe (sans ambiguïté), ainsi que des caractéristiques intermédiaires pour certains allophones qui pourraient alors être considérés comme ambisyllabiques (sauf pour /w/ qui a tendance à se rapprocher du contexte suivant).

Notons que rares sont les études portant sur la resyllabification des consonnes intervocaliques aux frontières de mots. Ce point est toutefois soulevé dans l'étude de Mullooly (2003) dans laquelle il observe le -r- de liaison dans une variété d'anglais non-rhotique grâce également à des mesures en articulographie électromagnétique (EMA), système similaire à EMMA. Cette méthode permet par conséquent de mesurer les mouvements, dans le plan sagittal, de la mandibule, de la langue et parfois du voile du palais. L'auteur utilise des stimuli du type *car out* avec modification de l'emplacement de l'accent emphatique. Les résultats sont très variables mais ils révèlent que l'environnement accentuel influence l'articulation du -r- intervocalique : l'auteur observe une tendance à ce que le *r* de liaison soit plutôt en attaque lorsque le deuxième mot porte l'accent.

Dans le domaine de la synthèse vocale, la syllabification en anglais soulève également des questions qui peuvent être mises en parallèle avec les résultats obtenus en phonétique articulatoire. Coleman (1992 ; 1994) explique que l'ambisyllabité (ou « chevauchement ») est nécessaire pour reproduire la parole dans le cadre du système de synthèse IPOX. Le découpage en constituants syllabiques [attaque + noyau + coda] suffirait à diviser les séquences phonétiques en constituants phonologiques si la parole était représentée par une suite bien ordonnée d'objets phonétiques élémentaires ; or, la parole n'est pas ainsi faite et les constituants peuvent se chevaucher. Ce chevauchement articulatoire des codas et des attaques dans les mots polysyllabiques est ensuite intégré au système YORKTALK au travers de l'ambisyllabité : elle permet ainsi de mieux rendre compte de l'organisation temporelle des syllabes.

Le principe d'ambisyllabité maximale est donc intégré au synthétiseur : un maximum de consonnes intervocaliques est ambisyllabique (en fonction des contraintes phonotactiques), à l'intérieur des mots mais également au-delà de leurs frontières. Par

opposition, Coleman montre que la concaténation des syllabes pose certains problèmes : l'interprétation phonétique des codas suppose un relâchement en fin de syllabe mais il n'est pas souhaitable de conserver ce relâchement lorsque la consonne codaïque est suivie d'une attaque de syllabe. Au contraire, on souhaite un passage souple de la coda à l'attaque et ce passage n'est possible qu'en intégrant des zones de chevauchement.

Le système YORKTALK est à l'origine du synthétiseur de parole PROSYNTH développé par les universités de York, Londres et Cambridge (Hawkins, House, Huckvale, Local et Ogden 1998). Local et Ogden (1997) y intègrent également le principe d'ambisyllabité maximale lorsque la structure le permet. Le synthétiseur prédit la durée des syllabes en fonction de leur relation avec les syllabes environnantes et de leur contexte ; l'interprétation temporelle de l'ambisyllabité est le chevauchement temporel et paramétrique d'une syllabe sur l'autre. Les consonnes intervocaliques sont ambisyllabiques du fait que les jointures entre les syllabes sont cruciales dans les groupes polysyllabiques (Ogden, Local et Carter 1999).

Les auteurs donnent trois raisons à l'intégration du principe d'ambisyllabité maximale :

- Aucun mot anglais ne peut avoir une rime avec une voyelle courte ouverte.
- L'ambisyllabité signifie que les *on-glides* sont des propriétés propres aux codas alors que les *off-glides* sont des propriétés propres aux attaques.
- La résonance secondaire des consonnes intervocaliques appartient à la fois à la voyelle précédente et à la voyelle suivante.

Cependant, l'ambisyllabité n'empêche pas les restrictions de cooccurrence d'opérer : les consonnes intervocaliques sont ambisyllabiques dans la mesure où les restrictions de cooccurrence sont respectées. Les auteurs donnent l'exemple du mot *loveless* dans lequel aucun segment n'est ambisyllabique puisque /vl/ n'est possible ni en attaque, ni en coda.

L'ambisyllabité est par conséquent intégrée aux différents systèmes de synthèse cités ci-dessus car elle permet de rendre compte plus facilement des effets segmentaux de coarticulation.

2.3.3.4. Travaux en psycholinguistique

La notion d'ambisyllabité pose également la question des procédés psycholinguistiques de segmentation syllabique, elle a par conséquent fait l'objet d'études en psycholinguistique.

Treiman et Danis (1988), ainsi que Treiman, Straub et Lavery (1994), ont travaillé sur cette problématique en anglais, au travers de la tâche d'inversion de syllabes dans des mots dissyllabiques. Ces mots varient sur trois plans : le patron rythmique (tantôt accentués sur la première syllabe, tantôt sur la deuxième), la catégorie à laquelle appartiennent les consonnes intervocaliques (liquide, nasale ou obstruante) et la complexité orthographique (consonnes orthographiques simples ou doubles).

Les stimuli varient selon ces caractéristiques spécifiques car les auteurs veulent tester l'influence de quatre facteurs principaux, supposés jouer un rôle dans la syllabification des consonnes intervocaliques :

- La position de l'accent
- La qualité de la voyelle précédente
- La qualité phonétique de la consonne intervocalique
- La complexité orthographique de la consonne intervocalique (simple ou double).

Les résultats montrent tout d'abord une influence très marquée du patron accentuel : pour les mots dissyllabiques avec accent sur la deuxième syllabe, les réponses sont majoritairement initiales. Avec les mots dissyllabiques accentués sur la première syllabe, la syllabification n'est pas régulière : les consonnes intervocaliques sont soit regroupées avec la syllabe précédente (position de coda), soit avec la syllabe suivante (position d'attaque), soit avec les deux syllabes (ambisyllabiques). Le choix de syllabification dans ce cas-là dépend d'autres critères que de celui de l'accentuation.

En effet, l'ambisyllabité dépend également de la qualité de la consonne intervocalique : les réponses ambisyllabiques sont plus fréquentes avec les liquides et les nasales, qu'avec les obstruantes, et de l'orthographe : l'ambisyllabité est plus fréquente quand le mot contient une succession de deux consonnes orthographiques identiques. Les auteurs observent également que la qualité de la voyelle influence la syllabification : si la première voyelle est relâchée et accentuée, elle aura tendance à former une unité avec la consonne intervocalique, alors que ce schéma est beaucoup moins fréquent dans le cas où la première voyelle est tendue et accentuée.

Derwing (1992) reproduit l'expérience de Treiman et Danis (1988) avec comme tâche l'insertion de pause. Les résultats obtenus pour l'anglais confirment les quatre facteurs influençant l'affiliation des consonnes intervocaliques constatés par Treiman et Danis. Ils mettent également en évidence une tendance à ce qu'une consonne soit coda d'une voyelle relâchée et accentuée, ceci variant en fonction de sa position dans la hiérarchie de sonorité (Selkirk 1984b) ou au contraire, en fonction de sa force consonantique. Derwing observe un autre facteur, non cité par Treiman et Danis : le facteur morphologique. Dans les mots bimorphémiques, les sujets ont fortement tendance à garder les morphèmes racines ensemble et à segmenter après la consonne (cf. également Wells 1990b), bien que des contre-exemples subsistent, comme par exemple les mots *sailor* et *icicle*.

2.3.3.5. Synthèse

Nous avons vu un certain nombre d'arguments en faveur ou défaveur de l'ambisyllabité, ainsi que de nombreuses études visant à déterminer les facteurs entraînant l'apparition de consonnes ambisyllabiques.

Un premier type d'argument important concerne les contraintes phonotactiques relatives à l'anglais. L'argument principal en faveur de l'ambisyllabité est le fait qu'on ne trouve pas de syllabe accentuée avec une voyelle courte ouverte en anglais, celle-ci doit être fermée par une ou plusieurs consonnes codaïques. Un deuxième argument phonotactique est celui de la longueur vocalique et le problème soulevé par Fudge (1969) dans les mots tels que *writing* où l'ambisyllabité permet de rendre compte du raccourcissement de la diphtongue /aɪ/ sous l'effet de l'occlusive non voisée /t/. En revanche, certains réfutent l'ambisyllabité (Fudge 1987b ; Blevins 1995) et Fudge cite l'exemple de mots tels que *filling*, dans lequel la consonne intervocalique devrait être le /l/ sombre si on accepte l'ambisyllabité ou un passage d'un /l/ sombre à un /l/ clair (Pulgram 1970). Or, rien dans les travaux sur l'ambisyllabité ne mentionne de "contre-règle".

Le deuxième type d'argument concerne les variations allophoniques de certaines occlusives intervocaliques (aspiration, glottalisation et *flapping*). Dans une approche telle que celle de Kahn (1976), l'ambisyllabité permet de justifier l'apparition (ou pas) de ces trois phénomènes ; cette approche est fortement remise en question par Selkirk

(1982). En effet, Selkirk pense que l'ambisyllabité n'a pas lieu d'être, et qu'il vaut mieux rendre compte des phénomènes d'aspiration, de glottalisation et de *flapping* au travers d'autres moyens (elle insère à ce sujet le trait [release] dans son analyse).

Le troisième type d'approche de la problématique se situe sous l'angle de la phonétique articulatoire. On cherche à savoir si les caractéristiques kinématiques des consonnes intervocaliques correspondent plutôt à celles en position codaïque ou à celles en position d'attaque. Il ressort des différentes études que nous avons citées que l'effet de la place de l'accent est déterminant : il influence grandement l'articulation des consonnes intervocaliques (Krakow 1989 ; Turk 1994 ; Nolan 1994 ; Mullooly 2003). Par exemple, les observations de Turk ont révélé que les caractéristiques kinématiques des occlusives intervocaliques post-accentuelles tendaient plus à ressembler à celles des occlusives en position finale de syllabe. En position pré-accentuelle, l'occlusive possède des caractéristiques kinématiques identiques aux consonnes en attaque (les résultats sont ici beaucoup plus tranchés).

Dans le cadre de la synthèse de la parole, Local et Ogden (1997) intègrent également le principe d'ambisyllabité : les jointures entre les syllabes apportent des informations dont il convient de tenir compte, comme les effets articulatoires et les phénomènes de résonance ; les deux voyelles adjacentes "colorent" la ou les consonne(s) intervocalique(s). Ainsi, l'ambisyllabité permet justement de conserver ces zones de chevauchement d'une syllabe sur l'autre et de conserver les informations qu'elles contiennent et qui, au final, contribuent au naturel de la parole synthétique des systèmes YORKTALK et PROSYNTH.

Dans un quatrième temps, nous avons vu que les conditions d'occurrence des consonnes ambisyllabiques sont observées dans diverses expériences psycholinguistiques. On sait que leurs occurrences dépendent de :

- L'emplacement de l'accent lexical (par rapport aux consonnes intervocaliques)
- La complexité orthographique des consonnes intervocaliques (simples ou doubles)
- La qualité des consonnes intervocaliques
- La qualité des voyelles adjacentes
- La structure morphémique du mot

Ce que l'on pourrait reprocher aux approches présentées précédemment, c'est la portée limitée de leurs stimuli puisqu'il s'agit de mots dissyllabiques isolés (voire de logatomes) ou des mots cibles insérés dans une phrase porteuse. On peut par conséquent

leur reprocher le manque de naturel des stimuli utilisés. De plus, il serait intéressant d'observer le phénomène d'ambisyllabité dans des mots de plus de deux syllabes (notamment pour observer la syllabification des consonnes intervocaliques entre deux syllabes inaccentuées) et dans des mots insérés dans des énoncés plutôt que des mots isolés. De plus, certaines tâches soulèvent des questions théoriques comme celle d'inversion de syllabes (Treiman et Danis 1988) car elle pose un problème fondamental, celui de la modification de la structure rythmique des mots. En inversant les syllabes, on modifie le patron rythmique et il est possible que cette modification crée un biais dans la syllabification qui en résulte.

2.3.4. Syllabe vs syllabité

Après avoir présenté toute une série de travaux concernant les problèmes que pose la syllabification en anglais, il convient de remarquer que certains ont proposé de parler de SYLLABITE plutôt que de constituance syllabique, notamment dans une langue telle que l'anglais (Halliday 1970 ; Hirst 1983 ; Jassem 1983). En effet, les arguments en faveur de la syllabe concernent la syllabité plutôt que la constituance syllabique (Hirst 1983 : 34) et les langues telles que l'anglais sont caractérisées par la notion de pic plutôt que par celle de frontière. Autrement dit, les caractéristiques de pic sont plus importantes (et plus aisément perceptibles) que celles de frontière syllabique.

Pour Hirst, les constituants syllabiques (attaque et rime) plutôt que la syllabe représentent de meilleurs candidats dans la hiérarchie des constituants prosodiques de l'anglais. Cette approche s'explique entre autres au niveau temporel puisque certains allongements portent sur l'attaque ou la rime plutôt que sur la globalité de la syllabe. C'est notamment le cas de l'allongement final qui affecte la rime et l'accent qui influence la durée de l'attaque (Campbell 1992a).

C'est également dans cette perspective que l'intervalle délimité par chaque attaque vocalique a été proposé (Fant, Kruckenberg et Nord 1991a) ; de même que le synthétiseur de Barbosa et Bailly (1994) utilise les unités délimitées par les centres perceptifs (attaques vocaliques) plutôt que la syllabe ou que Ramus (1998) observe les différences rythmiques entre diverses langues en comparant les intervalles vocaliques et consonantiques.

2.4. Résumé et conclusion

Ce deuxième chapitre est consacré à la problématique de départ de ce travail, l'ambisyllabité. Ce chapitre récapitule tout d'abord les différentes conceptions de la syllabe en phonologie, un aperçu théorique qui nous a permis de mettre en relief les désaccords théoriques portant sur la syllabe en anglais même si son statut d'unité phonologique et prosodique n'est pas (ou peu) contesté. Prendre en considération la syllabe en tant que niveau de la structuration prosodique de l'anglais pose le problème fondamental de la syllabification et de la détermination des frontières syllabiques.

La deuxième section de ce chapitre est donc consacrée à la description des travaux s'intéressant à la syllabification et plus spécifiquement à l'ambisyllabité. Notons que ces approches sont de deux types ; la première est descriptive dans le sens où elles listent les arguments pour ou contre les consonnes ambisyllabiques à partir des contraintes phonotactiques, allophoniques et articulatoires. Le deuxième type d'approche concerne les études effectuées en psycholinguistique sur lesquelles on se fonde sur l'intuition des interlocuteurs natifs lorsqu'ils se prêtent aux tâches d'inversion de syllabes, d'insertion de pauses ou de duplication de syllabes.

Aucune étude à notre connaissance n'a testé expérimentalement et statistiquement la validité de l'ambisyllabité en anglais, sauf peut-être les travaux réalisés dans le cadre du projet PROSYNTH dans lequel la durée des syllabes est prédite en fonction des zones de chevauchement entre les syllabes. De plus, la plus grande partie des études décrites portent sur de la parole de laboratoire.

Il serait par conséquent intéressant de tester la pertinence de l'ambisyllabité sur de la parole authentique afin d'observer si les consonnes ambisyllabiques se démarquent au niveau des consonnes en attaque ou en coda. De plus, il serait intéressant de tester plus en détail cette problématique en production avec des énoncés entiers et un contexte de communication plus naturel.

Ce deuxième chapitre nous a permis d'aborder le niveau syllabique de la structuration prosodique de l'anglais et ainsi d'exposer la problématique que représente la syllabification en anglais et l'impact qu'elle pourrait avoir sur un modèle de prédiction de l'organisation temporelle. Le troisième type d'approche prédictive repose sur la prédiction de la durée des unités rythmiques, mais, comme nous l'avons soulevé dans le Chapitre 1, ce type d'approche pose le problème de la définition de l'unité rythmique en anglais. Cette deuxième problématique est traitée dans le chapitre suivant.

CHAPITRE 3.

STRUCTURATION RYTHMIQUE : DEFINITION DES UNITES

3.1. DEFINITION	104
3.2. DESCRIPTION DES UNITES RYTHMIQUES	107
3.2.1. Introduction	107
3.2.2. Modèle proposé par Jassem.....	109
3.2.2.1. Description	109
3.2.2.2. Approche similaire proposée par O'Connor	113
3.2.3. Le pied accentuel.....	115
3.2.3.1. Approche d'Abercrombie.....	115
3.2.3.2. Approche de Halliday.....	117
3.2.4. Comparaison des deux modèles	117
3.3. COMPARAISON ET PERSPECTIVES DES DEUX MODELES	122
3.3.1. Traitement différentiel des syllabes inaccentuées.....	122
3.3.1.1. Syllabes inaccentuées post-accentuelles	123
3.3.1.2. Syllabes inaccentuées pré-accentuelles.....	126
3.3.2. Rôle et fonction du mot.....	127
3.3.3. Confusion entre le niveau du pied et le niveau du mot	130
3.3.4. Facteur syntaxique.....	132
3.3.5. Limitation d'étendue des facteurs rythmiques	133
3.4. PHENOMENE ACCENTUEL EN ANGLAIS.....	134
3.4.1. Définition	135
3.4.2. Corrélats acoustiques de l'accent	136
3.4.3. Distinction entre accent (lexical) et proéminence	138
3.4.4. Prédiction de l'accent et de la proéminence.....	140
3.5. RESUME ET CONCLUSION.....	142

Dans le premier chapitre, nous avons défini les modèles prédictifs comme pouvant être segmentaux, syllabiques ou supra-syllabiques ; par « supra-syllabique », nous entendons les modèles cherchant à prédire la durée des unités de niveau supérieur à celui de la syllabe, principalement le mot et l'unité rythmique. Nous nous limitons ici aux prédictions fondées sur l'unité rythmique (cf. Nooteboom (1981) pour une prédiction fondée sur le mot).

Ainsi, le but des modèles supra-syllabiques est de prédire la durée des unités rythmiques pour ensuite prédire la durée des syllabes et/ou des segments. Cette approche permet de prendre en compte les caractéristiques rythmiques de la parole, chose plus délicate lorsque l'unité prédictive est le segment. Cependant, deux approches rythmiques existent et la question qu'ils soulèvent est la définition de l'unité rythmique en anglais. Ce troisième chapitre est par conséquent consacré à la description de ces deux modèles et de leur évaluation théorique en fonction des différentes études portant sur le rythme en anglais.

La première partie de ce chapitre porte sur la définition du rythme dans sa globalité puis plus spécifiquement dans le domaine de la parole. Les deux modèles supra-syllabiques (ou rythmiques) sont ensuite détaillés et comparés, à savoir dans un premier temps le modèle de Jassem (1952) puis celui d'Abercrombie (1964) et de Halliday (1967) ; les prédictions des deux modèles sont ensuite comparées aux différentes observations présentes dans la littérature. La dernière section porte sur le phénomène accentuel à la base de tout découpage rythmique.

3.1. Définition

Dans ce travail, nous nous intéressons à l'effet de la structuration rythmique sur l'organisation temporelle de la parole en anglais britannique, mais il est difficile d'ignorer que le rythme a une existence concrète en dehors du champ de la parole. En effet, les êtres humains ont une tendance marquée vers une attitude rythmique en tant que répétition (approximativement ou strictement) régulière d'un événement, quelle que soit sa nature. Nous allons donc nous arrêter sur cette notion de rythme afin de tenter de la définir et de la mettre en perspective dans le cadre de la parole.

En dehors de toute considération langagière, les êtres humains sont caractérisés et conditionnés par une certaine rythmicité ; elle nous concerne autant au niveau biologique que dans nos faits et gestes quotidiens avec par exemple le cas de la marche : dès lors que l'on commence à marcher, une cadence s'installe automatiquement. Parallèlement, notre corps répond à des cycles au sens concret du terme : au niveau de nos habitudes quotidiennes comme l'acte de se lever, de se coucher, de manger ou de dormir, mais également au niveau biologique avec entre autres l'activité cardiaque et respiratoire. Eriksson (1991) remarque qu'il existe deux types de rythme, celui que nous construisons à partir de notre connaissance du monde, telle que l'alternance rythmique entre le jour et la nuit, et le rythme que nous percevons tel que celui de la musique. Ainsi, le corps humain est conditionné par une activité rythmique et c'est donc naturellement qu'il va reporter cette rythmicité dans son activité langagière.

La rythmicité est présente à tous les stades de l'évolution des êtres humains. Bien avant que l'acte langagier n'apparaisse dans sa forme complexe, une rythmicité est déjà instaurée au travers des babillages. Le mouvement répétitif d'ouverture et de fermeture des mandibules impose un cadre rythmique sur lequel le langage va se fonder pour se développer : lorsque le langage émerge, il existe déjà une base rythmique. En fonction de ce cadre en quelque sorte prédéfini, la production langagière est naturellement influencée par des caractéristiques biologiques telles que, par exemple, la respiration, elle-même déjà cyclique (Di Cristo 2002).

La cadence imposée inconsciemment à des événements répond souvent à une rythmicité binaire. Dans le cas de la marche, il y a alternance entre jambe gauche et jambe droite, de même que le jour et la nuit donnant lieu au fait de se lever et de se coucher, répondent eux aussi à un rythme binaire. Cette binarité inhérente à notre environnement va également se refléter sur notre perception des événements. Plus

catégoriquement, cette binarité nous conduit à regrouper des éléments de manière binaire au niveau perceptif : la tendance à percevoir des séries d'événements objectivement réguliers comme des groupes binaires. On parle alors de rythmisation subjective (Fraisse 1974 : 74) puisque le regroupement binaire (voire ternaire) se fait de manière tout à fait subjective. On cite souvent l'exemple de l'horloge : le son produit par une pendule forme une séquence de battements parfaitement régulière ; toutefois, nous percevons ces battements comme s'ils étaient regroupés par paires, d'où le fameux *tic-tac* en français, reflétant cette binarité subjective.

Cette caractéristique binaire est souvent déterminée par l'alternance d'un événement fort et d'un événement faible, comme c'est le cas dans des événements tels que les battements du cœur ou les battements en musique. Cette alternance existe également dans la parole, dans laquelle on retrouve une répétition d'éléments avec une tendance au regroupement binaire. Selkirk (1984a) présente le Principe d'Alternance Rythmique (*Principle of Rhythmic Alternation*) selon lequel la parole est caractérisée par au moins un battement fort et au plus deux battements faibles entre deux battements forts.

C'est dans le cadre de cette alternance de battements forts et faibles que le rythme langagier est défini, les battements étant rattachés aux constituants syllabiques. Mais cette définition fondée sur l'alternance entre syllabes accentuées et syllabes inaccentuées a souvent été remise en question car elle peut paraître beaucoup trop restrictive dans le cadre de la parole (notamment de la parole naturelle). Le rythme est un concept plus complexe qu'une stricte alternance entre une syllabe accentuée et une ou deux syllabes inaccentuées, ne serait-ce que si on considère les successions de pieds monosyllabiques, ils sont nécessairement exclus si nous nous en tenons à la définition d'alternance stricte.

Ainsi, le rythme langagier est défini comme étant une succession de syllabes fortes et de syllabes faibles (Selkirk 1984a), ceci en fonction de la distribution rythmique de l'accent. Les syllabes avec des accents de même niveau ont tendance à revenir à distance plus ou moins égale, entraînant des patrons d'alternance d'éléments forts et faibles relativement réguliers. Apparaît alors une caractéristique supplémentaire du rythme, celle de dimension temporelle. C'est ainsi que Di Cristo (2002) définit le rythme comme étant l'organisation temporelle des syllabes métriquement fortes et des syllabes métriquement faibles ou, plus généralement, comme l'organisation temporelle des proéminences (Di Cristo (A.) et Hirst 1997).

Il ressort de ces différents points que la notion de phénomène accentuel est directement liée au rythme, mais se pose alors la question de savoir si l'organisation rythmique engendre la production d'accents ou si les accents imposent un certain rythme. Pour Bolinger (1981), l'accentuation et le rythme sont indépendants, mais surtout, ce serait plutôt le rythme qui imposerait des patrons accentuels, l'accentuation ne serait donc pas un procédé rythmique.

Nous venons d'aborder succinctement les notions de binarité et d'alternance en parole, mais la fonction principale du rythme est celle de structuration de la chaîne parlée. Dans une séquence donnée, le rythme impose une structuration temporelle des différents éléments sous forme de regroupement. En anglais par exemple, le principe d'organisation rythmique regroupe un ensemble de syllabes faibles avec la syllabe forte qui précède pour former une unité rythmique. Le même type de structuration existe en français, mais le regroupement s'effectue avec la syllabe accentuée suivante.

La composante rythmique organise la parole tant au niveau de la production que de la perception. En effet, nos productions langagières sont contraintes par la composante rythmique spécifique à notre langue, mais celle-ci est également présente au niveau perceptif puisque l'être humain a tendance à plaquer un type de rythme sur tout énoncé perçu, ce qui explique pourquoi, pour certains, le rythme est avant tout perceptif (Fraisse 1974). Plus généralement, tout interlocuteur développe un sens aigu de la perception du rythme et perçoit inconsciemment toute régularité, de nature langagière ou pas, mais tout particulièrement il est très sensible à toute déviation par rapport à une régularité donnée (Laver 1994 : 524).

Ajoutons enfin la distinction entre le mètre et le rythme (Di Cristo 2002). Pour Di Cristo, le mètre représente la structuration théorique des énoncés contrainte sous forme de gabarits métriques (niveau profond), alors que le rythme est l'actualisation de cette représentation rythmique mentale lors de la phase de production en fonction de toutes les contraintes impliquées par cette production, comme la structure syntaxique et les contraintes sémantiques et pragmatiques (niveau de surface). Ainsi, notre message planifié possède une structure rythmique qui sera nécessairement modifiée lors de la phase de production en raison de nombreuses contraintes.

Cette brève introduction sur le rythme nous permet de voir la portée du rythme pour les êtres humains, cette importance expliquant pourquoi elle se reflète sur leur activité

langagière. En effet, la structuration rythmique est une dimension capitale en situation de communication. Roach (1991 : 37) résume cette importance de la manière suivante :

« [rhythm] helps us to find our way through the confusing stream of continuous speech, enabling us to divide speech into words or other units, to signal changes between topic or speaker, and to spot which items in the message are the most important. »

Nous allons maintenant décrire les différentes unités rythmiques qui sont présentées dans la littérature dans le but de représenter l'organisation rythmique de l'anglais. Nous nous focaliserons sur l'anglais, mais nous ferons référence à d'autres langues lorsque cela nous semblera nécessaire.

3.2. Description des unités rythmiques

3.2.1. Introduction

L'unité rythmique généralement utilisée est le pied métrique, mais il convient de préciser qu'il existe une différence fondamentale entre le pied en phonétique et le pied en phonologie. En effet, en phonétique, selon la définition classique d'Abercrombie (1964) et de Halliday (1967), un pied accentuel englobe en anglais une syllabe accentuée ainsi que toutes les syllabes inaccentuées suivantes quel que soit leur nombre jusqu'à la syllabe accentuée suivante sans que cette dernière ne soit incluse. Ainsi, un pied accentuel comprend au minimum une syllabe accentuée et au maximum une syllabe accentuée accompagnée de la totalité des syllabes inaccentuées suivantes.

En revanche, en phonologie, le pied est défini en anglais comme une unité suprasyllabique généralement dissyllabique (voire trisyllabique) avec proéminence sur la première syllabe (entre autres Selkirk 1984a ; Nespor et Vogel 1986 ; Hayes 1995). Il s'agit donc d'une unité généralement plus petite que le mot qui permet de représenter la distinction entre les syllabes accentuées et les syllabes inaccentuées.

Il existe par conséquent une différence fondamentale entre le "pied phonétique" et le "pied phonologique" puisque le premier contient un nombre illimité de syllabes tandis que le second est limité à deux ou trois syllabes.

Le "pied phonologique" s'inscrit dans la représentation hiérarchique des énoncés, telle que celle proposée par Selkirk (1984a) avec les constituants prosodiques suivants :

- Syntagme intonatif (*Intonational Phrase*)
- Syntagme phonologique (*Phonological Phrase*)
- Mot prosodique (*Word*)
- Pied (*Foot*)
- Syllabe (*Syllable*)

La taille du mot prosodique est équivalente à celle du mot, permettant ainsi de pouvoir qualifier une partie du discours comme étant initiale, médiane ou finale de mot. Le syntagme phonologique regroupe un ou plusieurs mots lexicaux. Un syntagme intonatif possède un contour intonatif spécifique, l'énoncé peut correspondre à un ou deux syntagmes intonatifs.

Le pied (phonologique) permet de déterminer la position des syllabes accentuées et inaccentuées dans les mots et unités supérieures. Selkirk représente les énoncés au travers de grilles métriques agencées sous forme de niveaux selon la hiérarchie des constituants citée ci-dessus. Elle remarque toutefois qu'il est peu probable que le pied serve de domaine pour les règles phonologiques. D'après Nespor et Vogel (1986), les grilles métriques présentent des avantages dans la détermination des accents, il n'est toutefois pas évident qu'il faille éliminer le niveau du pied en tant que domaine de règles phonologiques. Cependant, Nespor et Vogel citent un ensemble de règles phonologiques qui requièrent le pied en tant que constituant phonologique. Ils citent entre autres le cas de l'aspiration des plosives non voisées en anglais : de telles consonnes sont aspirées si elles sont initiales de pied. Giegerich (1992) cite également le cas des enclitiques : le domaine formé par les enclitiques et les mots accentués "hôtes" précédents correspond à une unité phonologique unique qui est le pied phonologique.

La définition des pieds se fonde sur une terminologie empruntée à la versification classique. Dans la rhétorique traditionnelle, le nombre de syllabes dans un pied est fixe ; un trochée est un pied binaire composé d'une syllabe longue suivie d'une syllabe courte, un iambe un pied binaire avec une syllabe courte suivie d'une syllabe longue, un sponde est un pied binaire avec deux syllabes longues et un dactyle un pied ternaire avec une syllabe longue suivie de deux syllabes courtes. Ces différentes structures impliquent le principe d'eurythmie (Hayes 1995) selon lequel les accents doivent être séparés par un espacement régulier, c'est en ce sens que le pied est considéré comme une unité rythmique temporelle (Ewen et van der Hulst 2000). Dans les langues comme

l'anglais, les pieds sont généralement décrits comme étant caractérisés par l'isochronie des syllabes accentuées : elles reviennent à intervalles quasi réguliers (Pike 1945 ; Abercrombie 1964). Dans les langues telles que le français, l'isochronie porte sur toutes les syllabes, accentuées ou pas. Nous reviendrons sur la notion d'isochronie au cours du Chapitre 4.

Hayes reprend la terminologie classique avec comme critère non plus l'allongement des syllabes mais leur proéminence ; par exemple, le trochée est un pied binaire avec proéminence initiale. Hayes précise qu'il existe trois types différents de pied qui permettent de décrire l'organisation rythmique de toutes les langues. Ces trois structures sont les suivantes :

- TROCHEE SYLLABIQUE (*syllabic trochee*) : c'est un pied dissyllabique, avec proéminence sur la première syllabe et aucune référence à la qualité de la voyelle.
- TROCHEE MORAÏQUE : ce pied consiste soit en une série de deux mores dont la première est forte, soit une seule more forte. Une more est définie par Ladefoged (1975 : 251) comme étant une unité temporelle généralement composée d'une consonne et d'une voyelle ou uniquement d'une voyelle. Le japonais est typiquement une langue dont l'unité rythmique est la more.
- IAMBE : c'est un pied dissyllabique avec proéminence à droite et syllabe faible à gauche ou un pied monosyllabique avec une syllabe forte.

Cette première sous-partie nous a permis d'introduire la notion de pied, principalement en phonologie. La section suivante est consacrée à la définition du pied accentuel en phonétique selon les deux approches de représentation du rythme en anglais, celui proposé par Jassem (1952) et celui proposé par Abercrombie (1967) et Halliday (1967).

3.2.2. *Modèle proposé par Jassem*

3.2.2.1. Description

La première de ces deux approches est celle de Jassem (1951 ; 1952) qui propose une représentation de l'organisation temporelle de l'anglais britannique fondée sur la structure rythmique. Le but du modèle est de prédire la durée des unités rythmiques

pour ensuite déduire la durée des syllabes qui les composent. Les énoncés sont découpés en deux unités différentes, à savoir :

- L'UNITE RYTHMIQUE ETROITE (*Narrow Rhythm Unit* ou *NRU*) : elle débute avec une syllabe accentuée et englobe toutes les syllabes inaccentuées suivantes jusqu'à la fin du mot. La syllabe accentuée débutant une NRU peut posséder un accent primaire ou secondaire, voire ternaire selon les définitions de degrés d'accentuation.
- L'ANACROUSE (*ANacrusis* ou *ANA*) : elle comprend toutes les syllabes inaccentuées restantes, que l'on pourrait qualifier d'extra-métriques dans le sens où elles n'appartiennent pas à la NRU.

Une ANA et la NRU suivante forment ce que Jassem appelle l'UNITE RYTHMIQUE TOTALE (*Total Rhythm Unit* ou *TRU*) ou COMPLETE (*Complete Rhythm Unit* ou *CRU* ; Jassem 1995). Un tel découpage en NRU, ANA et TRU est illustré dans la Figure 6 ci-après.

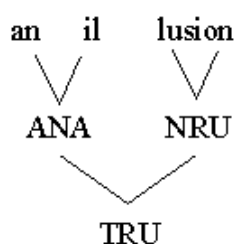


Figure 6. Découpage rythmique de l'énoncé *an illusion* selon le modèle de Jassem

Une TRU (ou CRU) peut être co-extensive avec un seul mot, mais elle correspond souvent à plus d'un mot lorsque l'ANA correspond à un mot isolé (comme *an illusion* où l'article indéfini est ANA) ou un mot peut-être découpé en plusieurs TRU (ou CRU) lorsque celui-ci possède plusieurs accents (comme *comprehensibility* composé de trois TRU puisqu'il possède trois types d'accent).

Pour un tempo donné, la longueur d'une NRU dépend du nombre de syllabes qui la compose : plus le nombre de syllabes augmente, plus la longueur de la NRU augmente, mais pas proportionnellement. En effet, une NRU de deux syllabes sera plus longue qu'une NRU d'une seule syllabe, mais pas deux fois plus longue ; c'est ce que Jassem entend par isochronie faible qu'il intègre à son modèle. Il l'illustre de la manière suivante (1983 : 427) :

NRU d'une syllabe : ———
 NRU de deux syllabes : ——— ———
 NRU de trois syllabes : ——— ——— ———

Dans cette illustration, les traits horizontaux représentent les durées relatives des syllabes contenues dans les NRU.

A l'intérieur d'une NRU polysyllabique, chaque syllabe a tendance à avoir une durée similaire : la longueur totale d'une NRU polysyllabique est approximativement divisée de manière égale par le nombre de syllabes. Les NRU sont régies par le principe d'isochronie faible (Hill, Jassem et Witten 1978 ; Hill, Witten et Jassem 1978) qui suppose une durée quasi similaire des NRU, quel que soit le nombre de syllabes contenu dans cette unité, ceci étant une tendance plutôt qu'une régularité stricte (cf. Chapitre 4). En revanche, les syllabes comprises dans une anacrouse ne sont pas régies par la notion d'isochronie faible ; elles ont comme caractéristique d'être aussi courtes que possible, plus précisément, aussi courtes qu'il est physiologiquement possible, dans la limite des contraintes articulatoires. Par conséquent, la longueur de l'anacrouse a tendance à être proportionnelle au nombre de syllabes qui la compose.

La NRU débute avec une syllabe accentuée et se termine avec la dernière syllabe inaccentuée du mot auquel appartient la syllabe accentuée. Toutefois, la NRU peut également englober des enclitiques, à savoir généralement des pronoms objet, sujet ou objet indirect dépendant du verbe précédent. Dans ces cas spécifiques, les enclitiques sont inclus dans la NRU et les syllabes qui les composent partagent les caractéristiques temporelles des autres syllabes appartenant à la NRU (et pas celles des ANA). Ainsi, dans un énoncé comme *pack it, it* est enclitique et forme avec le verbe une NRU dissyllabique.

Jassem propose une unité supplémentaire pour représenter les patrons mélodiques de l'anglais : l'anacrouse est cette fois regroupée avec l'unité rythmique précédente pour former l'UNITE TONALE (*Tonal Unit*), correspondant au pied abercrombien défini plus loin (section 3.2.3). Cette unité est définie par Jassem comme le domaine des contours mélodiques, elle est reprise dans le cadre de la modélisation des patrons mélodiques de l'anglais par Hirst (1998).

Le découpage et la répartition des durées syllabiques selon le modèle de Jassem permettent de distinguer certains énoncés phonémiquement identiques ; Jassem cite les

énoncés *summer dresses* et *some addresses* dont les regroupements en NRU et ANA sont présentés dans la Figure 7 ci-dessous.

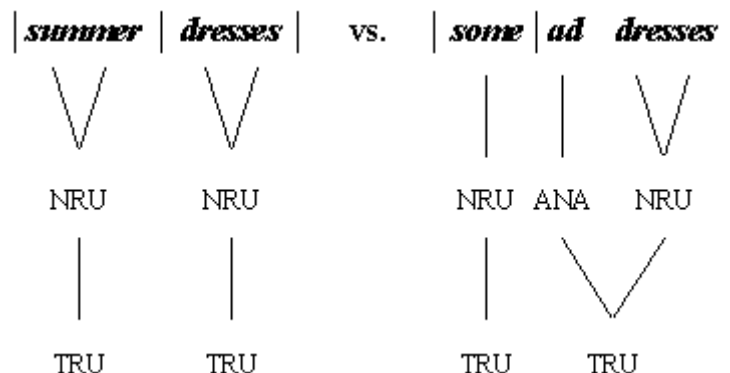


Figure 7. Représentation des énoncés *summer dresses* et *some addresses* selon le modèle de Jassem (1951)

La différence entre les deux énoncés est purement rythmique : dans les deux cas, la composition phonématique est identique (/sʌmədresɪz/), mais le découpage en NRU et en ANA permet de différencier les patrons de durée et donc de distinguer ces deux énoncés.

Dans *summer dresses*, chacun des deux mots est accentué sur la première syllabe, on a donc la structure ‘NRU + NRU’, les durées syllabiques relatives pouvant être représentées | — — | | — — | (chaque trait horizontal représente la durée relative des syllabes). Dans *some addresses*, la structure est ‘NRU + ANA + NRU’, elle est représentée | — — | · — — | (le point représente une anacrouse). Dans ce cas, la première NRU (/sʌm/) est monosyllabique et donc longue, alors que le deuxième groupe est composé d’une ANA (/əd/) suivie d’une NRU de deux syllabes (/dre.sɪz/) : l’anacrouse est très courte et les deux syllabes de la NRU se partagent sa durée globale. Par conséquent, /ʌ/ est plus long dans le deuxième énoncé, /ə/ plus long dans le premier énoncé, /e/ et /ɪ/ approximativement de durée égale dans les deux cas.

Jassem (1949) avait à ce sujet proposé de représenter les regroupements rythmiques grâce à la transcription IPA. Les jointures rythmiques sont représentées par un espace séparant les unités rythmiques totales, ceci donnant les deux transcriptions suivantes : /'sʌm ə 'dresɪz/ (*summer dresses*) et /'sʌm ə 'dresɪz/ (*some addresses*).

Jassem (1952 : 33) précise que la NRU est *doublement* caractérisée par la notion d'isochronie faible : d'un point de vue externe (les différentes NRU d'un énoncé ont tendance à avoir une durée similaire) et interne (durée similaire des syllabes composant la NRU). Cette organisation temporelle des NRU résulte de la rythmicité de la parole et de la nécessité de compresser un ensemble de phones pour qu'ils correspondent à la durée d'une unité rythmique (Jassem 1995 : 16).

3.2.2.2. Approche similaire proposée par O'Connor

O'Connor (1967 ; 1980) et O'Connor et Arnold (1973) décrivent un modèle qui présente des similitudes avec celui de Jassem. Le modèle se fonde sur les syllabes accentuées, autour desquelles les autres syllabes (inaccentuées) s'organisent ; la similitude réside principalement dans la distinction entre les deux types de syllabes inaccentuées (correspondant aux anacrouses et aux syllabes inaccentuées appartenant à la NRU précédente) et du regroupement rythmique qui s'ensuit. En effet, les énoncés sont découpés en deux types d'unités : les syllabes inaccentuées pré-accentuelles (parfaitement équivalentes à l'anacrouse) et le groupe accentuel (*stress group*), composé d'une syllabe accentuée éventuellement suivie d'une ou plusieurs syllabes inaccentuées appartenant au même groupe de mots (*word group*) représenté par un mot unique ou un mot suivi d'enclitique(s).

Les syllabes inaccentuées pré-accentuelles sont produites aussi rapidement que possible alors que les syllabes inaccentuées post-accentuelles appartenant au groupe accentuel se partagent équitablement la durée du groupe accentuel avec la syllabe accentuée qui les précède. Le groupe accentuel précédé d'une ou plusieurs syllabes inaccentuées pré-accentuelles forment une unité rythmique (*rhythm unit*), parfaitement équivalente à la TRU dans l'approche de Jassem. Le groupe accentuel est également caractérisé par ce que Jassem a qualifié d'isochronie interne et externe, avec toutefois une différence puisque O'Connor définit l'isochronie en tant que principe strict.

Le regroupement des différentes syllabes inaccentuées se fait selon quatre principes (1967 : 126) :

- Toute syllabe inaccentuée en début de groupe de mot se regroupe avec le groupe accentuel suivant (et est prononcée aussi rapidement que possible)

- Toute syllabe inaccentuée se trouvant dans le même mot que la syllabe accentuée appartient au même groupe rythmique
- Toute syllabe inaccentuée liée grammaticalement avec le mot précédant mais formant un mot isolé (typiquement un enclitique) appartient au même groupe rythmique
- En cas de doute, les mots inaccentués se regroupent avec le groupe rythmique précédent.

Notons que Knowles (1987) propose un regroupement similaire mais au niveau du mot. Dans un mot donné, les syllabes inaccentuées peuvent être pré-accentuelles (*leading syllables*) ou post-accentuelles (*trailing syllables*) ; les premières sont réduites et particulièrement courtes (voire supprimées comme dans le mot *police*) alors que les deuxièmes sont plus allongées (en fonction de leur poids).

O'Connor, dans une édition ultérieure (1980), justifie le regroupement adopté dans son modèle sur la base de la vitesse à laquelle les syllabes inaccentuées sont prononcées. Les syllabes inaccentuées précédant une syllabe accentuée sont prononcées particulièrement rapidement, comme les deux premières syllabes de l'énoncé *it was better* ; en revanche, la syllabe post-accentuelle *-tter* a une durée similaire à celle de la syllabe accentuée *be-*. Ce regroupement sur la base du débit est relativement aisé dans un tel exemple, mais devient insoluble dans des énoncés plus complexes. Même si le découpage est identique à celui de 1967, il est plus délicat de fonder celui-ci sur le débit. Il est plus souhaitable de commencer par découper les énoncés en unités puis de faire une prédiction de leurs caractéristiques temporelles plutôt que de partir de l'information de durée (ou plutôt de rapidité d'énonciation) et de déduire le regroupement en unités.

Ce patron de durées syllabiques correspondant au modèle de O'Connor est identique à celui de Jassem, à l'exception peut-être de la définition de l'isochronie qui est stricte pour O'Connor. Il est toutefois intéressant de remarquer que le modèle de O'Connor est postérieur à celui de Jassem et que O'Connor ne fait en aucun cas référence au modèle de Jassem.

3.2.3. *Le pied accentuel*

Nous venons de présenter le modèle proposé par Jassem dont le but est de représenter l'organisation rythmique et temporelle de l'anglais au travers des unités rythmiques étroites et des anacrouses. Postérieure à ce modèle, s'est développée l'approche d'Abercrombie (1964 ; 1967) puis de Halliday (1967 ; 1970), devenue la référence dans la majorité des études sur le rythme de l'anglais. En effet, les références à la structuration rythmique de l'anglais s'appuient de manière quasi systématique sur le modèle d'Abercrombie, souvent sans connaître l'existence du modèle de Jassem.

3.2.3.1. *Approche d'Abercrombie*

Pour Abercrombie (1964 ; 1967), il existe des relations de quantité entre les syllabes en anglais et ces relations sont fondamentales en phonétique et en phonologie ; il semble donc impossible de faire une description prosodique de l'anglais sans prendre en compte la quantité des syllabes (principe ensuite repris par Knowles 1987). Le pied (accentuel) est l'unité dans laquelle la syllabe fonctionne rythmiquement ; les énoncés sont divisés en pieds par les battements (*beats*) isochrones des accents, leur durée étant approximativement égale. Chaque pied débute avec une syllabe accentuée et se termine à la syllabe accentuée suivante, mais sans l'inclure. Notons que cette approche est fondée sur la terminologie des études sur le rythme en musique puisque le découpage se fait en fonction des temps forts.

Abercrombie cite l'exemple suivant, dans lequel les barres verticales représentent les frontières de pied (1964 : 217) :

| *This is the* | *house that* | *Jack* | *built* |

La quantité de chaque syllabe est proportionnelle à la longueur totale du pied dans lequel la syllabe se trouve ; elle est également relative à la quantité de toute autre syllabe dans le pied. Abercrombie remarque que la quantité de la syllabe ne dépend pas directement de la quantité de la voyelle, ni de l'accent ; de même que le pied ne dépend pas systématiquement des frontières de mot (contrairement au découpage selon le modèle de Jassem).

Abercrombie distingue trois types de pied dissyllabique, représenté grâce aux symboles suivants : '∪' représente une syllabe courte, '∩' une syllabe moyenne et '—' une syllabe longue. Les trois types de pied sont les suivants (1964 : 219) :

- Type A : pied ‘court-long’. Il s’agit d’une syllabe courte suivie d’une syllabe longue, noté $\cup -$. On a par exemple | *shilling* |, | *never* |, | *atom* |, | *cuckoo* |.
- Type B : pied ‘égal-égal’. Il s’agit d’un pied composé de syllabes de longueur égale, noté $\cap \cap$. On a par exemple | *greater* |, | *firmly* |, | *centre* |, | *matches* |.
- Type C : pied ‘long-court’. Il s’agit d’un pied composé d’une syllabe longue suivie d’une syllabe courte, noté $- \cup$. On a par exemple | *Grey to* |, | *tea for* |, *per* | *haps I* | *did*. Dans ce type de pied, la quantité $- \cup$ dépend de la présence d’une frontière de mot dans le pied. La structure phonématique des syllabes ne détermine pas leur quantité, contrairement aux types A et B.

Ces trois catégories de pied sont caractérisées par Abercrombie au travers des structures phonématiques suivantes (1964 : 220) :

- (1) (C)V₁CV(C)
- (2) (C)VCC(C)V(C)
- (3) (C)V₂(C)V(C)

C et V représentent respectivement n’importe quelle consonne et voyelle, (C) une consonne facultative, V₁ une voyelle courte et V₂ une voyelle longue ou une diphtongue. (1) correspond plutôt au pied de type A, alors que (2) et (3) au pied de type B, autrement dit V₁ entraîne des quantités de type ‘ $\cup -$ ’ (court-long) et V₂ de type ‘ $\cap \cap$ ’ (égal-égal).

Abercrombie remarque que la présence d’une frontière de mot ne représente pas toujours les quantités de syllabes auxquelles on pourrait s’attendre. Par exemple, | *stop her* | est de type A alors que | *take it* | est de type B : les structures phonématiques ont déterminé les quantités syllabiques comme si les frontières de mots étaient absentes.

Ces deux exemples posent la question des enclitiques, ils se présentent dans les cas où un verbe est suivi d’un pronom objet, sujet ou objet indirect appartenant au même pied. Ces enclitiques forment avec la syllabe accentuée précédente (et toute autre syllabe inaccentuée suivante facultative) le pied abercrombien. Il n’est donc pas nécessaire de les distinguer des autres syllabes inaccentuées puisqu’ils se comportent rythmiquement de manière identique.

3.2.3.2. Approche de Halliday

Tout comme la structure du pied décrite par Abercrombie, le pied dans l'approche de Halliday (1970) contient une ou plusieurs syllabes, la première syllabe étant toujours accentuée. Halliday préfère toutefois parler de syllabe "saillante" plutôt que de syllabe accentuée étant donné l'emploi quelquefois vague de la notion d'accentuation. Par conséquent, le pied comprend généralement une syllabe saillante ou une syllabe saillante suivie d'une ou plusieurs syllabes non saillantes.

Nous insistons sur le mot "généralement", car il existe également des pieds qui commencent par ce que Halliday appelle un battement silencieux (*silent beat*) ou ictus silencieux (*silent ictus*) pour Abercrombie (1967). Chaque pied doit logiquement débiter avec une syllabe saillante, mais il existe des syllabes inaccentuées en début d'énoncé qui ne sont précédées d'aucune syllabe accentuée, on dit alors qu'il existe un battement silencieux avant la syllabe inaccentuée, formant ainsi un pied. Halliday cite l'exemple suivant :

// ^ it's / Arthur //

it's est non saillant, il est par conséquent précédé d'un battement silencieux (représenté par un caret), permettant à cet ensemble de former un pied. Ainsi, un pied peut commencer par un battement silencieux sans pour autant perturber le rythme.

Comme nous l'avons remarqué plus tôt, Abercrombie, ainsi que Halliday, fondent leur description sur les approches généralement adoptées pour rendre compte du rythme en musique, la terminologie est ainsi reprise de ces différentes études. Tout comme Abercrombie, Halliday décrit les pieds comme étant de durée relativement stable quelle que soit leur taille (isochronie faible).

3.2.4. Comparaison des deux modèles

La différence essentielle entre ces deux types d'approche réside dans le traitement des syllabes inaccentuées et donc dans la délimitation des unités rythmiques.

- Pour Jassem, les syllabes inaccentuées sont soit pré-accentuelles (anacrouses), soit post-accentuelles (elles appartiennent à la NRU), les deux possédant des différences rythmiques et temporelles fondamentales. Les syllabes en anacrouse sont des proclitiques, les syllabes inaccentuées post-accentuelles des enclitiques.

- Pour Abercrombie et Halliday, toutes les syllabes inaccentuées suivent la syllabe accentuée et sont englobées dans le pied, sauf les syllabes inaccentuées en début d'énoncé qui sont précédées d'un battement (ou ictus) silencieux pour Halliday.

Cette différence de regroupement engendre des prédictions temporelles totalement divergentes au niveau des syllabes accentuées et inaccentuées.

Si on reprend les énoncés *summer dresses* et *some addresses* (Figure 8 et Figure 9 ci-après), les regroupements en unités rythmiques sont différents.

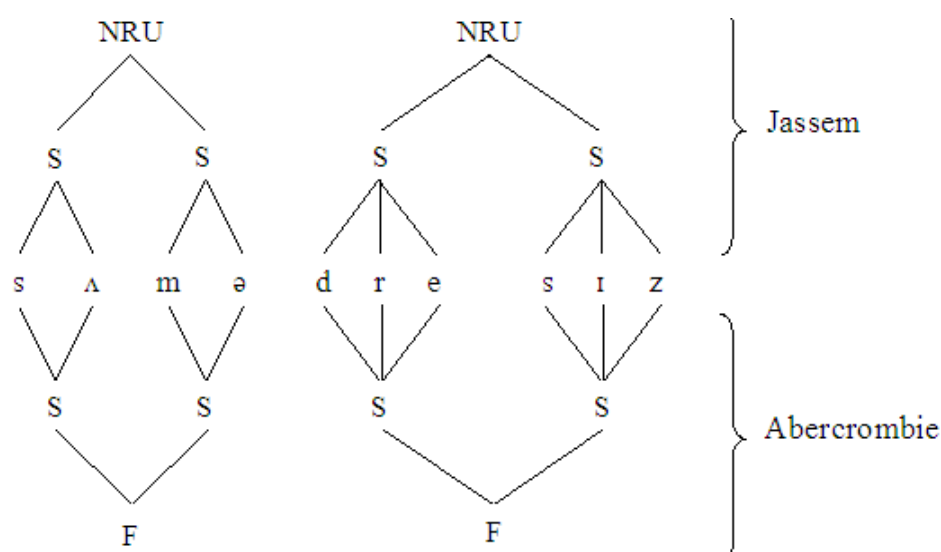


Figure 8. Représentation de l'énoncé *summer dresses* selon les deux modèles rythmiques ('S' correspond à 'Syllable' et 'F' à 'Foot')

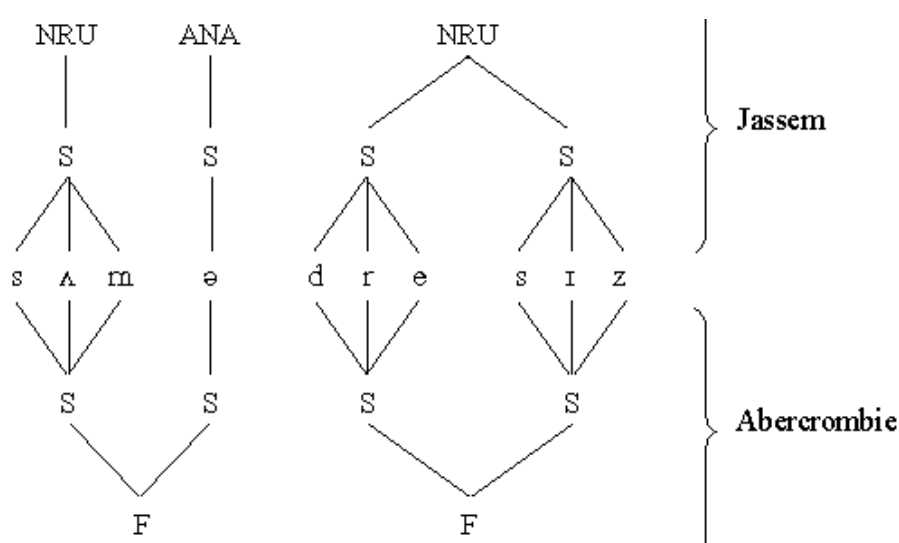


Figure 9. Représentation de l'énoncé *some addresses* selon les deux modèles supra-syllabiques

La première remarque que l'on peut faire à partir de ces deux figures est que la présence d'une frontière de mot influence le regroupement en unités rythmiques dans le modèle de Jassem, à l'instar du modèle d'Abercrombie. En revanche, si la présence d'une frontière de mot n'a pas d'impact sur la structure et l'étendue du pied accentuel, elle en a une sur la structure des syllabes : le rattachement de la consonne /m/ dépend de l'emplacement de la frontière de mot (après ou avant le schwa). Cette différence de quantité se reflète sur la catégorisation des pieds : type A pour *summer* et type C pour *some a-*.

En conséquence, il convient de préciser que le modèle de Jassem permet de différencier des énoncés du type *summer dresses* et *some addresses* de par la prise en compte des frontières de mots dans le regroupement rythmique contrairement au modèle d'Abercrombie. En revanche, s'il n'existe pas de différence entre ces deux énoncés au niveau de la taille du pied, il en existe une au niveau syllabique et cette différence se reflète sur le type de pied abercrombien.

Ce qui est tout particulièrement intéressant, ce sont les exemples tels que *player* et *play a* dans lesquels le regroupement syllabique est identique et seuls les regroupements rythmiques diffèrent (Figure 10 ci-après).

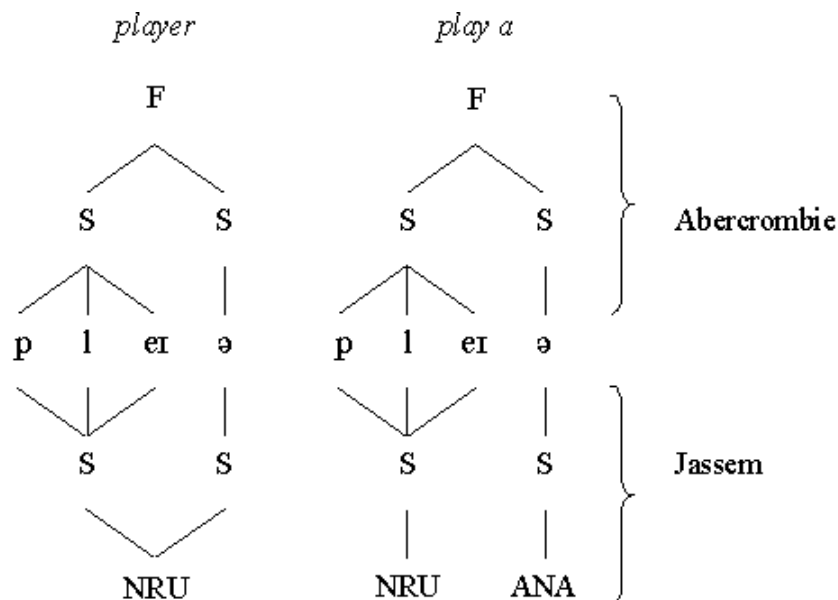


Figure 10. Regroupements syllabique et rythmique des énoncés *player* (1) et *play a* (2) selon l'approche d'Abercrombie (partie haute de la figure) et selon l'approche de Jassem (partie basse de la figure).

Dans ces énoncés, les structures syllabiques sont identiques, seule la structure des unités rythmiques diffèrent. Le modèle d'Abercrombie ou de Halliday ne prévoit aucune différence (partie haute de la figure), alors que d'après le modèle de Jassem, il existe une différence fondamentale au niveau du regroupement rythmique de par la présence d'une frontière de mot dans l'énoncé *play a* (partie basse de la figure).

Citons enfin la série d'énoncés donnés par Jassem, Hill et Witten (1984) dont le but est d'illustrer la différence entre les deux types d'approche :

John's pleased
John was pleased
John would be pleased
John would have been pleased
John would have been extremely pleased

D'après Abercrombie et Halliday, la syllabe /dʒɒn/ devrait être de plus en plus courte pour s'ajuster au nombre croissant de syllabes inaccentuées dans le pied (de zéro à quatre). Pour Jassem, la longueur de /dʒɒn/ n'est pas affectée par la présence (ou absence) de syllabes inaccentuées suivantes, car la TRU finit à la fin de 'John'. La Figure 11 ci-après illustre les différences de structuration rythmique et de prédiction des durées syllabiques (relatives).

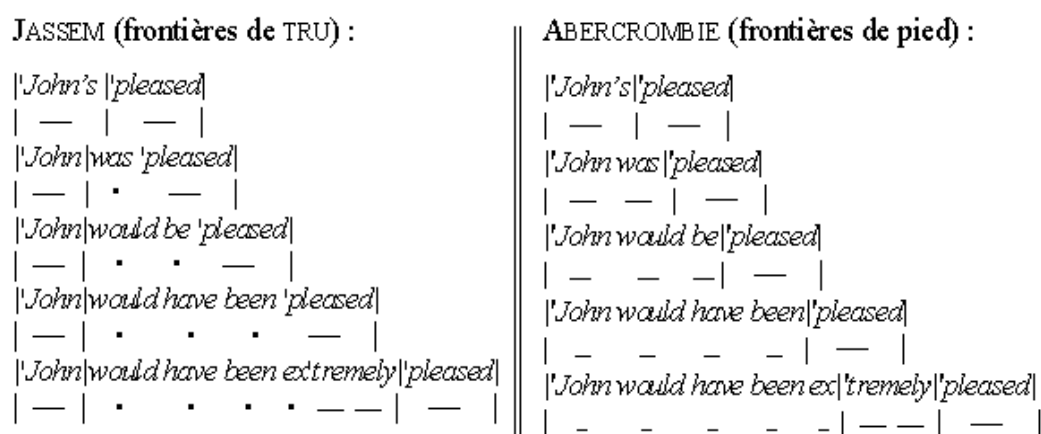


Figure 11. Représentation d'énoncés identiques en fonction du modèle de Jassem (colonne de gauche) et de celui d'Abercrombie (colonne de droite)

La transcription donnée dans la Figure 11 illustre les différences de répartition des durées syllabiques en fonction de l'unité à laquelle elles appartiennent (ANA, NRU ou pied accentuel). La longueur de chacun des traits représente la durée relative de chaque

syllabe, adaptée en fonction de sa durée prédite ; les points représentent les ANA. Nous faisons ici référence aux durées relatives des différentes syllabes et non aux durées brutes ; en effet, les durées brutes de syllabe prennent en compte des facteurs supplémentaires tels que les caractéristiques intrinsèques des consonnes et des voyelles, ainsi que de l'influence du contexte phonémique. Cette représentation nous permet de visualiser l'écart de prédiction des durées syllabiques entre les deux modèles.

Cette différence est observée par Jassem, Hill et Witten (1984) : les auteurs testent les deux types d'unité afin de déterminer lequel est le plus approprié pour rendre compte du rythme de l'anglais britannique. Ils observent que la durée moyenne d'une NRU représente plus de deux fois celle d'une ANA et que la durée moyenne des phones dans les ANA équivaut à 0.7 fois la durée moyenne des phones d'une NRU. Il ressort également que la vitesse d'élocution des phonèmes est environ 1.43 plus rapide pour les ANA que pour les syllabes dans une NRU. Une deuxième expérience montre qu'il existe très peu d'isochronie dans l'ANA, alors qu'il semble y avoir une tendance distincte vers un patron isochrone dans la NRU. Un certain degré d'isochronie est également présent dans le pied abercrombien, mais les résultats confirment que les syllabes inaccentuées pré- et post-accentuelles doivent être traitées différemment. Les auteurs rejettent par conséquent la structure du pied abercrombien car il empêche de distinguer temporellement les deux types de syllabes inaccentuées, alors que la différence ressort comme fortement significative statistiquement.

Une dernière remarque peut être faite concernant l'effet de l'accent ; dans les deux modèles, les syllabes comprises dans la NRU ou dans le pied accentuel se partagent équitablement la durée de l'unité rythmique, quel que soit leur statut accentuel. Cette caractéristique est surprenante étant donné qu'il est largement reconnu que la présence d'un accent affecte la durée de la syllabe sur laquelle il tombe et des phonèmes qui la composent (entre autres Lehiste 1970 ; Klatt 1987 ; Fant et Kruckenberg 1993 ; Fant, Kruckenberg et Nord 1991a ; cf. section 3.4). On peut alors se demander comment se fait la différence au niveau perceptif et productif entre les syllabes accentuées et inaccentuées dans les unités rythmiques, puisque les syllabes accentuées ont une durée relative similaire à celle des syllabes inaccentuées suivantes dans la NRU et le pied accentuel. Jassem, Hill et Witten (1984) expliquent que la perception des accents découle de l'organisation temporelle spécifique des énoncés : les syllabes accentuées sont perçues comme tel grâce à la répartition des durées syllabiques. Dans le cas où

toutes les syllabes d'un énoncé ont une durée équivalente, on reconnaît les accents grâce à des patrons intériorisés (mémorisés). Ainsi, un énoncé tel que *dinner's ready* représenté | — — | — — | sera perçu comme deux NRU de deux syllabes, et non quatre NRU d'une seule syllabe, car dans ce cas-là, les durées syllabiques relatives seraient plus longues (mais pas deux fois plus longues).

Les modes de fonctionnement caractérisant chacun des deux modèles convergent quant à la méthode adoptée pour prédire l'organisation temporelle de la parole ; en effet, ils cherchent à prédire l'organisation temporelle de la parole en fonction du rythme plutôt que de fonder leur prédiction sur l'influence de la syntaxe et des caractéristiques intrinsèques des phonèmes (cf. citation de Cutler et Isard 1980 citée dans le Chapitre 1). Malgré cette méthodologie commune, ils ne font pas la même prédiction du fait qu'ils ne découpent pas les énoncés selon le même modèle ; celui de Jassem distingue les syllabes inaccentuées pré- et post-accentuelles supposées différer dans leurs caractéristiques temporelles alors qu'une telle distinction n'est pas appliquée dans celui d'Abercrombie ou de Halliday. Notons que malgré cette différence fondamentale, les deux modèles ont comme point commun le concept d'isochronie faible, dans le domaine de la NRU et du pied accentuel.

3.3. Comparaison et perspectives des deux modèles

Cette section est consacrée à la comparaison de ces deux modèles avec les différents résultats que l'on trouve dans la littérature.

3.3.1. Traitement différentiel des syllabes inaccentuées

La différence fondamentale entre les deux modèles d'Abercrombie/ Halliday et de Jassem réside dans le traitement différentiel des syllabes inaccentuées pré- et post-accentuelles.

Dans les deux modèles, l'ajout de syllabes inaccentuées dans le pied et la NRU engendre un degré de compression des syllabes contenues dans cette unité rythmique (accentuées et inaccentuées). Cette remarque fait l'unanimité dans la littérature, mais la différence fondamentale est de savoir si ce sont toutes les syllabes inaccentuées suivantes qui influencent la durée de la syllabe accentuée (pied accentuel) ou seulement

les syllabes inaccentuées appartenant à la NRU. Autrement dit, la problématique qui sous-tend cette divergence est celle du rôle des frontières de mot dans le regroupement rythmique.

3.3.1.1. Syllabes inaccentuées post-accentuelles

D'après Lehiste (1972), l'addition de syllabes à droite comme à gauche de la syllabe accentuée crée un raccourcissement, mais bien plus marqué à droite, d'où le besoin d'élaborer un modèle *asymétrique* de raccourcissement polysyllabique. Le modèle de Lindblom et Rapp (1973) prédit le raccourcissement d'une syllabe accentuée par les syllabes inaccentuées précédentes et suivantes, à condition qu'elles appartiennent toutes au même mot que la syllabe accentuée ; il s'exerce ainsi un phénomène de compensation au sein du mot. Il existe par conséquent deux types de contrainte dans ce modèle : une compensation proactive et une compensation rétroactive. La première fait référence au raccourcissement proportionnel d'un élément en fonction du nombre de syllabes qui suit cet élément dans l'unité ; la deuxième représente le raccourcissement proportionnel d'un élément en fonction du nombre de syllabes qui précède dans l'unité. Les auteurs précisent que les effets proactifs sont beaucoup plus marqués au niveau de la syllabe.

Les travaux de Huggins (1975) visent en partie à observer l'effet du nombre de syllabes inaccentuées sur la syllabe accentuée *précédente* ainsi que l'effet de la présence d'une frontière de mot. Dans l'énoncé

‘*cheese(s) (a)bound(ed) (ab)out*’

la durée de la voyelle accentuée de *bound* est raccourcie par les syllabes inaccentuées suivantes qu'elles appartiennent au même mot ou non, l'effet des syllabes inaccentuées à l'intérieur du mot étant toutefois plus marqué. Il existe cependant une différence d'effet entre les deux mots *cheese* et *bound* que Huggins explique d'un point de vue syntaxique : les mots *cheese* et *bound* sont séparés par une frontière syntaxique majeure séparant sujet et prédicat, contrairement aux deux derniers mots. Ces résultats suggèrent par conséquent la pertinence d'un modèle fondé sur le pied métrique, tout en considérant les frontières syntaxiques mineures et majeures.

Fowler (1977) réplique le même type d'expérience que Huggins avec six séries de six énoncés, dans lesquels le nombre de syllabes inaccentuées entre les syllabes accentuées varie de zéro à deux et dans lesquels l'emplacement de la frontière de mot est variable. Une des séries de stimuli est la suivante, avec les patrons accentuels respectifs entre parenthèses (A : accentué, I : inaccentué et # : frontière de mot) :

*The **fact** started the argument (AA)*

*The **fact** has started the argument (A#IA)*

*The **factor** started the argument (AI#A)*

*The **fact** has restarted the argument (A#IIA)*

*The **factor** restarted the argument (AI#IA)*

*The **factory** started the argument (AII#A)*

Fowler observe les résultats suivants :

1. Dans **AA**, la durée de la première voyelle accentuée est plus longue que toute autre voyelle accentuée correspondante dans un autre patron de manière significative.
2. L'emplacement de la frontière de mot ne joue pas un rôle capital dans les patrons **A#IA** et **AI#A**.
3. La première voyelle accentuée dans **A#IIA** est nettement plus longue que dans **AII#A**.

Le résultat 1. est en accord avec les deux types d'approches que nous avons vues puisque dans les deux cas, la syllabe accentuée d'un pied monosyllabique est plus longue que la syllabe accentuée d'un pied polysyllabique. Elle est également en accord avec Bolinger (1963 ; 1981) : une voyelle accentuée suivie d'une autre voyelle accentuée sera allongée. Il ressort de 2. que la frontière de mot n'influence pas les patrons **A#IA** et **AI#A**, ce qui va dans le sens du modèle d'Abercrombie (au niveau du regroupement rythmique), mais pas celui de Jassem. En revanche, 3. confirme le modèle de Jassem, pas celui d'Abercrombie. Notons ici que la présence d'une frontière de mot n'est pas importante lorsque la syllabe accentuée est suivie d'une seule syllabe inaccentuée, alors qu'elle l'est lorsque la syllabe accentuée est suivie de deux syllabes inaccentuées.

Globalement, il ressort de l'étude de Fowler qu'une syllabe accentuée est raccourcie par l'ajout de syllabes inaccentuées entre cette syllabe et la syllabe accentuée suivante, mais ce raccourcissement est atténué par la présence d'une frontière de mot, nettement plus marqué en cas de deux syllabes inaccentuées. Ces résultats vont dans le sens de

ceux obtenus par Huggins et présentent donc une hypothèse intermédiaire entre les deux modèles que nous comparons, à savoir que la frontière de mot joue un rôle dans la répartition des durées syllabiques, toutefois la présence d'une frontière de mot atténue mais ne bloque pas les différents effets de raccourcissement.

Les résultats obtenus par Rakerd, Sennet et Fowler (1992) testent la validité de la structure de pied abercrombien dans lequel les syllabes inaccentuées se regroupent avec la syllabe accentuée précédente : il n'est donc pas ici question de symétrie ou d'asymétrie de l'effet des syllabes inaccentuées puisque les syllabes inaccentuées sont nécessairement post-accentuelles. Les auteurs confirment que l'ajout de syllabes inaccentuées dans le pied abercrombien raccourcit la syllabe accentuée de ce même pied mais pas les syllabes accentuées au-delà de la frontière de pied. Ces résultats confirment la validité de la structure de pied abercrombien, bien que cette étude ne cherche pas à détecter quelque différence entre les diverses syllabes inaccentuées comme le fait le modèle de Jassem, autrement dit, elle ne cherche pas à tester l'influence de la présence d'une frontière de mot.

Campbell (1996) confirme également cette tendance. A l'intérieur d'un pied accentuel, la présence et le nombre de syllabes inaccentuées influent sur la durée de celles-ci ainsi que sur la durée de la syllabe accentuée précédente, mais pas sur la durée de la syllabe accentuée suivante, et ceci même lorsque les syllabes en position finale d'unité intonative ne sont pas prises en compte. De plus, en cas de succession de deux syllabes accentuées, les deux sont allongées et les syllabes inaccentuées en début d'unité intonative (typiquement des anacrouses) ont tendance à être très raccourcies. Il existe donc une réduction significative de la durée du pied en fonction du nombre de syllabes inaccentuées suivant la syllabe accentuée mais pas en fonction du nombre de syllabes inaccentuées précédant la syllabe accentuée. Ces résultats confirment par conséquent le regroupement des syllabes en pieds avec tête à gauche (1996 : 228). L'allongement des syllabes accentuées adjacentes suggère une compensation pour aboutir à une durée d'intervalle similaire à celle que l'on aurait en cas de syllabes inaccentuées intermédiaires ; nous reviendrons sur ce type de régularité dans le chapitre suivant.

3.3.1.2. Syllabes inaccentuées pré-accentuelles

Fowler (1977) conduit une expérience similaire à celle citée plus haut afin d'étudier l'influence des syllabes inaccentuées sur la syllabe accentuée suivante. Elle observe ici l'effet de l'ajout d'un préfixe au mot contenant la syllabe accentuée cible (*'the fact **restarted** the argument'*), l'effet de l'ajout d'une syllabe inaccentuée sous forme de mot (*'David **has** signed the paper'*) et également l'effet de l'ajout d'un suffixe au mot précédant (*'the factor **started** the argument'*). Fowler précise que les effets observés dans les résultats sont faibles mais fiables ; il ressort que la voyelle accentuée semble être plus longue lorsqu'elle est précédée de deux syllabes inaccentuées (IIA) que lorsqu'elle est précédée d'une autre syllabe accentuée (AA). Ces résultats suggèrent qu'une syllabe inaccentuée influe à la fois sur la syllabe accentuée précédente et sur la syllabe accentuée suivante. Rappelons toutefois que ces effets sont faiblement marqués.

Fowler conclut que les contraintes temporelles sont *bidirectionnelles* car elles sont à la fois proactives et rétroactives : une voyelle accentuée est raccourcie en fonction du nombre de syllabes inaccentuées qui précède et qui suit. Toutefois, les effets proactifs de compression sont plus marqués que les effets rétroactifs. Fowler (1981) explique que cette différence de raccourcissement reflète une cohésion coarticulatoire asymétrique entre la syllabe accentuée et la ou les syllabe(s) inaccentuée(s) environnantes. Cette caractéristique du modèle de Fowler remet en question les modèles de Jassem et d'Abercrombie, bien qu'elle confirme la nécessité du traitement différentiel des syllabes inaccentuées à droite et à gauche de la syllabe accentuée.

Les travaux de Turk et White (1999) et de Turk et Shattuck-Hufnagel (2000) présentent l'étendue des effets de l'allongement accentuel sur la syllabe accentuée, ainsi que sur les syllabes inaccentuées à sa droite et à sa gauche. Le domaine d'allongement accentuel débute avec l'attaque de la syllabe accentuée et s'étend jusqu'à la fin du mot (domaine correspondant par conséquent à la NRU). Dans ce domaine, les auteurs observent que les syllabes inaccentuées à gauche sont nettement plus courtes que les syllabes inaccentuées post-accentuelles. De plus, les auteurs observent que l'allongement des syllabes inaccentuées à droite est relativement proche de celui de la syllabe accentuée dans les mots de deux et trois syllabes (contrairement à celui des syllabes inaccentuées pré-accentuelles). Ces résultats correspondent à la prédiction faite selon le modèle de Jassem puisque le domaine d'allongement est identique à la NRU : les syllabes dans la NRU se partagent la durée totale de la NRU.

Turk et Shattuck-Hufnagel (2000) expliquent les différents allongements ou raccourcissements au niveau du mot par un phénomène de raccourcissement polysyllabique et d'égalisation du ratio entre les syllabes (*syllable-ratio equalization*). Ce deuxième correspond à un phénomène d'égalisation de durée syllabique à l'intérieur de l'unité (le mot) et permet d'expliquer que, dans un stimulus tel que '*tuna#choir*', les durées relatives de /tju:n/ et /ə/ sont plus proches que celles des mêmes syllabes dans les stimuli '*tune#a#choir*' et '*tune#acquire*' (prédiction identique à celle du modèle de Jassem). Le modèle d'Abercrombie ne prédit aucune différence de regroupement rythmique, bien qu'ils diffèrent en terme de regroupement syllabique (/tju:n.ə/ vs /tju:n.ə/).

Nous venons de voir qu'il existe un certain nombre d'études qui montrent que l'effet de la syllabe inaccentuée à droite est plus marqué, mais que, pour certaines études, la syllabe inaccentuée à gauche exerce également un effet sur la syllabe accentuée. Cette asymétrie est absente des modèles de Jassem et d'Abercrombie, bien que celui de Jassem traite les deux types de syllabes inaccentuées de manière distincte.

3.3.2. Rôle et fonction du mot

Il convient à ce stade de se poser la question du statut et de la fonction du mot dans l'organisation rythmique de l'anglais puisqu'il apparaît dans plusieurs études sans pour autant appartenir à la hiérarchie classique des constituants prosodiques de l'anglais.

Dans son étude, Pike (1945) fait référence au mot puisque le début et la fin d'une unité rythmique correspondent presque toujours au début et à la fin d'un mot. Pike prend par conséquent le mot comme unité inférieure, dont les frontières délimitent les unités rythmiques. Comme nous l'avons précisé précédemment, d'autres travaux font référence au mot, notamment Jassem, Lindblom et Rapp, Huggins, Giegerich, Turk et White, et Turk et Shattuck-Hufnagel.

La définition de la NRU (Jassem) dépend des frontières de mots mais Jassem précise que le mot n'a aucunement le statut d'unité phonétique ou phonologique : c'est une unité purement grammaticale. Toutefois, au niveau de l'organisation temporelle et accentuelle, certaines relations apparaissent entre les mots en tant que constituants grammaticaux et éléments phonologiques, ceci expliquant la référence au mot dans le

modèle. Même si le mot n'est pas nécessairement une unité sémantique, il n'en reste pas moins une unité porteuse de sens (Jassem 1952 : 34). En revanche, pour Abercrombie, les frontières de mot n'influencent pas le regroupement rythmique en pieds accentuels (bien qu'elles restent déterminantes pour le découpage syllabique).

Coker, Umeda et Brownman (1973) remarquent que les mots fortement accentués sont bien distincts dans la chaîne parlée, ce qui indique que les mots jouent bien un rôle dans la perception des énoncés, ne serait-ce que pour distinguer les parties fortes des parties faibles (en terme d'accentuation). En revanche, les mots faibles (inaccentués) agissent tels des syllabes appartenant à un mot "global" polysyllabique rapidement prononcé (par opposition aux parties fortes). Autrement dit, un mot accentué se comporte comme tel, mais il se regroupe avec d'autres lorsqu'il n'est pas accentué. Cette remarque peut être mise en parallèle avec le modèle de Jassem puisque les mots accentués sont à la base de la NRU alors que les mots inaccentués vont se regrouper entre eux (ainsi qu'avec les initiales inaccentuées des mots accentués) pour former des anacrouses et ne seront donc pas considérés comme des unités individuelles.

De plus, un tel découpage permet le plus souvent de distinguer les mots lexicaux (typiquement une NRU ou une partie de NRU) et les mots grammaticaux (anacrouses). Il est effectivement peu probable que les syllabes appartenant à un mot lexical et les syllabes appartenant à un mot grammatical aient des caractéristiques temporelles équivalentes (Van Santen 1992b ; Campbell 2000). Le modèle de Jassem permet la distinction de ces deux catégories de mots au niveau temporel, contrairement au modèle d'Abercrombie.

Lindblom et Rapp (1973) observent les propriétés temporelles du suédois sur de la parole réitérée et les résultats de leurs diverses expériences montrent que la durée de chaque syllabe accentuée est raccourcie par les syllabes inaccentuées qui l'entourent à condition qu'elles fassent partie du même mot. Ainsi, la durée de la voyelle accentuée diminue de manière régulière lorsque la longueur du mot auquel elle appartient augmente, autrement dit, le modèle ne prédit aucun raccourcissement des segments accentués par les syllabes inaccentuées au-delà des frontières de mot.

Comme nous l'avons vu précédemment, Huggins (1975) tente de déterminer si la variable utilisée est le mot ou le pied métrique : si le mot est la variable en question, l'ajout de syllabes inaccentuées au-delà de la frontière de mot ne devrait pas influencer la durée de la syllabe accentuée. En revanche, si le pied est la variable utilisée, le

raccourcissement devrait être proportionnel au nombre de syllabes inaccentuées ajoutées dans le pied.

Les résultats de Huggins sont intermédiaires : ils suggèrent que la frontière de mot joue un rôle dans l'organisation rythmique (contrairement au pied abercrombien), mais que les syllabes inaccentuées influencent la durée des syllabes accentuées même au-delà des frontières de mot. Huggins remarque toutefois que cet effet est bloqué lorsque la frontière de mot coïncide avec une frontière syntaxique majeure. Ceci suggère une triple différenciation de l'effet des syllabes inaccentuées : les syllabes inaccentuées pré-accentuelles, les syllabes inaccentuées post-accentuelles appartenant au mot et post-accentuelles au-delà de la frontière de mot.

Turk et Sawusch (1997) remarquent que la structure rythmique est considérablement influencée par l'emplacement des frontières de mot et montrent que l'allongement accentuel s'étend sur la syllabe suivant la syllabe accentuée quand celle-ci appartient au même mot accentué. Turk et White (1999) posent comme frontières du domaine de l'allongement accentuel le bord gauche de la syllabe accentuée puis les bords gauche et droit des mots. Ainsi, leurs limites sont comparables à celles du modèle de Jassem : la première frontière isole l'anacrouse de la NRU, la deuxième frontière termine la NRU à droite et englobe l'anacrouse à gauche (ou une partie de l'anacrouse en cas de mot(s) inaccentué(s) précédent(s) et forme ainsi la TRU). Notons que certains effets d'allongement résiduel apparaissent dans les résultats de Turk et White, ce qui les amène à conclure que ces frontières de domaine d'allongement accentuel atténuent mais ne bloquent pas totalement l'étendue de l'allongement.

Il ressort de ces diverses remarques que le mot, même s'il ne représente pas une unité phonétique ou phonologique, est caractérisé par des effets de frontières qui rentrent en jeu lors de la structuration rythmique (Nakatani, O'Connor et Aston 1980). Par conséquent, les frontières de mot plus que le mot en tant qu'unité semblent jouer un jeu important dans la structuration rythmique et l'organisation temporelle de l'anglais. Toutefois, un manque d'homogénéité ressort des divers travaux dont le but est de tester la fonction des frontières de mot dans l'organisation temporelle des unités rythmiques en anglais. Certains travaux constatent l'absence d'effets spécifiques aux frontières de mot ; d'autres observent des effets intermédiaires, à savoir des effets de raccourcissement polysyllabique dû à l'ajout de syllabes inaccentuées atténués par les frontières de mot mais pas bloqués. Deux remarques sont à retenir :

- Il ne s'agit pas du mot en tant qu'unité phonétique ou phonologique qui rentre en jeu dans la structuration temporelle des unités rythmiques, mais plutôt de ses frontières.
- Les divers effets observés aux frontières de mot suggèrent que la présence d'une telle frontière ait une valeur d'atténuation plutôt que de blocage total des différents effets temporels en jeu.

3.3.3. *Confusion entre le niveau du pied et le niveau du mot*

Suite à ces diverses remarques quant au rôle du mot dans la structuration rythmique des énoncés, Zhang (1996) met en garde contre la confusion entre le niveau du pied et le niveau du mot. L'auteur cite à ce sujet les travaux de Lindblom et Rapp (1973) et remarque que les résultats obtenus lors de leurs expériences sur le suédois sont faussés car les deux auteurs ne distinguent pas le niveau du pied de celui du mot.

Zhang décide de tester sur de l'anglais les effets observés par Lindblom et Rapp en séparant distinctement les deux niveaux. La méthode utilisée comporte deux contextes différents :

- La structure de pied est intacte mais la longueur du mot est modifiée.
- La longueur de mot est conservée tout en variant la structure de pied.

Comme nous l'avons vu précédemment, Lindblom et Rapp font référence à des compensations rétroactives et proactives ; Zhang teste ces deux contraintes à la fois sur de la parole réitérée, puis sur une liste de mots isolés.

Dans le premier contexte, la structure de pied ne change pas, mais le nombre de syllabes dans le mot correspondant varie. Dans cette condition, Zhang ne retrouve aucune compensation proactive au niveau du mot (parole réitérée et liste de mots), autrement dit les durées syllabiques ne diminuent pas en compensation du nombre de syllabes qui suit dans le mot lorsque la structure de pied reste identique. Zhang tire la même conclusion au sujet de la compensation rétroactive : l'hypothèse n'est pas confirmée lorsque les deux niveaux sont distingués.

Dans le deuxième contexte, la taille du pied augmente mais la longueur du mot reste constante. La durée de la syllabe accentuée a tendance à décroître quand la longueur de pied augmente, mais pas la syllabe inaccentuée suivante. Autrement dit, il existe un effet évident de la longueur de pied sur la durée de la syllabe accentuée, mais il est difficile de dire si cet effet s'applique à toutes les syllabes du pied. Zhang explique qu'il

pourrait s'agir d'un procédé de raccourcissement *sélectif*, par opposition à une compensation qui toucherait toutes les syllabes. Ce procédé pourrait être une propriété du rythme fondé sur le pied (*foot-timing*). Par conséquent, la durée de la syllabe accentuée diminue lorsque la longueur du pied augmente, et ceci indépendamment de la longueur de mot.

En conséquence, une fois les deux niveaux séparés, Zhang ne retrouve pas les effets observés par Lindblom et Rapp et conclut que les deux types de compensation au niveau du mot sont des artefacts du raccourcissement au niveau du pied. Ainsi, un grand nombre d'effets dus à la longueur de mot sont réinterprétés comme étant des effets au niveau du pied.

Zhang observe également que la durée de la syllabe inaccentuée pré-accentuelle (extra métrique) est plus courte que la syllabe inaccentuée post-accentuelle, de même qu'elle a un effet allongeant sur la durée de la syllabe accentuée et cet effet est indépendant de la longueur de mot. L'effet allongeant de cette syllabe inaccentuée à gauche se combine aux autres effets internes au pied, à savoir le raccourcissement de la durée de la syllabe accentuée en fonction du nombre de syllabes dans le pied. Ceci signifie que les deux procédés sont totalement indépendants l'un de l'autre et que leurs effets peuvent donc être séparés linéairement.

Sur cette base, Zhang propose un modèle fondé sur la distinction de Hayes (1995) entre les langues iambiques et les langues trochaïques ; en effet, Hayes observe que les stimuli avec un contraste d'intensité sont groupés avec la proéminence initiale (pied trochaïque), alors que les stimuli avec un contraste de durée sont groupés avec la proéminence finale (pied iambique). La procédure proposée dans le modèle de Zhang consiste en un parallélisme de pieds (*parallel footing*), fondé sur le même principe que la phonologie autosegmentale : deux niveaux sont utilisés pour représenter les deux types de pied, le niveau trochaïque et le niveau iambique. Le modèle permet le chevauchement de pieds puisqu'il fonctionne sur deux niveaux séparés.

Le but du modèle de Zhang est de prédire les durées syllabiques en prenant en compte trois facteurs : le type de pied (trochaïque et/ou iambique), le niveau d'accentuation (accent primaire, secondaire ou inaccentué) et le nombre de syllabes dans le pied.

3.3.4. Facteur syntaxique

Il apparaît dans certaines études que la présence d'une frontière de mot a une incidence sur l'organisation temporelle des énoncés ; on peut alors se demander si les caractéristiques syntaxiques de ces frontières contribuent également à influencer les patrons temporels.

Cooper et Paccia-Cooper (1980) observent ce qu'ils appellent le raccourcissement trochaïque (raccourcissement de la syllabe accentuée par ajout de syllabes inaccentuées) et essaient de déterminer si les frontières syntaxiques bloquent cette règle de raccourcissement trochaïque TSR (*Trochaic Shortening Rule*). Dans les paires de stimuli du type :

‘*Clint* # *until*’
‘*Clinton* # *till*’

‘*Clin*’ dans ‘*Clinton* # *till*’ est beaucoup plus court que ‘*Clint*’ dans ‘*Clint* # *until*’, la présence de la frontière syntaxique dans le deuxième stimulus semble donc bloquer le raccourcissement trochaïque. Cooper et Paccia-Cooper étendent ce test à d'autres stimuli, dans des énoncés plus complexes et dans d'autres contextes syntaxiques. Ils observent que le blocage de la TSR est influencé par le statut de la frontière syntaxique, mais que les effets de la TSR ne sont pas cumulatifs : l'effet n'est pas augmenté lorsqu'on a concordance entre plusieurs types de frontière syntaxique.

Cooper et Paccia-Cooper testent plus amplement la TSR par l'intermédiaire de la palatalisation et du *flapping* : ces deux phénomènes se trouvent au-delà des frontières de mot, la palatalisation opérant également au-delà des frontières de syntagme. L'avantage de ces deux effets est qu'ils sont audibles et donc plus facilement contrôlables et observables que le raccourcissement au niveau du pied. Les deux phénomènes sont effectivement bloqués par les frontières majeures entre deux mots.

Les frontières fortes sont ensuite catégorisées par les auteurs en fonction d'un index de profondeur de branchage (branchage hiérarchique de nœuds syntaxiques). Ainsi, lorsque cet index est supérieur ou égal à deux, la frontière est qualifiée de forte. Ce travail se base donc sur la structure syntaxique et plus précisément sur la présence et la nature d'une frontière entre deux mots.

Notons que les raccourcissements observés au niveau syntaxique par Paccia et Paccia-Cooper peuvent être également expliqués d'un point de vue rythmique, plus

particulièrement sous l'angle du modèle de Jassem. Les deux stimuli vus plus haut correspondent aux schémas temporels suivants :

$$\begin{array}{cc} \textit{Clint until} & \textit{Clinton till} \\ | - | \cdot - | & | - - | - | \end{array}$$

‘*Clin*’ est plus court dans ‘*Clinton*’ puisqu’il se partage la durée globale de la NRU avec la syllabe inaccentuée suivante, contrairement à la NRU monosyllabique ‘*Clint*’.

De manière similaire, Giegerich (1992) explique les répartitions de durée syllabique en terme de facteur syntaxique. Il reprend les énoncés ambigus cités par Abercrombie :

- 1) Take **Grey** to London
- 2) Take **Greater** London

Selon la définition classique (Abercrombie), les compositions phonémiques et le regroupement en pieds sont identiques dans les deux cas (avec toutefois une divergence de type de pied), la distinction se fait en revanche sur les durées relatives des voyelles dans *Grey to* et *Greater*. La voyelle /eɪ/ dans 1) est plus longue que /ə/ (avec un ratio d'environ 2:1), alors que la voyelle /eɪ/ dans 2) a une durée voisine de celle de /ə/. Giegerich explique cette différence en terme de structure syntaxique : dans le premier cas, la voyelle /eɪ/ est suivie d'une frontière de mot et subit par conséquent un allongement, alors que dans le deuxième cas, la voyelle n'est suivie d'aucune frontière, elle n'est donc pas allongée. Par conséquent, l'information syntaxique importante ici est la présence d'une frontière de mot à l'intérieur du pied.

Les deux énoncés cités plus haut sont différenciés dans le modèle de Jassem en terme de regroupement en unités rythmiques : 1) NRU + ANA et 2) NRU (dissyllabique). Dans le modèle d'Abercrombie, il s'agit de deux pieds différents mais distingués sur la base des qualités vocaliques : 1) pied de Type C et 2) pied de Type B (section 3.2.3.1).

3.3.5. *Limitation d'étendue des facteurs rythmiques*

Jusqu'ici, nous nous sommes concentrée sur les modifications temporelles à l'intérieur de l'unité rythmique, mais selon certaines études, les effets dépassent les frontières de ces unités. En effet, l'organisation rythmico-temporelle de la parole se planifie à un niveau global, pas une unité après l'autre. C'est d'ailleurs ce qui conduit Jassem à expliquer que la NRU est doublement isochrone (en terme d'isochronie faible) : au

niveau de la durée des syllabes qui la composent (isochronie interne) ainsi qu'au niveau de la durée des NRU successives dans un énoncé (isochronie externe). Notons qu'il en est de même pour le pied accentuel.

Fourakis et Monahan (1988) observent que la planification et la production des pieds métriques sont effectuées à la fois en concaténation et en non-concaténation. Autrement dit, la durée de la syllabe accentuée est influencée par la structure de pied à laquelle elle appartient, mais également par celle des pieds environnants. Par conséquent, quand la structure du pied est modifiée, on ajoute des syllabes (inaccentuées) de manière concaténative avec les conséquences locales que l'on vient de voir, mais cet ajout se répercute sur l'ensemble de l'énoncé de manière non-concaténative sous la forme de divers raccourcissements. Cette répercussion se traduit par une accélération du débit des syllabes et une décélération du débit des pieds métriques. Les procédés de production utilisent des fonctions d'anticipation (*look-ahead*) avec une planification globale, ceci expliquant l'influence des pieds entre eux. Par conséquent, la production des syllabes influence le pied en question mais également les pieds voisins ; un modèle de prédiction de l'organisation rythmique doit tenir compte à la fois des paramètres locaux et globaux.

De manière similaire au niveau du mot, Lehiste (1974) ainsi que Rakerd, Sennet et Fowler (1992) remarquent qu'un mot a tendance à raccourcir lorsque l'énoncé augmente en taille, de même que sa durée diminue plus il est éloigné de la fin de l'énoncé. Par conséquent, la durée des mots dépend de l'organisation temporelle locale, mais également de l'organisation temporelle de l'ensemble de l'énoncé. On imagine donc qu'un phénomène d'ajustement de la durée s'exerce au sein de l'unité rythmique, mais également au niveau de l'ensemble des énoncés ou unités intonatives. Cette dimension supplémentaire, si elle s'avère significative, doit être prise en compte dans un modèle qui considère les caractéristiques des pieds environnants, au moins au niveau de l'unité intonative.

3.4. Phénomène accentuel en anglais

Cette section permet de définir brièvement l'accent en anglais, celui-ci étant à la base des regroupements rythmiques (quel que soit le modèle). Il s'agit ici uniquement de l'accentuation non emphatique, à savoir tout accent considéré dans un contexte neutre.

Il convient en tout premier lieu de distinguer ce que Bolinger appelle *stress* et *accent*. Cette distinction terminologique n'existe pas en français et nous proposons d'adopter la distinction proposée par Di Cristo (1999) entre 'accent lexical' (*stress*) et 'proéminence' (*accent*).

3.4.1. Définition

Définir le concept d'accent est en soi problématique. Globalement, on pourrait le qualifier de mise en relief particulière d'une partie d'un énoncé. La notion d'accentuation se présente sous deux facettes : celle de la production et celle de la perception. La production concerne l'effort articulatoire particulier fourni par le locuteur pour mettre en relief une partie de l'énoncé. L'autre facette concerne la sensation auditive de l'interlocuteur qui fait qu'il perçoit telle ou telle partie de l'énoncé comme plus saillante. Par conséquent, l'accent est une réalité autant pour le locuteur que pour l'interlocuteur, à savoir tant au niveau productif que perceptif (Lehiste 1970). Le locuteur structure son discours en organisant la répartition des accents et l'interlocuteur utilise cette répartition pour faire un premier décodage du message qui lui est envoyé puisque les parties du message avec un contenu sémantique important sont mises en relief.

On attribue généralement cinq fonctions différentes à l'accent (résumées par Astésano 2001), ces différents types de fonctions s'appliquant ou non selon les langues.

– La première de ces fonctions est la FONCTION RYTHMIQUE qui concerne les langues dans lesquelles il existe une alternance entre les syllabes accentuées et les syllabes inaccentuées pour éviter les conflits accentuels (*stress clashes*). Ces derniers sont les cas où deux syllabes accentuées se retrouvent adjacentes dans un énoncé, ces situations engendrant généralement des déplacements d'accent (*stress shift*) à condition que la syllabe sur laquelle l'accent se déplace soit accentuable. C'est par exemple le cas dans l'énoncé anglais souvent cité *thirteen men* dans lequel le patron accentuel théorique est *thir'teen 'men*, mais on observe souvent un déplacement du premier accent sur la première syllabe *thir* (syllabe accentuable), permettant ainsi un patron accentuel plus équilibré et évitant le conflit accentuel entre les deux syllabes consécutives *teen* et *men*. Liberman et Prince (1977) remarquent toutefois que les déplacements d'accents sont optionnels et dépendent entièrement des locuteurs.

- La deuxième fonction est la fonction HIERARCHIQUE : le phénomène accentuel permet la distinction entre les différents niveaux de saillance. Ainsi, la hiérarchie des différentes proéminences d'un énoncé fera que la proéminence nucléaire (proéminence majeure d'un énoncé) sera placée au sommet de cette hiérarchie de proéminence.
- L'accent possède également une fonction DEMARCATIVE : il indique alors les frontières de mots ou de syntagmes grâce aux accents initiaux et terminaux ; c'est le cas dans les langues telles que le tchèque et le français, caractérisées respectivement par un accent initial et final.
- La quatrième fonction est la fonction CONTRASTIVE : l'accent met en relief des unités accentuables, par contraste avec les unités non accentuables.
- La dernière fonction est la fonction CULMINATIVE : chaque mot lexical comporte au moins un accent. Cette fonction concerne les langues à rythme accentuel comme l'anglais.

A ces cinq fonctions, on pourrait ajouter la VALEUR DISTINCTIVE concernant uniquement les langues à rythme accentuel. Celle-ci permet de distinguer les paires de mots telles que le substantif *object* et le verbe *object*, le premier étant accentué sur la première syllabe et le deuxième sur la deuxième syllabe. On parle de valeur distinctive car l'emplacement de l'accent permet l'identification du nom ou du verbe.

Ajoutons également que l'on dit du phénomène accentuel en anglais qu'il est LIBRE car il ne tombe pas systématiquement sur la même syllabe quel que soit le mot. Il est libre en anglais, à l'instar d'autres langues dans lesquelles il peut être fixe comme en français par exemple, où l'accent tombe généralement sur la dernière syllabe d'un mot (d'où sa valeur démarcative). De ce fait, on a pendant longtemps pensé que l'accent était totalement imprévisible en anglais ; depuis, un certain nombre d'études ont établi une liste de règles applicables aux mots polysyllabiques permettant de déterminer l'emplacement des accents lexicaux (Chomsky et Halle 1968), notamment dans un cadre pédagogique (Ginésy 1990 ; Duchet 1994 ; Viel 1995).

3.4.2. *Corrélat acoustiques de l'accent*

Même s'il existe dans la littérature un consensus qui consiste à dire que l'accent est associé à une augmentation de la fréquence fondamentale, de la durée et de l'intensité, la définition acoustique de l'accent reste très disparate. Aucun de ces paramètres ne peut

servir d'indice accentuel de manière univoque car l'accentuation relève d'une combinaison de divers indices. On peut toutefois citer un certain ordre (décroissant) d'importance des trois corrélats majeurs de l'accent (Fry 1955 ; 1958) :

1. Fréquence fondamentale
2. Durée
3. Intensité

Fry (1955 ; 1958) a été le premier à observer et comparer les caractéristiques acoustiques de l'accent lexical. Ses conclusions sont que l'intensité et la durée participent activement à la perception de l'accent (avec toutefois prééminence pour la durée), mais la fréquence fondamentale apparaît comme étant le plus important de ces paramètres.

Pour Bolinger (1958), l'indice primaire de l'accent est la proéminence mélodique et la durée est un indice fonctionnant de pair avec la fréquence fondamentale ; en revanche, l'intensité ne joue aucun rôle dans la perception de l'accent. De ce fait, Bolinger ne parle pas d'accent (en tant que "*stress*"), mais plutôt d'accent mélodique (*pitch accent*) lorsqu'il s'agit de l'énoncé.

Lehiste (1970 : 128) reproche au travail de Bolinger de ne pas tenir compte des caractéristiques intrinsèques des phonèmes (cf. également Couper-Kuhlen 1986). En effet, bon nombre d'études s'accordent à dire que les trois paramètres sont les corrélats acoustiques de l'accent, mais les caractéristiques intrinsèques des différents phonèmes représentent des indices supplémentaires dont il faut tenir compte. Certaines voyelles sont intrinsèquement plus longues que d'autres (voyelles ouvertes vs voyelles fermées) ; il en va de même pour l'intensité et la fréquence fondamentale. Ces caractéristiques intrinsèques des voyelles sont en relation avec leur qualité et ne doivent pas être ignorées.

Si la fréquence fondamentale est le plus souvent présentée comme étant le paramètre acoustique le plus important de l'accent, certains travaux ont tenté de démontrer le contraire. Sluijter et van Heuven (1993) montrent sur le néerlandais que, contrairement à l'acceptation générale, l'intensité est un indice important de l'accent. Lors de la production d'un accent, les variations de l'intensité sont irrégulières et seules les fréquences hautes sont amplifiées. Ainsi, en observant les bandes de hautes fréquences, les auteurs obtiennent une augmentation de 3 à 11dB entre une voyelle inaccentuée et une voyelle accentuée, ceci étant au-dessus du seuil de perception.

Fant et Kruckenberg observent que la durée est l'indice acoustique prééminent de l'accent en suédois (1993) et en anglais (Fant, Kruckenberg et Nord 1991a). Contrairement à Sluijter et van Heuven (1993), Fant et Kruckenberg observent une augmentation de 2dB en contexte accentué et concluent que le rôle de l'intensité, en plus de déterminer le niveau de la hauteur, est de compléter le rôle des contours de la fréquence fondamentale. Les auteurs observent qu'en anglais la durée moyenne des syllabes accentuées est de 300ms contre 140ms pour les syllabes inaccentuées, à savoir une durée en moyenne deux fois plus longue (le même type de ratio est obtenu en suédois). Lehiste (1970) précise que la durée est un indice indéniable d'accentuation, mais elle n'est pas universelle et dépend de la langue dont il est question ; en estonien par exemple, les syllabes inaccentuées sont généralement plus longues que les syllabes accentuées. Notons que ces diverses remarques sont en contradiction avec la prédiction des deux modèles rythmiques de Jassem et d'Abercrombie/ Halliday.

L'effet de l'accent se répercute également sur la réalisation des différents segments selon qu'ils appartiennent à une syllabe accentuée ou inaccentuée (Greenberg, Carvey et Hitchcock 2002). Les segments en position d'attaque et de noyau sont la plupart du temps réalisés avec leur forme canonique en contexte accentué, alors qu'ils ont tendance à dévier de cette forme en contexte inaccentué. De plus, la distribution des noyaux vocaliques diffère largement selon qu'il s'agit d'une syllabe accentuée ou inaccentuée. La distribution est relativement égale dans les syllabes accentuées avec toutefois une préférence pour les voyelles basses (Greenberg, Chang et Hitchcock 2001 ; Hitchcock et Greenberg 2001) ; en revanche, les noyaux vocaliques inaccentués sont majoritairement des voyelles hautes antérieures et des voyelles hautes centrales.

3.4.3. Distinction entre accent (lexical) et proéminence

La base sur laquelle la distinction entre accent et proéminence se fonde fait l'objet de nombreux débats. Nous allons ici présenter les bases principales de cette distinction.

L'accent est défini en terme de FORCE RELATIVE (Jones 1918, 1972 ; Kingdon 1958) : une syllabe accentuée est plus saillante que les syllabes environnantes. De ce fait, elle ne peut devenir accentuée que si elle est entourée d'autres syllabes. Ce dernier point est largement critiqué par Jassem (1952 : 29) car cette approche exclut certaines syllabes : seuls les mots de plus d'une syllabe sont concernés et ceci signifie que l'on ne

tient pas compte des mots monosyllabiques prononcés de manière isolée, il faut qu'ils soient insérés dans un énoncé pour rentrer en compétition avec les autres syllabes. Ainsi, un mot monosyllabique prononcé de manière isolée ne peut pas être qualifié d'accentué ou d'inaccentué. Or, selon la définition de Palmer citée par Jassem (1995 : 24), le mot est la plus petite unité capable de fonctionner comme un énoncé complet¹⁶.

L'accent lexical et la proéminence sont également distingués sur la base du DOMAINE ACCENTUEL (Kingdon 1958 ; Cutler 1984). Kingdon distingue l'accent de mot et l'accent de phrase, le premier correspond au degré relatif de force utilisé pour réaliser les différentes syllabes d'un mot polysyllabique (accent lexical), son domaine est la *syllabe*. L'accent de phrase correspond au degré de force utilisé pour réaliser certains mots d'un énoncé (proéminence), son domaine est le *mot*. En ce sens, tous les mots polysyllabiques ont un accent de mot, mais dans un énoncé, seuls les mots avec un degré de force marqué ont un accent de phrase, qu'ils soient monosyllabiques ou polysyllabiques. Les deux niveaux ou domaines ne sont pas indépendants l'un de l'autre puisque l'accent de phrase tombe généralement sur une syllabe avec un accent de mot. Cette distinction met en valeur les fonctions hiérarchiques et contrastives de l'accent (Cutler 1984). Notons que Jassem (1952) et Bolinger (1980) préfèrent ne pas parler d'accent de mot/ phrase dans le sens où les 'mots' et les 'phrases' sont des catégories grammaticales et doivent donc être évitées.

La distinction entre accent et proéminence se fait également en fonction du caractère abstrait ou concret du NIVEAU D'ANALYSE. Pour Bolinger (1958), l'accent dit lexical est un accent potentiel à un niveau abstrait alors que la proéminence correspond à la réalisation effective de cet accent. La distinction de Jassem et Gibbon (1980) est similaire à celle de Bolinger ; l'accent lexical est défini comme une propriété phonologique des mots et représente un emplacement potentiel pour la proéminence. Cette approche possède donc un niveau d'analyse "abstrait" (accents lexicaux potentiels) et un niveau d'analyse "concret" avec le choix de réalisation des proéminences. La réalisation d'un accent lexical est donc une réalisation *potentielle* de proéminence.

Pour d'autres, l'accent est défini en fonction de l'organisation RYTHMIQUE (Jassem 1952 ; Jassem et Gibbon 1980 ; Couper-Kuhlen 1986 ; Wells 1991). Ainsi, l'accent en

¹⁶ « [the word is] the smallest speech unit capable of functioning as a complete utterance. »

anglais est essentiellement rythmique dans le sens où nous percevons les syllabes comme accentuées quand elles coïncident avec un battement rythmique (on parle de *stress-as-rhythm*). Notre production des syllabes accentuées est organisée de façon à produire des patrons rythmiques. Couper-Kuhlen (1986 : 33) souligne que la théorie de l'accent fondée sur le rythme permet de rendre compte des déplacements d'accent (*stress shift*) et de l'accent silencieux (*silent stress*).

Si la distinction entre accent et proéminence se fonde sur des approches différentes, il n'en reste pas moins que ces approches ne sont pas diamétralement opposées. On peut tout à fait considérer l'accent et la proéminence en termes de niveaux d'analyse (abstrait *vs* concret) et de domaines (syllabe *vs* mot). De plus, ces deux approches ne sont pas opposées à celle de Jassem et Gibbon, Couper-Kuhlen et Wells pour qui l'accent est essentiellement rythmique.

3.4.4. Prédiction de l'accent et de la proéminence

L'accent et la proéminence ne se situent pas au même niveau de représentation, mais ils sont totalement interdépendants. Malgré cette interdépendance, ils doivent être distingués notamment si on se place dans l'optique de prédiction de leur occurrence et de leur emplacement.

La prédiction de l'accent de mot en anglais est relativement aisée, même si celui-ci est libre. On a pourtant longtemps pensé qu'il n'était pas prévisible et que les apprenants devaient apprendre tous les mots de langue avec leur patron accentuel respectif (Jones 1972). Depuis, un ensemble de règles d'accentuation permet de déterminer la place de l'accent de mot notamment en fonction de sa structure morphologique (suffixes forts ou contraignants, faibles et mixtes¹⁷, préfixes de natures différentes), de sa catégorie grammaticale, du nombre de syllabes qu'il contient et de sa structure phonologique (Ginésy 1990 ; Roach 1991 ; Duchet 1994 ; Viel 1995 ; Lilly et Viel 1998).

Selon Cutler et Isard (1980) et Cutler (1984), le patron accentuel de chaque mot d'une langue appartient à la représentation phonologique de l'ensemble des mots du

¹⁷ Si on reprend la terminologie de Ginésy (1990) par exemple, Les suffixes forts déterminent l'emplacement de l'accent de mot (comme *-ic* par exemple) alors que les suffixes faibles ne permettent pas de prévoir l'accent (ex. : *-ment*). Les suffixes mixtes sont intermédiaires dans le sens où ils sont forts dans certains contextes et faibles dans d'autres (ex. : *-al*).

lexique mental. Autrement dit, les patrons accentuels ne découlent d'aucune règle mais sont stockés mentalement avec les mots, au même titre que leur composition phonémique. Cutler observe à ce sujet que les interlocuteurs s'aident des patrons accentuels pour la reconnaissance et la compréhension des mots, ils vont notamment être primordiaux lorsque les mots sont mal entendus ou que les circonstances induisent des erreurs d'interprétation.

Il en résulte qu'il est possible de prédire l'emplacement des accents lexicaux. En revanche, la prédiction des proéminences est nettement plus problématique car elles dépendent de facteurs sémantico-pragmatiques. Dans un cas neutre, la proéminence principale tombe par défaut sur la syllabe accentuée la plus à droite dans un énoncé (noyau ou syllabe nucléaire). Ceci est toutefois très théorique dans le sens où la proéminence nucléaire peut porter sur absolument n'importe quelle syllabe de l'énoncé, la plus à droite ou non, une syllabe accentuée ou non. Astésano (2001) oppose à ce sujet l'approche des linguistes générativistes (Chomsky et Halle 1968) pour qui la position de la syllabe nucléaire dépend de choix lexico-sémantiques de la part du locuteur (application de la Règle de l'Accent Nucléaire en fonction de la structure syntaxique) et l'approche des linguistes pragmaticiens (Bolinger 1972) pour qui aucune règle ne permet de prédire l'emplacement du noyau : le locuteur est absolument libre de placer le noyau sur n'importe quelle syllabe, indépendamment de tout facteur syntaxique. On peut ici faire référence à la célèbre citation de Bolinger selon laquelle il faut être télépathe pour pouvoir prédire l'emplacement des proéminences dans un énoncé.

Malgré ce flou, la tendance universelle neutre est de placer le noyau vers la fin de l'énoncé dans le but de faciliter l'accès à l'information et d'optimiser la communication (Astésano 2001). Il faut souligner que l'organisation des proéminences relève entièrement de la volonté du locuteur et reflète la structure informationnelle de l'énoncé. C'est cette caractéristique qui rend l'organisation des proéminences plus difficilement prédictible que l'organisation accentuelle des énoncés.

3.5. Résumé et conclusion

Ce chapitre présente la deuxième problématique soulevée dans ce travail, à savoir la définition de l'unité rythmique en anglais. Nous avons présenté et comparé deux modèles différents, celui de Jassem et celui d'Abercrombie/ Halliday. Leur but commun est de prédire les durées syllabiques à partir de l'organisation rythmique de la parole.

Les deux approches diffèrent fondamentalement dans le traitement des syllabes inaccentuées, la prise en compte des frontières de mot et par voie de conséquence dans le regroupement des syllabes en unités rythmiques, entraînant ainsi une prédiction divergente des durées syllabiques. Jassem regroupe les syllabes en unités rythmiques étroites et anacrouses alors que Abercrombie regroupe les syllabes en pieds accentuels. Nous avons ensuite présenté certains arguments dans les différents travaux présents dans la littérature que nous avons comparés aux deux modèles de Jassem et d'Abercrombie. La dernière section de ce chapitre est ensuite consacrée à la définition et à la question de la prédiction de l'accent en anglais, celui-ci étant à la base des deux types de regroupements en unités rythmiques.

Il serait par conséquent intéressant de comparer les deux modèles et de déterminer lequel permet une meilleure prédiction de la durée des unités rythmiques et des syllabes/ phonèmes qui les composent. Comparer et évaluer les deux modèles revient à tester l'importance des frontières de mot sur la structuration rythmique ainsi que l'effet de l'accent sur la durée des syllabes et phonèmes.

Dès lors que l'on s'intéresse à la structuration rythmique, une problématique supplémentaire se pose, celle de l'isochronie. En effet, le point commun des deux modèles, outre leur orientation de prédiction de l'organisation temporelle à partir du rythme, est la notion d'isochronie faible. Celle-ci suppose la récurrence régulière des intervalles accentuels, à savoir la NRU ou le pied accentuel selon l'approche. Nous allons approfondir cette problématique dans le chapitre suivant.

CHAPITRE 4.

L'ISOCHRONIE : « ILLUSION PERCEPTIVE » OU REALITE ?

4.1. DEFINITION	146
4.2. PEUT-ON PARLER D'ISOCHRONIE EN PRODUCTION ?	147
4.2.1. Facteurs influençant la durée des intervalles accentuels.....	147
4.2.2. Isochronie stricte ou isochronie faible ?.....	150
4.2.3. Problème lié à l'isochronie : comment la mesurer ?	152
4.3. L'ISOCHRONIE EN TANT QUE PHENOMENE PERCEPTIF	154
4.4. CATEGORISATION DES LANGUES	157
4.4.1. Historique.....	157
4.4.2. Échec de la catégorisation classique des langues en fonction du rythme .	160
4.4.3. Approches alternatives du rythme de la parole	161
4.4.4. Exemple de différence rythmique intra-langue.....	167
4.4.5. Identification automatique des langues	167
4.5. RESUME ET CONCLUSION.....	171

Dans le chapitre précédent, nous avons présenté le mode de fonctionnement des modèles que nous avons qualifiés de "supra-syllabiques" dans le premier chapitre. Nous avons présenté principalement deux types de modèles rythmiques divergeant sur la structure des unités rythmiques utilisées ainsi que les différents arguments en faveur ou défaveur de ces approches.

Une caractéristique de la structuration rythmique en production est celle de l'isochronie des unités rythmiques, différant selon la langue dont il s'agit. Cette caractéristique est d'ailleurs le point commun majeur des deux modèles de Jassem et d'Abercrombie/ Halliday puisque les deux prévoient que la durée des unités rythmiques (NRU et pied accentuel) est approximativement égale quel que soit le nombre de sous-constituants. Toutefois, cette notion d'isochronie soulève de nombreux débats ; nous consacrerons par conséquent ce chapitre à cette problématique. Nous utiliserons ici le terme de pied (accentuel) par défaut en tant qu'unité rythmique de base en anglais, sans distinction entre le pied abercrombien et la NRU de Jassem, ceci dans le but de simplifier les diverses descriptions.

Nous définirons tout d'abord le concept d'isochronie. Nous verrons ensuite les implications et les problèmes que pose la notion d'isochronie en production. La troisième section est consacrée à l'aspect perceptif de l'isochronie. Nous verrons dans la section suivante le débat fondamental qu'elle pose, à savoir celui de la catégorisation des langues en fonction de leurs caractéristiques rythmiques, fortement remis en question dans la littérature. Nous décrirons alors les approches alternatives proposées pour rendre compte de ces distinctions.

4.1. Définition

L'isochronie est le principe selon lequel la chaîne parlée est divisée en un certain nombre d'unités de durée égale. Classé (1939) est un des premiers à citer et à observer cette régularité temporelle dans la parole ; il remarque toutefois qu'une isochronie parfaite ne peut être réalisée que dans des conditions bien précises, avec à la fois une structure phonémique et une structure grammaticale similaires. C'est ensuite Pike (1945) qui définit le rythme comme étant la récurrence isochrone d'unités, cette définition étant reprise par la suite par Abercrombie (1967). Cette définition aboutit à une distinction catégorique entre les langues dont la récurrence porte sur la syllabe (langues isosyllabiques) et celles dont la récurrence porte sur la syllabe accentuée (langues iso-accentuelles).

Abercrombie (1965) fonde cette distinction sur des critères physiologiques de production : toutes les langues sont caractérisées par des battements respiratoires (*chest pulses*) et des battements accentuels (*stress pulses*). Les premiers correspondent aux souffles d'air sortant des poumons donnant lieu à la contraction et la relaxation des muscles impliqués dans les mouvements respiratoires. Les battements accentuels correspondent eux à des contractions plus fortes des muscles respiratoires renforçant ainsi certains battements respiratoires. Par conséquent, lorsque les battements respiratoires reviennent à intervalles réguliers, la langue est qualifiée d'isosyllabique. En revanche, lorsque ce sont les battements accentuels qui sont isochrones, il s'agit d'une langue iso-accentuelle. Nous reviendrons sur cette distinction entre les langues dans la section 4.4.

Lehiste (1977) passe en revue la littérature sur l'isochronie en anglais et remarque qu'il existe peu de consensus entre spécialistes au sujet de l'isochronie. Certains ont affirmé que l'isochronie était bien présente dans la parole comme par exemple Pike (1945), Jassem (1952 ; 1983), Abercrombie (1965) et Halliday (1967). Uldall (1971) observe qu'il existe un phénomène d'isochronie dans les pieds de une, deux et trois syllabes, particulièrement en parole lente, mais pas dans ceux de quatre syllabes. Selkirk (1984a) précise qu'il existe bien une impression d'isochronie, cette impression reflétant l'organisation cognitive de la parole en temps réel. Pour Hayes (1995), les pieds sont régis par le principe d'eurythmie : les syllabes accentuées sont séparées par des espacements réguliers.

En revanche, d'autres auteurs ont complètement rejeté l'isochronie comme par exemple Shen et Peterson (1962), Bolinger (1965), O'Connor (1965), Lea (1974), Faure, Hirst et Chafcouloff (1980), Nakatani, O'Connor et Aston (1981), Dauer (1983), Eriksson (1991), Nooteboom (1991) et Tatham et Morton (2002).

L'isochronie a mené de nombreux linguistes à faire référence à une « impression de régularité » (Faure, Hirst et Chafcouloff 1980 : 72). Pour illustration, citons Pike (1945) :

« The units tend to follow one another in such a way that the lapse of time between the beginning of their prominent syllable is somewhat uniform » (1945: 34)

Abercrombie (1965) :

« Feet of (approximately) even length » (1965: 28)

Halliday (1967) :

« Tendency for salient syllables to occur at roughly regular intervals of time » (1967: 12)

Et Jassem (1984) :

« a given number of syllables to be of equal length » (1984: 426)

Ce manque de consensus observé par Lehiste et le nombre de travaux contredisant le principe d'isochronie nous mènent à nous interroger sur la validité de l'hypothèse d'isochronie en production.

4.2. Peut-on parler d'isochronie en production ?

4.2.1. Facteurs influençant la durée des intervalles accentuels

Comme nous l'avons vu plus tôt, les premiers travaux sur l'isochronie en anglais postulent une durée quasi constante des intervalles accentuels. Cette isochronie implique que dans les langues comme l'anglais, les syllabes comprises dans les intervalles sont compressées ou étirées pour que la durée totale du regroupement de ces syllabes s'ajuste à la durée (quasi) constante de l'intervalle accentuel, et ceci quelle que soit la composition segmentale des syllabes en question. L'isochronie accentuelle

suppose donc une capacité de compression et d'élasticité des syllabes, et par voie de conséquence, des phonèmes qui les composent.

Le problème est que ces travaux souffrent du manque de mesures physiques exactes (Buxton 1983) ; de plus, la plupart des travaux ultérieurs rapportent que la durée des intervalles accentuels n'est pas régulière mais plutôt qu'elle dépend du NOMBRE DE SYLLABES qui la compose (Bolinger 1965 ; Faure, Hirst et Chafcouloff 1980 ; Nooteboom 1981 ; Nakatani, O'Connor et Aston 1981 ; Dauer 1983 ; Eriksson 1991 ; Fant, Kruckenberg et Nord 1991a et b ; Van Santen 1997 ; Tatham et Morton 2002). La durée des intervalles accentuels augmente linéairement en fonction du nombre de syllabes ajoutées dans l'intervalle, il n'y a donc pas d'isochronie.

Faure, Hirst et Chafcouloff (1980) testent l'isochronie en comparant 114 intervalles délimités par l'attaque de chaque voyelle accentuée. Il ressort que la durée des intervalles varie en fonction du nombre de syllabes comprises dans chacun des intervalles, contrairement à la définition même de l'isochronie. Il est donc évident que les syllabes accentuées ne sont pas séparées par des intervalles temporels approximativement égaux.

Les auteurs observent ensuite que les syllabes ne subissent aucune compression : les syllabes accentuées sont simplement plus longues que les syllabes inaccentuées. Les auteurs attribuent des durées syllabiques fixes à partir des observations de leur corpus : les syllabes accentuées durent 0.22 seconde contre 0.14 seconde pour les syllabes inaccentuées. Les auteurs en viennent à la conclusion que c'est la perception des syllabes accentuées par opposition à celle des syllabes inaccentuées qui donne l'impression de percevoir les intervalles accentuels comme étant égaux en durée.

Cette conclusion est en accord avec les travaux de Roach (1982) et de Dauer (1983) ; notons toutefois que l'attribution d'une durée fixe à la syllabe accentuée va à l'encontre des prédictions des modèles rythmiques que nous avons présentés dans le chapitre précédent. Quel que soit le modèle, la durée de la syllabe accentuée diminue lorsque l'unité augmente en taille, tout comme celle des syllabes inaccentuées. Mais, comme le montre Eriksson (1991) par la suite, une augmentation linéaire de la durée des pieds en fonction du nombre de syllabes (comme l'observent Faure, Hirst et Chafcouloff) n'est pas incompatible avec une diminution progressive de la durée de ces syllabes.

L'influence du nombre de syllabes dans l'intervalle accentuel ou pied est l'argument principal à l'encontre de l'hypothèse d'isochronie stricte. Nous pouvons également mentionner l'influence de la POSITION DU PIED dans une unité phonologique plus grande comme l'unité intonative ou l'énoncé. De par les effets de frontière, certaines études rapportent un allongement d'environ 20% des unités rythmiques en fin d'énoncé en parole lue (Fox 2000 ; Tatham et Morton 2002). Si cet allongement final s'avère systématique quel que soit le type de discours, il est évident que les intervalles ne peuvent tous être absolument égaux en durée. De manière similaire, Tatham et Morton (2002) observent un raccourcissement des unités rythmiques suivant une pause, ceci compromettant un peu plus l'existence de l'isochronie (stricte). Les auteurs observent que la durée des unités rythmiques est très variable, mais il est cependant intéressant de remarquer que ces variations ne semblent pas être hasardeuses, les patrons de durée des unités semblent relativement similaires dans tous les énoncés de leur corpus.

Les travaux sur l'isochronie ont également engendré des COMPARAISONS INTER-LANGUES dont l'organisation rythmique est supposée différer : les intervalles accentuels sont supposés avoir une durée quasi-égale dans les langues à rythme accentuel et les syllabes sont conditionnées par le principe d'isochronie dans les langues à rythme syllabique (cf. section 4.4). Dauer (1983) compare l'anglais, le thaï, l'espagnol et l'italien et observe la durée des intervalles accentuels et des syllabes dans les quatre langues et découvre qu'il n'y a aucune différence significative entre les durées d'intervalle, quelle que soit la langue. Ce résultat rejoint les travaux de Fant, Kruckenberg et Nord (1991a et b) : il n'y a aucune différence significative de durée entre les intervalles accentuels en anglais, suédois et français. L'espacement moyen entre chaque syllabe accentuée est de 555ms en français, 565ms en anglais et 555ms en suédois, la durée de ces intervalles augmentant en fonction du nombre de phonèmes dans les trois langues.

L'absence d'isochronie stricte peut également trouver son origine dans les fonctions multiples du phénomène accentuel. Hayes (1995) remarque que l'emplacement des accents n'a pas uniquement une fonction rythmique ; il peut par exemple marquer la structure morphologique et syntaxique ou signaler la distribution du focus. Par conséquent, le rythme joue un rôle important dans l'emplacement des accents, mais il n'est pas le seul facteur à l'influencer : ces emplacements résultent d'une combinaison de différents facteurs. Il est donc difficilement imaginable que tous les patrons

accentuels soient parfaitement réguliers, du fait même de cette interaction. L'influence du rythme est certes majeure mais pas unique.

Un dernier facteur influençant la durée des unités rythmiques est le DEBIT de parole ; nous avons évoqué dans le premier chapitre l'étendue des effets dus aux variations de débit et les difficultés que pose le contrôle d'un tel facteur. Nous pouvons supposer qu'un débit rapide influence tout autant la durée des unités rythmiques.

4.2.2. *Isochronie stricte ou isochronie faible ?*

Toutes les approches que nous venons de citer montrent que l'isochronie stricte n'existe pas en anglais. Toutefois, il convient de préciser que rares sont ceux qui défendent une définition aussi stricte de l'isochronie, il s'agit plutôt d'isochronie FAIBLE (Jassem 1952 et 1983 ; Abercrombie 1964 ; Halliday 1967 ; Uldall 1971 ; Catford 1977 ; Fowler 1977 ; Buxton 1983 ; Hoequist 1983a ; Eriksson 1991 ; Kohler 1991). En effet, la durée des intervalles séparant chaque attaque de syllabe accentuée est APPROXIMATIVEMENT égale, mais pas strictement égale (Faure, Hirst et Chafcouloff 1980). Notons à ce sujet que, comme le souligne Eriksson (1991), les définitions de l'isochronie que nous avons vues plus haut dégagent cette idée d'approximation (nos soulignements) :

« ... somewhat uniform » (Pike), « ... roughly regular » (Halliday), « feet of (approximately) equal length » (Abercrombie)

Plutôt que l'isochronie, Catford (1977) préfère parler d'isodynamisme : chaque pied implique environ la même quantité de pouvoir d'initiation (*initiator power*). Plus globalement, l'accent anglais est caractérisé par une augmentation soudaine du pouvoir d'initiation, puis un déclin progressif de ce pouvoir jusqu'au battement (ou accent) suivant. Le pouvoir d'initiation se répartit donc dans des intervalles (ou quanta) plus ou moins égaux ; il est donc ici aussi question d'isodynamisme faible. Catford cite à ce sujet l'expérience de Halliday (1967) dans laquelle ce dernier demande à des sujets de lire des phrases en parlant fort dans le but de mesurer le rapport entre les durées de pieds de une, deux et trois syllabes. Ce rapport est de 5:6:7, ce qui signifie que l'isodynamisme n'est pas absolu (il faudrait pour cela un rapport de 1:1:1), mais il existe une tendance marquée vers une certaine régularité.

D'après Fowler (1977), les locuteurs montrent une certaine résistance aux déviations par rapport à une organisation accentuelle potentiellement régulière. Divers degrés de

compression dus à l'ajout de syllabes inaccentuées sont en faveur d'une régularité accentuelle, mais faible. La compression du matériel phonétique entre deux accents existe mais ces effets de compression ne sont pas assez marqués pour soutenir une version stricte de l'isochronie. Séguinot (1979) observe que la durée des groupes rythmiques n'augmente pas proportionnellement au nombre de syllabes qu'ils contiennent : la durée de l'intervalle augmente mais pas de manière constante, notamment pour les groupes rythmiques courts. On peut ainsi parler de tendance vers l'isochronie.

Buxton (1983) remarque qu'un certain degré d'isochronie peut être observé si d'autres techniques de mesure sont considérées. Buxton cite les travaux de Hill, Witten et Jassem (1978) dans lesquels certaines unités rythmiques sont plus isochrones que d'autres, notamment les unités en position finale d'énoncé et celles comprenant l'accent principal de l'énoncé. Ainsi, les énoncés anglais sont caractérisés par une structure rythmique permettant une meilleure prédictibilité des syllabes accentuées (par rapport aux syllabes inaccentuées) ainsi que des syllabes en fin d'énoncé (par rapport aux syllabes non finales). Buxton insiste sur le fait que les syllabes accentuées sont les plus pertinentes dans la parole : elles contiennent le contenu informationnel et les sujets se concentrent sur ces syllabes (le cerveau semble d'ailleurs anticiper les syllabes accentuées de manière relativement fiable). Ces diverses remarques suggèrent que l'isochronie ne constitue plus un phénomène global, mais plutôt un phénomène conditionné par un contexte de taille, de position et de niveau d'accentuation ; on revient donc sur l'idée de conditionnement mis en avant par Classé. La question que l'on peut donc soulever est de savoir si, avec un tel conditionnement, l'isochronie peut exister en parole spontanée.

Kohler (1991) montre également qu'il existe une tendance marquée vers une isochronie faible. Cette tendance est observée en allemand et l'auteur suppose qu'il en sera de même en anglais et en néerlandais. La compression des syllabes s'observe uniquement dans les intervalles mono et dissyllabiques. Au-delà de deux syllabes, on atteint un plafond et il n'apparaît plus de compression. De plus, l'auteur observe un ajustement de la durée de la syllabe inaccentuée en fonction de la durée de la syllabe accentuée précédente, ceci abondant dans le sens d'une tendance vers une durée régulière des intervalles accentuels.

Notons enfin que l'existence d'une succession régulière, ou quasi régulière, de battements est obligatoire pour l'organisation de patrons rythmiques (Selkirk 1984a). Si une telle régularité n'existait pas, nous ne pourrions pas par définition parler de rythme. Toutefois, l'isochronie ne doit pas être une condition absolue pour qu'un énoncé soit qualifié de rythmique : la notion de patron accentuel (récurrence d'événements forts et faibles) est tout aussi importante que le concept d'isochronie pour qualifier un énoncé de rythmique. Toutes les syllabes ne sont pas équivalentes rythmiquement, seules certaines le sont. De la même manière qu'il existe en musique une certaine liberté rythmique, la parole ne doit pas être absolument isochrone pour qu'elle soit qualifiée de rythmique ; il existe des variations inévitables par rapport à une norme isochrone stricte, dues à divers facteurs tels que le nombre de syllabes dans chaque groupe mesuré, la position de l'unité, *etc.* (Fox 2000). Notre perception du rythme est souple : elle est globale et compense naturellement tout écart par rapport à une régularité stricte potentielle.

4.2.3. Problème lié à l'isochronie : comment la mesurer ?

Tester l'isochronie pose certains problèmes. Lehiste (1977) soulève trois points à ce sujet.

- De nombreuses études n'ont pas abouti sur l'observation de quelque phénomène d'isochronie dans la parole tout d'abord parce que les styles utilisés sont très variés : les corpus sont tantôt de la prose, tantôt de la versification ; or, ces différents styles rendent les comparaisons délicates.
- Aucun cadre méthodologique uniforme n'a été établi afin de tester l'isochronie dans la parole, ici aussi rendant les comparaisons difficiles.
- Les différents auteurs ne définissent jamais la quantité d'écart acceptable par rapport à l'isochronie stricte pour que l'on qualifie un énoncé de régulier. Autrement dit, nous savons ce que représente l'isochronie stricte, mais nous ne savons pas combien de variation est acceptable pour qu'un énoncé soit qualifié de faiblement isochrone. Toutefois, Fowler (1977) et Dauer (1983) reprochent à Lehiste de ne pas non plus déterminer les limites de variation que l'on peut utiliser pour distinguer une production faiblement isochrone d'une production non isochrone.

Fowler cite également la question de la vitesse d'élocution : certains rapportent que le débit doit être lent pour obtenir un effet d'isochronie car les durées syllabiques sont plus contrôlées ; en revanche, d'autres prétendent que le débit doit être rapide car il correspond mieux à la parole naturelle et donc au rythme authentique de la parole.

En réponse au problème de quantification de l'isochronie, Couper-Kuhlen (1993) s'emploie à observer les écarts par rapport à une production parfaitement isochrone, ceci dans un corpus radio. L'anglais spontané n'est pas uniformément isochrone, mais l'auteur observe que des portions entières de parole de son corpus se trouvent être isochrones. Ces portions correspondent généralement à des structures parallèles ou à des discontinuités syntaxiques, mais il est impossible de prédire quelle portion sera ou ne sera pas isochrone. À partir de ces données, Couper-Kuhlen essaie de déterminer l'écart au-delà duquel un énoncé est considéré comme non isochrone. Les intervalles sont mesurés d'attaque vocalique en attaque vocalique. L'auteur estime qu'il faut que la durée d'un intervalle soit modifiée de plus de 20% pour qu'il ne soit plus perçu comme isochrone avec toutefois quelques précautions car certaines séquences sont isochrones avec des écarts de plus de 20%.

Parallèlement à ces différents points, se pose également la question de la définition des limites de l'intervalle à mesurer (Fowler 1977 ; Roach 1982 ; Eriksson 1991 ; Cummins et Port 1998). La plupart considère les intervalles commençant par l'attaque de la syllabe accentuée ou l'attaque de la voyelle accentuée, que ce soit en production ou en perception (Eriksson 1991). Toutefois, pour Fowler (1977), mesurer les intervalles entre les attaques vocaliques de syllabe accentuée n'est pas exact, car ceci revient à évaluer une version beaucoup trop forte de l'isochronie. De plus, l'attaque vocalique comme marqueur de début d'intervalle est remise en question, notamment par ceux qui mettent en avant la théorie du centre perceptif P (*P-center*) ou battements accentuels (Morton, Marcus et Frankish 1976 ; Marcus 1979). Le centre P est défini comme le point qui correspond au moment psychologique de perception de chaque syllabe. Il est l'instant dans le signal de parole que les interlocuteurs utilisent pour aligner le signal et les battements artificiels qui leur sont présentés (dans un cadre expérimental), celui-ci étant influencé par le contexte phonétique et la durée de la voyelle.

Mais la théorie du centre perceptif est sujette à de nombreuses critiques. Eriksson (1991) et Couper-Kuhlen (1993), par exemple, expriment quelques réserves au sujet des

conditions d'expérimentation du centre perceptif, principalement parce que les résultats portent sur de la parole de laboratoire, plus exactement sur des listes de mots isolés, voire de non-mots, souvent monosyllabiques, dans des conditions expérimentales favorisant l'isochronie. Il semble difficile de savoir si cette théorie des centres perceptifs est réellement applicable à de la parole continue. De plus, Couper-Kuhlen remarque que les centres perceptifs ne sont en fait que des positionnements relatifs et qu'il est impossible de déterminer leur emplacement de manière exacte, compromettant ainsi la validation de l'hypothèse de l'isochronie. Enfin, ces études ne se tournent pas vers la question du débit de parole, or il est probable que l'emplacement des centres perceptifs sera influencé par les variations de débit.

4.3. L'isochronie en tant que phénomène perceptif

Nous avons vu qu'en production, la parole ne répondait pas à un patron isochrone strict. Pourquoi alors percevons-nous la parole comme étant (faiblement) isochrone ? Plus généralement, pourquoi percevons-nous la parole comme régulière alors qu'elle ne l'est pas physiquement ? Cette isochronie subjective, ou « illusion perceptive » (Classé 1939), est à l'origine de diverses études en perception.

Allen (1975) observe que les interlocuteurs sous-estiment la durée des intervalles longs et surestiment la durée des intervalles courts, c'est ce qui explique qu'ils perçoivent les intervalles comme étant plus réguliers qu'ils ne le sont réellement (cf. également Donovan et Darwin 1979). Pour Lehiste (1977), suivie de Selkirk (1984a) et Couper-Kuhlen (1993), la durée de chaque intervalle accentuel varie, mais les différences de durée ne sont pas perceptiblement significatives. Lehiste suppose donc que l'isochronie est un phénomène perceptif, étant donné que le concept d'isochronie impliqué par l'organisation rythmique ne se retrouve pas dans le signal acoustique.

Lehiste présente des stimuli à des sujets et leur demande d'identifier les intervalles les plus longs et les intervalles les plus courts ; l'auteur observe qu'il faut augmenter (ou diminuer) un intervalle de 30 à 100 ms pour qu'il soit jugé comme le plus long (ou le plus court). Les interlocuteurs ont tendance à compenser partiellement les contraintes de durée et perçoivent donc la parole comme étant plus rythmique qu'elle ne l'est réellement. Lehiste cite Coleman (1974) qui montre que, malgré l'augmentation de la durée des intervalles accentuels en fonction du nombre de syllabes inaccentuées, les

interlocuteurs ont tendance à percevoir les intervalles accentuels comme étant plus isochrones qu'ils ne le sont vraiment. La notion d'isochronie répond par conséquent à une dichotomie entre le niveau de la perception et de la production.

Lehiste précise que les locuteurs ont toutefois tendance à produire une certaine isochronie qui se reflète dans le traitement et l'ajustement des contraintes de durée, comme par exemple le raccourcissement des consonnes dans un groupe consonantique ou le raccourcissement de la durée de la racine monosyllabique d'un mot lorsque des suffixes lui sont ajoutés. D'après Lehiste, ces ajustements de la durée des segments visent à ce que la durée des mots atteigne une durée moyenne commune, il y a donc bien une recherche de la régularité.

Fowler (1977) remarque que des preuves moins directes de la régularité accentuelle ressortent des tests de perception : les accents sont organisés de manière régulière et prédictible, même si cette organisation n'est pas nécessairement isochrone. Les sujets savent non seulement qu'une syllabe accentuée arrive, mais ils savent quand elle va se présenter. Ces résultats confirment, selon Fowler, l'existence d'une version faible de la régularité accentuelle : les intervalles entre syllabes accentuées sont contraints d'une manière régulière et prédictible (cf. également Buxton 1983). De plus, D'après Bolinger (1981 : 33), les portions accentuelles de la parole sont plus faciles à produire et à comprendre lorsqu'on a une tendance rythmique régulière. Cette restitution rentre par conséquent en jeu dans le décodage et la compréhension du message.

Ainsi, il semblerait que les interlocuteurs plaquent une régularité sur un signal qui n'est pas forcément régulier. Cette remarque soulève un point qui nous paraît important concernant l'organisation rythmique de la parole. Si l'interlocuteur restitue un rythme régulier dans un signal dont le rythme est irrégulier, c'est qu'il possède probablement un patron rythmique régulier interne auquel il se réfère constamment (Astésano 2001). Ce patron expliquerait pourquoi l'interlocuteur sous-estime les intervalles longs et surestime les intervalles courts. Un tel patron, si celui-ci existe effectivement, va influencer autant la perception de la parole que la production langagière ; ceci signifie que chaque individu possède une forme régulière sous-jacente qui est modifiée lors de la phase de production langagière (Cutler 1980) mais restituée lors de la phase de perception ; on revient ici sur la distinction entre mètre et rythme (Di Cristo 2002).

Hoequist (1983a) explique à ce sujet qu'il existe une relation complexe entre le signal acoustique et ce que nous percevons. Cette complexité traduit le fait que le

chercheur observant la durée des intervalles dans un signal n'utilise probablement pas le même point de référence qu'un interlocuteur jugeant inconsciemment la durée de ces mêmes intervalles. Notons que c'est dans l'optique de la résolution de cette complexité que la théorie des centres perceptifs a été proposée.

La déviation entre le patron régulier sous-jacent et la production irrégulière résulte en partie de l'influence de divers facteurs tels que les caractéristiques intrinsèques des phonèmes utilisés, des contraintes articulatoires, rythmiques, contextuelles, *etc.* On a une déformation lors de la phase de production qui fait que les énoncés produits ne sont pas isochrones : il existe donc un paradoxe entre les phases de pré-production et de production. Par conséquent, la rythmicité, si elle existe, se situe au niveau perceptif et au niveau de la pré-production, mais pas au niveau de la réalisation acoustique (Benguerel et D'Arcy 1986 : 244). Bertinetto (1989) en conclut que l'isochronie est une question d'image intériorisée de la parole.

Eriksson (1991) observe la perception du rythme dans la parole au travers de tâches de reproduction du rythme. Des stimuli de parole ou de bruits sont présentés à des sujets et ils doivent reproduire le rythme au moyen de petites tapes et juger la longueur et la régularité des intervalles. Il ressort que les sujets perçoivent la parole comme plus régulière qu'elle ne l'est, sauf lorsque les stimuli ne sont pas de la parole. Eriksson explique que dans ce type de démarche, la complexité acoustique et le poids sémantique des stimuli rendent la tâche plus difficile, ceci expliquant pourquoi les stimuli avec du bruit ne sont pas perçus comme plus réguliers qu'ils ne le sont réellement (contrairement à de la parole). Pour Lehiste, tel est le cas pour la parole car, dit-elle, nous sommes incapables de détecter la plupart des différences et nous attendons une certaine isochronie. En revanche, Eriksson explique ces réponses par le fait que la complexité de la tâche rend le rythme difficile à percevoir et les sujets ont comme réponse à cette complexité une sorte de réflexe correspondant à un rythme généralement régulier. Ces deux interprétations ne sont pas contradictoires si on les considère sous l'angle de la dichotomie entre pré-production régulière et production irrégulière.

4.4. Catégorisation des langues

4.4.1. Historique

L'isochronie est directement liée à une impression de régularité et de cette impression est née la distinction des différentes langues du monde en fonction de leurs caractéristiques rythmiques. Pour traiter la question de l'isochronie, il convient de se tourner maintenant vers cette problématique de catégorisation des langues.

Loyd James (1940) décrit le son produit par les langues ensuite apparentées aux langues à rythme accentuel (comme l'anglais) comme étant le son d'un message en morse, par opposition au son de mitraillette qui caractérise les langues dites à rythme syllabique (comme le français). Pike (1945) distingue également les deux types de rythme : dans certaines langues, les syllabes reviennent à intervalles réguliers, alors que dans d'autres, ce sont les syllabes accentuées. Ceci rejoint la définition que nous avons donnée plus tôt : le rythme est fondé sur les syllabes accentuées, et tel est le cas pour toutes les langues à rythme accentuel comme l'anglais ; la durée de chaque unité rythmique dépend de la présence d'un accent plutôt que du nombre de syllabes. Par opposition, dans les langues à rythme syllabique comme le français, le rythme est lié à la syllabe, auquel cas ce sont les syllabes plutôt que les accents qui ont tendance à revenir à intervalles réguliers.

Cinq principes fondamentaux ressortent de cette application du principe d'isochronie. Pour les langues ACCENTUELLES (Pike 1945) :

- la durée des intervalles accentuels est (quasi) égale,
- elle est indépendante du nombre de syllabes dans l'intervalle et
- l'intervalle est caractérisé par un certain degré de compression des syllabes.

Pour les langues SYLLABIQUES :

- les syllabes sont de durée (quasi) égale et
- la durée des intervalles est proportionnelle au nombre de syllabes.

Abercrombie (1967) précise que la structure rythmique des langues est fondée sur l'isochronie des unités rythmiques, à savoir les intervalles accentuels pour les langues à rythme accentuel (*stress-timed*) comme l'anglais, le russe et l'arabe, et les syllabes pour les langues à rythme syllabique (*syllable-timed*) comme le français, le torugu et le yoruba. Toutes les langues appartiennent nécessairement à l'une ou l'autre catégorie.

Dans cette catégorisation des langues, il faut également citer l'existence d'une troisième catégorie : celle des langues à rythme moraïque, représentées principalement par le japonais, dont l'unité est la more (cf. section 3.2.1). Ladefoged (1975) propose ce troisième type de rythme fondé sur la more car ses propriétés rythmiques se distinguent de celles des langues à rythme syllabique et des langues à rythme accentuel. Ainsi, le rythme repose sur l'isochronie des mores, tout comme le rythme des langues à rythme accentuel se fonde sur l'isochronie (théorique) des syllabes accentuées.

Malgré des terminologies et/ ou des méthodologies divergentes, de nombreuses études se sont intéressées à la catégorisation des langues en fonction de leurs caractéristiques rythmiques. Par exemple, Wenk et Wioland (1982) opposent les langues « *codachrones* » (*trailer-timed languages*) dans lesquelles la syllabe accentuée est finale de groupe rythmique et les langues « *capochrones* » (*leader-timed languages*) dans lesquelles la syllabe accentuée est initiale de groupe rythmique. Selkirk (1984a) représente cette distinction au travers d'une grille métrique. Une grille métrique consiste en une série de croix *x*, agencée hiérarchiquement au-dessus des énoncés, permettant de représenter les différents niveaux d'accentuation. La structure rythmique est représentée par les trois premiers niveaux de la grille : le premier niveau représente les syllabes, le deuxième niveau les éléments proéminents (accents primaires et secondaires) et le troisième niveau l'accent principal dans le mot phonologique. La différence entre les deux catégories de langues réside dans le deuxième niveau de la grille, celle-ci est illustrée dans Figure 12 ci-après.

x	x	x x x x	x x x x
x x x x	x	x x x x	
a. the manager's here		b. il popolo	

Figure 12. Exemples de grilles métriques pour un énoncé anglais et un énoncé italien (Selkirk 1984a :40-41)

En anglais, seules les syllabes accentuées bénéficient d'une croix supplémentaire alors qu'en italien, toutes les syllabes bénéficient d'une croix supplémentaire. Ainsi, l'isochronie réside dans les intervalles délimités par les croix du niveau 2 : entre les syllabes accentuées en anglais et entre les syllabes en italien. Par conséquent, la

distinction entre les différentes catégories de langue peut être représentée au niveau basique de l'organisation rythmique.

Les psychologues s'intéressent également à la perception du rythme des langues ; Rice (1992) demande à des sujets de nationalité différente de grouper des sons synthétiques dont le rythme est régulier. Les résultats montrent que les éléments avec un contraste d'intensité sont regroupés avec l'élément proéminent initial (typiquement les langues iambiques comme l'anglais) et que les éléments avec un contraste de durée sont regroupés avec l'élément proéminent final (typiquement les langues trochaïques comme le français). On retrouve donc les résultats de Hayes (1995, section 3.2.1).

Ramus donne un tableau récapitulatif de différentes langues selon la catégorie rythmique à laquelle elles ont été affiliées (Figure 13).

Langues	Références
Langues accentuelles	
Anglais	Pike (1945)
Arabe	Abercrombie (1967)
Néerlandais et les langues germaniques en général	Ladefoged (1975)
Russe	Abercrombie (1967)
Polonais	Rubach et Booij (1985)
Thaï	Luangthongkum (1977) cité par Dauer (1983)
Langues syllabiques	
Espagnol	Pike (1945)
Français	Abercrombie (1967)
Italien	Bertinetto (1981)
Catalan	Mehler et coll. (1993)
Grec	Arvaniti (1994)
Telegu	Abercrombie (1967)
Yoruba	Abercrombie (1967)
Hindi	O'Connor (1973), cité par Dauer (1983)
Langues moraïques	
Japonais	Ladefoged (1975)
Tamoul	Steever (1987)
Alyawarra, Aranda Et Paiute	Haraguchi et Tanaka (1992) cités par Nazzi (1997)

Figure 13. Regroupement des différentes langues selon leurs caractéristiques rythmiques (Ramus 1999 : 41)

Ce tableau ne sera pas exploité dans le détail mais il permet de voir les différents regroupements des langues étudiées dans la littérature en fonction de la catégorisation en langues à rythme accentuel, à rythme syllabique et à rythme moraïque.

4.4.2. *Échec de la catégorisation classique des langues en fonction du rythme*

Cette distinction rythmique des langues longtemps attestée ne fait plus l'unanimité car les études comparatives présentes dans la littérature n'ont pu confirmer que la distinction se faisait sur la base seule de l'existence de l'isochronie des intervalles accentuels, syllabiques et moraïques.

Roach (1982) reprend les deux conclusions d'Abercrombie que nous avons citées plus tôt et les teste sur de la parole spontanée. Cette expérience vise à catégoriser les différentes langues selon l'une ou l'autre des deux définitions. Tester ces deux hypothèses pose des problèmes techniques tels que la question du débit de parole, la délimitation des unités intonatives et la mesure des intervalles accentuels ou, plus exactement, la détermination du début des intervalles accentuels. A ce sujet, Roach décide d'opter pour le début phonologique de la syllabe qui porte l'accent. A l'intérieur des unités intonatives, l'auteur abandonne les syllabes inaccentuées initiales pour éviter le problème de pieds commençant par un battement silencieux, ainsi que les syllabes finales en raison de leur allongement final.

Les résultats ne confirment aucune des deux hypothèses d'Abercrombie : les syllabes n'ont pas tendance à avoir une durée égale dans les langues à rythme syllabique et les intervalles accentuels n'ont pas plus tendance à avoir une durée égale dans les langues à rythme accentuel. Il n'est donc pas possible de séparer les deux groupes de langues sur la base de la variation de la durée syllabique dans les langues à rythme accentuel et de l'irrégularité des battements accentuels dans les langues à rythme syllabique. Roach conclut que cette distinction entre les langues est en fait auditive et subjective : une langue se fonde sur la syllabe si elle "sonne" comme telle. Pour Roach, il s'agit peut-être simplement du fait que les langues à rythme syllabique ont tendance à avoir une structure syllabique moins complexe et que les langues à rythme accentuel ont tendance à voir leurs voyelles inaccentuées réduites. Cette hypothèse est confirmée par les travaux de Dauer (1983) sur lesquels nous reviendrons plus loin.

Wenk et Wioland (1982) et Wenk (1983) expliquent que syllabes ne sont pas isochrones en français : elles sont produites et perçues dans les groupes rythmiques, tout comme celles de l'anglais. Le rythme en français dépend plutôt d'unités rythmiques plus grandes telles que le syntagme phonologique. De plus, les expériences de Scott, Isard et Boysson-Bardies (1985) montrent que les sujets natifs anglais et français se comportent de façon similaire au niveau perceptif : les deux groupes entendent les intervalles accentuels comme plus réguliers qu'ils ne le sont réellement. Cette attitude similaire contredit toute distinction catégorique vis-à-vis de leur langue respective.

En ce qui concerne l'isochronie moraïque, alors que certains considèrent la more comme une unité avec une durée constante (Port, Al-ani et Maeda 1980), il existe là aussi peu de preuves expérimentales en faveur de la régularité temporelle de la more en japonais (Beckman 1992), résultats ensuite confirmés par Hoequist (1983a et b). De manière plus générale, Hoequist n'obtient aucune preuve en faveur de l'hypothèse de rythmicité accentuelle, syllabique et moraïque. L'auteur observe des tendances vers l'isochronie pour certaines unités mais aucune isochronie systématique, quelle que soit la langue observée.

Ces diverses observations remettent par conséquent en question la catégorisation classique des langues, du moins lorsque celle-ci se fonde sur l'isochronie des intervalles accentuels, syllabiques et moraïques. Aucune étude n'a pu confirmer l'hypothèse d'isochronie des unités temporelles (Laver 1994). Face à cet échec de catégorisation, de nouvelles approches sont proposées afin de mettre en évidence d'autres patrons rythmiques systématiques.

4.4.3. *Approches alternatives du rythme de la parole*

Précisons que la classification des langues n'est pas remise en question, simplement elle n'est plus envisagée sous l'angle des caractéristiques rythmiques, à savoir sur l'isochronie des intervalles accentuels ou des syllabes. La problématique de classification est désormais considérée de manière différente.

Bolinger (1981) et Dasher et Bolinger (1982) sont parmi les premiers à proposer une approche rythmique différente puisque le rythme n'est plus prédit à partir de la notion d'accent, mais à partir de celle de la qualité des voyelles. La durée des syllabes dépend de la quantité de la voyelle suivante en fonction de la Loi d'Emprunt (*Borrowing Rule*)

qui dit qu'une syllabe avec une voyelle réduite "emprunte" du temps à la syllabe précédente comportant une voyelle pleine. En conséquence, le rythme représente une propriété dérivée d'autres propriétés phonologiques de la langue. L'accentuation et le rythme sont le résultat de la présence d'une voyelle pleine (ou réduite) et puisque la qualité de la voyelle est donnée, elle est utilisée pour prédire le patron accentuel. A partir de cette approche, les auteurs considèrent que les deux types de rythme ne sont pas diamétralement opposés mais peuvent coexister dans une même langue (Bolinger 1981 pour l'anglais ; Wenk et Wioland 1982, Astésano 2001 pour le français). Pour Bolinger, le rythme syllabique relève de l'agencement temporel des voyelles pleines et des voyelles faibles alors que le rythme accentuel est caractérisé par les patrons de fréquence fondamentale. Wenk et Wioland défendent le même point de vue, mais dans le cas du français.

Les travaux de Dauer (1983) représentent un grand tournant dans les études portant sur la classification des langues puisque ses résultats sont à l'origine de nombreuses autres études (Hoequist 1983b ; Wenk 1983 ; Bertinetto 1989 ; Nespor 1990 ; Fant, Kruckenberg et Nord 1991a et b ; Cutler 1991 ; Eriksson 1991). Dauer s'appuie sur les travaux de Roach (1982) et suppose que la différence entre les langues ne réside pas dans la durée des intervalles, de même que les accents ne reviennent pas plus régulièrement dans les langues comme l'anglais que dans des langues comme l'espagnol. La différence est plutôt liée à la composition interne de l'intervalle, à savoir sa durée et le nombre de syllabes qu'il contient. Plus spécifiquement, les langues du monde varient selon trois dimensions.

1. LA STRUCTURE SYLLABIQUE

La durée des syllabes varie considérablement dans les langues à rythme accentuel parce que la composition syllabique est plus complexe. En français et en espagnol, la syllabe la plus fréquente est de forme CV alors que les combinaisons sont nettement plus variées et surtout complexes en anglais, avec la conséquence directe d'une augmentation de la durée. Black et Singh (1968 : 118) observent en effet que 70% des syllabes en italien sont de type CV contre seulement 20.3% en anglais. La complexité syllabique combinée à la composition segmentale accroissent la distinction entre les syllabes accentuées et inaccentuées, c'est notamment le cas en anglais.

2. LA REDUCTION DES VOYELLES

Elle se présente la plupart du temps sous la forme de centralisation vocalique et représente un autre moyen de maximiser la différence entre les syllabes accentuées et les syllabes inaccentuées dans les langues à rythme accentuel (Cutler 1991). Ceci représente d'ailleurs une des difficultés majeures des sujets français lors de l'apprentissage de l'anglais car la réduction des voyelles inaccentuées est absente de leur langue maternelle. La différence de réduction est approfondie et illustrée par Wenk (1983) : en fonction de la caractéristique de "tête à droite" ou "tête à gauche" des pieds (Hirst et Di Cristo 1984), l'articulation des syllabes inaccentuées diffère. Wenk illustre les différences de tension au moyen d'une courbe rythmique (Figure 14 ci-après).

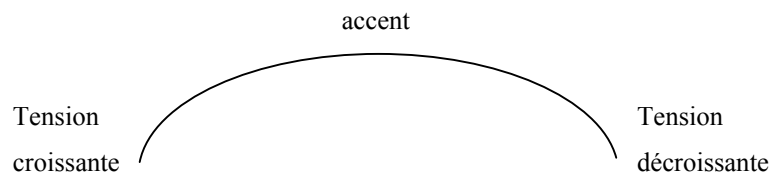


Figure 14. Courbe rythmique représentant les différences de tension dans le cadre de la syllabe (Wenk 1983: 157)

Les syllabes accentuées sont toujours situées au maximum de la tension articulatoire. En revanche, dans les langues codachrones comme le français (pied avec tête à droite), les syllabes inaccentuées se situent au début de la courbe rythmique avec par conséquent une tension articulatoire croissante (d'où le caractère tendu de ces syllabes). Dans les langues capochrones (pied avec tête à gauche), les syllabes inaccentuées se situent en fin de courbe rythmique avec un relâchement de la tension musculaire (d'où leur réduction).

Cutler (1991) précise que cette différence majeure entre les deux langues reflète deux stratégies fondamentalement différentes de segmentation de la chaîne parlée : les sujets français segmentent la parole de syllabe en syllabe, alors que la segmentation des sujets anglophones se fonde sur l'opposition entre syllabes fortes et syllabes faibles.

3. L'IMPORTANCE DE L'ACCENT

Plus l'accent a de l'importance dans une langue, plus celle-ci aura tendance à appartenir à la catégorie des langues à rythme accentuel. Dans ces langues, l'accent est caractérisé

par une combinaison d'allongement, de modification du contour mélodique, de changement de hauteur et de qualité ; il est donc plus saillant perceptiblement que dans les langues à rythme syllabique. Les langues sont donc plus ou moins accentuelles selon un continuum (cf. également Hoequist 1983a et b). Les langues sont ainsi localisées sur un axe représentant le degré d'importance de l'accent (Dauer 1983, ensuite repris par Nespor 1990). Ce continuum est illustré dans la Figure 15 ci-après.

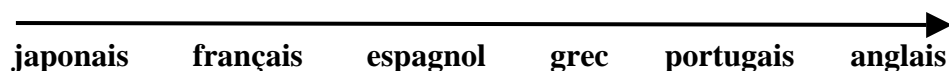


Figure 15. Axe représentant le degré d'importance de l'accent dans différentes langues (Dauer 1983: 60)

L'allongement accentuel va de pair avec cet argument car l'effet de l'accent a tendance à être plus marqué dans les langues dites accentuelles. Fant, Kruckenberg et Nord (1991a et b) précisent que la distinction entre les langues ne réside pas dans l'espacement régulier des syllabes ou des accents, mais dans l'importance perceptive des syllabes lourdes accentuées. Nous avons tendance à percevoir leurs successions comme quasi-périodiques de par leur proéminence perceptive bien plus marquée que celle des syllabes inaccentuées (cf. également Faure, Hirst et Chafcouloff 1980). En l'absence d'une telle importance perceptive des syllabes accentuées, la sensation auditive penchera vers les langues à rythme syllabique.

Les langues ne sont donc plus placées dans des groupes fermés mais situées sur un continuum avec aux extrêmes les langues à rythme accentuel et les langues à rythme moraique, les langues à rythme syllabique se trouvant le long du continuum en fonction de leurs caractéristiques phonétiques et phonologiques (Dauer 1983). Ceci est en accord avec la remarque de Delattre (1966) selon laquelle les différentes langues ne doivent pas être classées catégoriquement mais situées sur un continuum de valeurs contrastées.

Hoequist (1983b) propose un critère supplémentaire de classification des langues, celui d'effet d'adjacence. La durée des syllabes accentuées est fortement influencée par les syllabes inaccentuées adjacentes, mais uniquement en anglais (voire en espagnol, mais à moindre degré). Cet effet d'adjacence est similaire au continuum de Dauer puisqu'on a le cas des langues avec effet d'adjacence très marqué à un extrême du

continuum, puis le cas des langues avec absence d'effet d'adjacence à l'autre extrême. Hoequist propose ainsi de distinguer les langues selon leur façon stricte ou pas d'agencer leur durée, et pas uniquement selon le degré d'importance de l'accent. Par exemple, le japonais est caractérisé par un contrôle strict de la durée des mores, il est d'ailleurs qualifié de "langue à durée contrôlée" (*duration-controlling language*). L'anglais est lui caractérisé par un contrôle au niveau du pied et donc d'un degré important d'influence des syllabes inaccentuées sur la durée des syllabes accentuées (effet d'adjacence) ; c'est une "langue à durée compensée" (*duration-compensating language*). L'opposition est par conséquent binaire et toutes les langues situées entre ces deux pôles partagent des caractéristiques avec l'une ou l'autre catégorie. La classification ne se situe donc pas sur un continuum unique mais sur une combinaison de continuums.

Bertinetto (1989) présente une synthèse étendue des facteurs permettant de distinguer les diverses langues (1989 : 108) :

- (a) Réduction vocalique OU articulation pleine des syllabes inaccentuées
- (b) Certitude OU incertitude dans le décompte des syllabes
- (c) Accélération du tempo par la compression des syllabes inaccentuées OU compression proportionnelle
- (d) Structure syllabique complexe avec incertitude relative des frontières syllabiques OU structure syllabique simple avec des frontières bien définies
- (e) Tendance aux syllabes lourdes sous l'accent OU absence de cette tendance
- (f) Flexibilité dans le placement de l'accent OU rigidité
- (g) Densité relative des accents secondaires OU pas.

Ces différents facteurs se regroupent avec ceux de Dauer, notamment les facteurs (a), (d) et (f). On ne parle donc plus d'opposition dichotomique mais plutôt d'orientation scalaire. Les langues sont plus ou moins caractérisées par les facteurs et selon le degré d'importance de chacun de ces sept facteurs, les langues possèdent plus ou moins les caractéristiques de rythme iso-accentuel ou iso-syllabique.

Nespor (1990) ajoute également un argument supplémentaire, celui de l'existence de langues dont les caractéristiques rythmiques sont intermédiaires entre celles des langues accentuelles et syllabiques. Les langues comme le catalan, le polonais et le portugais (tout particulièrement le portugais brésilien) partagent des propriétés phonologiques généralement attribuées à l'une et l'autre catégorie. Vogel ajoute également comme

facteur le déplacement accentuel lorsque deux syllabes accentuées sont adjacentes (Vogel, Bunnell et Hoskins 1995) ; un tel déplacement existe en anglais et en italien (lorsque les deux syllabes sont strictement adjacentes) mais pas dans d'autres langues. Nespor en conclut que la différence entre les langues réside dans des propriétés phonologiques non rythmiques plutôt que dans des caractéristiques rythmiques.

A partir de ces différentes remarques, Eriksson (1991) se fonde sur la proposition de Faure, Hirst et Chafcouloff (1980) d'un modèle dans lequel la durée des intervalles dépend linéairement du nombre de syllabes dans l'intervalle, avec une durée constante ajoutée qui représente l'allongement de la syllabe accentuée. Eriksson teste ensuite ce modèle sur les données de Dauer ; la prédiction de son modèle est quasi parfaite : la durée de l'intervalle semble augmenter à la même vitesse pour les différentes langues comparées et la seule différence semble résider dans le terme constant lié à la syllabe accentuée (200ms pour l'anglais et la langue thaï, 100ms pour le français, l'espagnol et l'italien). La durée de l'intervalle augmente en fonction du nombre de syllabes à la même vitesse pour toutes les langues, mais la durée totale de l'intervalle dépend des différences de durée des syllabes accentuées propres aux langues.

Notons enfin qu'à partir d'une méthodologie différente, Grabe et Low (2002) confirment l'absence de distinction stricte entre les langues dites iso-accentuelles, isosyllabiques et moraïques mais plutôt un chevauchement entre les groupes. Les auteurs calculent pour dix huit langues la variabilité temporelle dans des intervalles acoustico-phonétiques successifs en utilisant le *PVI (Pairwise Variability Index)* qui indique le niveau de variabilité dans les mesures successives. L'originalité de cette approche est de ne pas considérer la durée des unités phonologiques (phonèmes et/ ou syllabes) mais plutôt la durée des voyelles et des intervalles entre les voyelles (en excluant les pauses). Les résultats confirment un chevauchement des catégories.

Il ressort de ces différentes présentations que la catégorisation des langues du monde sur la base de l'isochronie accentuelle, syllabique ou moraïque ne tient pas. Il existe bien des différences entre ces types de langues, mais ces différences ne sont pas catégoriques, elles sont plutôt représentées par un continuum. Toutes ces approches s'accordent sur les caractéristiques permettant de distinguer les langues portant sur l'influence du phénomène accentuel sous la forme d'allongements et de réductions vocaliques, des différences sur le plan des structures syllabiques (plus complexes dans

les langues comme l'anglais), et des effets de raccourcissement sous l'effet de syllabes inaccentuées précédentes et suivantes. Par conséquent, la distinction repose sur des différences phonétiques et phonologiques plutôt que sur l'isochronie accentuelle ou syllabique.

4.4.4. Exemple de différence rythmique intra-langue

Les perspectives de recherche concernant les différences inter-langues se sont au fur et à mesure développées et certains se sont penchés sur la comparaison des caractéristiques à l'intérieur d'une même langue. En anglais par exemple, Deterding (2001) a montré qu'il existe une différence rythmique fondamentale entre l'anglais britannique standard et l'anglais de Singapour : les syllabes réduites, le plus souvent sous la forme de centralisation vocalique vers le phonème schwa, sont plus fréquentes en anglais britannique et cette tendance contribue à la différence entre les deux types de rythme. En revanche, les durées syllabiques varient considérablement entre les deux variétés, d'où l'appariement de l'anglais de Singapour à la catégorie des langues à rythme syllabique, contrairement à l'anglais britannique. C'est également le cas pour l'anglais de Hawaï qui est apparenté aux langues à rythme syllabique (Bolinger 1981).

La méthode de Grabe et Low (2002) est également appliquée sur de l'anglais britannique et de l'anglais de Singapour et les résultats obtenus abondent dans le sens de ceux de Deterding : on observe plus de variabilité vocalique en anglais britannique. Les deux langues réduisent les voyelles inaccentuées mais les voyelles réduites en anglais de Singapour sont moins centralisées et plus longues que celles de l'anglais britannique. Notons que malgré ces divergences, l'anglais de Singapour reste éloigné des langues isosyllabiques.

4.4.5. Identification automatique des langues à partir des particularités rythmiques

Ramus (1999) et Ramus, Nespor et Mehler (1999) présentent un modèle dont le but est d'expliquer comment se fait l'apprentissage des langues par les nouveau-nés. Les auteurs suggèrent que le type de rythme auquel chaque langue appartient aide les nouveau-nés à acquérir un certain nombre de propriétés phonologiques. Globalement, Ramus postule que le nouveau-né perçoit les caractéristiques rythmiques dans le signal grâce auquel il a accès à la structure syllabique, puisque chaque catégorie de langue est

caractérisée par un certain degré de complexité syllabique. L'approche adoptée ici se fonde sur les études de Roach (1982), Dasher et Bolinger (1982) et Dauer (1983).

Rappelons que pour Dauer, la distinction entre les langues réside dans ¹⁾ la structure syllabique, ²⁾ la réduction des voyelles inaccentuées et ³⁾ l'importance de l'accent. Selon Ramus, c'est la combinaison des deux premiers qui donne l'impression de syllabes beaucoup plus saillantes dans les langues à rythme accentuel, du fait même que leurs syllabes accentuées sont plus complexes et possèdent des voyelles non réduites (par opposition aux voyelles inaccentuées). Ces différentes remarques permettent de percevoir le rythme d'une façon tout à fait différente et Ramus se fonde sur cette nouvelle conception pour développer son modèle rythmique.

Le but global est de quantifier les propriétés phonologiques de chaque langue afin de reconnaître et de catégoriser son rythme (Ramus 1999 ; Ramus, Nespor et Mehler 1999). Le but ultime de ces travaux est de formuler des hypothèses sur l'acquisition du langage : montrer que le nouveau-né peut discriminer les types de langue sur la base de leur rythme. Huit langues sont ici observées (Ramus 1999) : trois à rythme accentuel (l'anglais, le néerlandais et le polonais), quatre à rythme syllabique (le français, l'espagnol, le catalan et l'italien) et une à rythme moraique (le japonais).

Ramus teste dans un premier temps l'hypothèse de catégorisation des langues ; pour cela, il segmente manuellement son corpus en intervalles vocaliques et consonantiques. Cette segmentation découle de l'hypothèse que les nouveau-nés distinguent uniquement les consonnes des voyelles, plutôt que chacun des phonèmes de la langue. En découlent trois variables :

- $%V$: proportion des intervalles vocaliques dans la phrase
- ΔV : écart type des durées d'intervalles vocaliques dans la phrase
- ΔC : écart type des durées d'intervalles consonantiques dans la phrase

Les deux variables $%V$ et ΔC permettent d'obtenir les classes traditionnelles de rythme : les langues à structure syllabique plus complexe ont un $%V$ plus faible et un ΔC plus élevé par rapport aux langues à structure syllabique simple. Une série d'expériences sous forme de tâches de catégorisation d'énoncés avec entraînement confirment les trois classes obtenues plus tôt, et ceci à la fois pour les sujets adultes et les nouveau-nés. Ramus obtient des résultats parallèles à ceux de Dasher et Bolinger et de Dauer en utilisant les intervalles vocaliques et consonantiques. Par conséquent, son modèle de perception fondé sur la variable $%V$ permet de simuler les expériences de discrimination

des langues. Ce modèle de perception du rythme s'appuie sur la supposition que le cerveau segmente la parole en intervalles vocaliques et consonantiques, pour ensuite déduire le type de rythme en fonction de leur durée.

Dans un deuxième temps, Ramus effectue des tests de discrimination entre différentes paires de langues, en partant cette fois d'énoncés synthétisés grâce à l'algorithme MBROLA (Dutoit, Pagel, Pierret, Bataille et van der Vrecken 1996). L'originalité de son approche est qu'il synthétise des phrases sous quatre formes différentes :

La forme *saltanaj* : chaque phonème est remplacé par un autre phonème de même mode d'articulation (les fricatives sont remplacées par [s], les voyelles par [a], les liquides par [l], les occlusives par [t], les nasales par [n] et les semi-voyelles par [j]). On conserve ici les caractéristiques phonotactiques, l'intonation et le rythme.

La forme *sasasa* : toutes les consonnes sont remplacées par [s] et les voyelles par [a]. On possède ici les informations intonatives et rythmiques, mais plus celles relatives aux règles phonotactiques.

La forme *sasasa plat* : on conserve la forme *sasasa* mais la fréquence fondamentale est fixée à 230Hz pour tous les énoncés. Le rythme est désormais le seul indice présent.

La forme *aaaa* : tous les phonèmes sont remplacés par [a], la courbe de F0 est conservée. Le seul indice est désormais celui de l'intonation.

Ces phrases synthétisées permettent d'isoler certains indices, tout particulièrement celui du rythme. Au cours des tests de discrimination entre l'anglais et le japonais, les sujets français (natifs) distinguent très bien les deux langues dans les trois premières tâches mais pas dans la tâche *aaaa* (intonation seulement) ; le rythme est par conséquent un bon indice de discrimination entre l'anglais et le japonais. Ramus réalise la même expérience avec des sujets anglophones : cette fois, la discrimination s'effectue également dans la tâche *aaaa* : il faut donc que les sujets soient natifs dans une des deux langues pour pouvoir discriminer l'anglais et le japonais sur la base de l'intonation.

Ramus s'intéresse ensuite aux deux langues intermédiaires que sont le polonais et le catalan. Il les compare aux deux langues de référence, l'anglais et l'espagnol, grâce aux tâches *sasasa* et *sasasa plat*. Les trois paires de langue testées sont anglais/espagnol, espagnol/polonais et polonais/anglais. En contexte *sasasa* (intonation et rythme), les

sujets distinguent uniquement les paires anglais/espagnol et espagnol/polonais, alors qu'en contexte *sasasa plat* (rythme uniquement), ils distinguent les trois paires. Par conséquent, le polonais ne s'apparente à aucune des deux catégories – notons qu'il se distingue de l'anglais en terme de rythme seulement. Ainsi, Ramus suppose que, dans le contexte *sasasa*, l'intonation a gêné la perception du rythme ; ces résultats confirment le caractère intermédiaire du polonais. En revanche, les sujets discriminent le catalan de l'anglais et du polonais, mais pas de l'espagnol. Le catalan n'est donc pas intermédiaire : il s'apparente aux langues à rythme syllabique.

Ces résultats confirment donc les trois classes de rythme, mais Ramus postule l'existence d'une quatrième classe comprenant le polonais (Nespor 1990). On peut toutefois reprocher à ces travaux de ne se fonder que sur les enregistrements d'un seul locuteur et de ne porter que sur de la parole contrainte (dite de laboratoire) ; on peut se demander si les résultats seraient identiques avec de la parole naturelle.

Farinas (2002) présente également un modèle d'Identification Automatique des Langues (IAL) à partir des caractéristiques rythmiques. La première étape de ce modèle est le découpage automatique du signal en "pseudo-syllabes", celles-ci correspondant au regroupement des portions non-vocaliques avec les portions vocaliques détectées automatiquement dans le signal. L'avantage technique évident vis-à-vis de l'approche de Ramus est l'automatisation de la tâche. Les paramètres rythmiques pris en compte sont la durée totale des segments consonantiques (D_C), la durée totale des segments vocaliques (D_V) et la complexité de la pseudo-syllabe en comptant le nombre de segments consonantiques (N_C). Ces trois paramètres tiennent compte du critère de distinction de complexité syllabique, mais également du degré différent de l'importance de l'accent au travers des allongements différentiels des pseudo-syllabes et des segments consonantiques et vocaliques à l'intérieur de cette unité.

Le modèle est évalué sur le corpus lu multilingue MULTEXT, composé de cinq langues différentes que sont l'anglais, l'allemand, le français, l'italien et l'espagnol. Le taux d'identification correcte augmente en fonction de la durée des énoncés utilisés ; pour un énoncé de 21 secondes, le taux d'identification correcte est de 78%. Notons que la matrice de confusion révèle de nombreuses confusions entre l'anglais, l'allemand et l'italien et entre l'italien et l'espagnol. En revanche, cette matrice confirme la distinction des langues à rythme accentuel et à rythme syllabique.

4.5. Résumé et conclusion

Dans ce quatrième chapitre, nous avons abordé la thématique de l'isochronie. Il ressort des différentes études citées que l'hypothèse de l'isochronie stricte ne tient pas : il n'existe aucune preuve physique de l'égalité absolue des intervalles accentuels en anglais, des syllabes en français et des mores en japonais. En anglais, la durée des intervalles accentuels est influencée par de nombreux facteurs qui font que lors de la phase de production de ces intervalles, leur durée ne peut être parfaitement égale. Cependant, malgré les déformations dues aux contraintes de production, la parole est perçue comme régulière. Il existe par conséquent un décalage entre la production des énoncés et la perception qu'en ont les interlocuteurs : il ne semble donc pas nécessaire que la parole soit parfaitement isochrone pour être perçue comme rythmique.

Les suppositions concernant l'isochronie ont conduit à la catégorisation des langues du monde en fonction du domaine concerné par l'isochronie. Il résulte de cette approche une triple distinction des langues, selon que leurs caractéristiques rythmiques soient iso-accentuelles, iso-syllabiques ou iso-moraïques. Cette hypothèse a fait l'objet de nombreux travaux et de nombreuses études ont montré que la catégorisation des langues sur la base de l'isochronie ne tenait pas. La distinction rythmique entre les diverses langues repose plutôt sur une différence au niveau de la complexité de la structure syllabique, de la réduction vocalique du noyau des syllabes inaccentuées et du degré d'importance des syllabes accentuées.

C'est d'ailleurs à partir de ces paramètres que se fait la distinction des modèles d'identification automatique des langues. Il convient également de préciser que la différenciation n'est plus binaire mais se représente plutôt sur un continuum avec aux deux extrêmes d'une part les langues avec une opposition forte entre les syllabes accentuées et inaccentuées (comme l'anglais) et d'autre part les langues dans lesquelles cette opposition est peu marquée (comme le japonais). Ce continuum permet la reconnaissance de certaines langues dont les caractéristiques sont intermédiaires entre ces deux extrêmes comme par exemple le polonais (Nespor 1980). L'hypothèse de différenciation des langues sur la base des caractéristiques rythmiques n'est pas directement liée à la problématique générale de ce travail, mais nous intéresse particulièrement car tester ces différents paramètres de distinction renseigne sur la structuration phonologique et rythmique de la langue anglaise ainsi que sur les répercussions potentielles au niveau temporel.

Dans le cadre de ce travail, il est important de souligner que la majorité des travaux présentés ici testent l'hypothèse de l'isochronie sur l'intervalle accentuel délimité par les syllabes accentuées, à savoir le pied abercrombien. Or, si on suit l'approche de Jassem, l'isochronie faible porte sur la NRU, pas sur la TRU, encore moins sur l'unité tonale, unité équivalente au pied accentuel ; ceci signifie que pour observer un degré faible d'isochronie dans la parole, il faut exclure les anacrouses. Cette piste doit être explorée avant de rejeter catégoriquement tout phénomène d'isochronie (strict ou faible) en production, tout comme Jassem, Hill et Witten (1984) l'avaient d'ailleurs montré. Il se peut que la régularité existe bien en production, mais à un niveau différent de celui de l'intervalle accentuel.

Cette première partie théorique nous a permis d'introduire notre sujet portant sur le rythme et la structuration prosodique de l'anglais britannique contemporain. Deux problématiques sont soulevées dans ce travail, celle de l'ambisyllabité (et de syllabification) et celle de la définition de l'unité rythmique. Elles relèvent d'un cadre de recherche plus global représenté par la prédiction de l'organisation temporelle de la parole.

Nous nous sommes dans un premier temps concentrée sur la problématique de l'ambisyllabité, procédé qui nous semble intéressant mais qui, à notre connaissance, n'a pas été testé statistiquement, notamment sur de la parole naturelle. Nous avons ensuite abordé la question de la structuration rythmique de l'anglais ; nous avons décrit principalement deux modèles rythmiques proposés pour l'anglais britannique, le modèle classique d'Abercrombie/ Halliday et celui de Jassem. La différence entre les deux réside dans le regroupement différent des syllabes en unités rythmiques, donnant donc des prédictions de durée différentes.

Cette deuxième partie est consacrée aux tests des deux problématiques d'ambisyllabité et de définition de l'unité rythmique avec au préalable une présentation de la base de données utilisée pour ces différents tests. Nous allons donc présenter tout d'abord la base de données Aix-MARSEC utilisée pour la partie expérimentale (Chapitre 5). Nous testerons ensuite les deux modèles rythmiques de

Jassem et d'Abercrombie/ Halliday sur Aix-MARSEC (Chapitre 6). Nous nous intéresserons enfin à la question d'ambisyllabité au travers de divers tests statistiques sur Aix-MARSEC et d'une expérience en production (Chapitre 7).

PARTIE II.

BASE DE DONNEES ET RESULTATS EXPERIMENTAUX

CHAPITRE 5.
CONSTITUTION ET EXPLOITATION
D'UNE BASE DE DONNEES ORALES : PROJET AIX-MARSEC

5.1. QUEL(S) CORPUS ?	179
5.1.1. Types de corpus.....	180
5.1.2. Qu'entend-on par parole "spontanée" ?.....	183
5.1.3. Transcription et annotation de corpus oraux	185
5.1.4. Disponibilité des corpus	187
5.2. DESCRIPTION DU CORPUS	188
5.2.1. Le corpus d'origine : du SEC au MARSEC	188
5.2.2. Homogénéisation des données et pré-traitements	190
5.3. TRANSCRIPTION PHONETIQUE DU CORPUS	190
5.3.1. Phonétisation brute.....	191
5.3.1.1. Méthodologie	191
5.3.1.2. Traitements spécifiques.....	192
5.3.2. Optimisation de la phonétisation.....	195
5.3.2.1. Méthodologie	195
5.3.2.2. Règles d'élision de phonèmes	197
5.3.2.3. Evaluation de l'efficacité des règles d'élision.....	199
5.4. ALIGNEMENT AUTOMATIQUE DES PHONEMES.....	200
5.4.1. Méthodologie	200
5.4.2. Evaluation et perspectives d'amélioration de l'alignement automatique..	201
5.5. AUTRES NIVEAUX D'ANALYSE.....	202
5.5.1. Regroupement en syllabes et en constituants syllabiques.....	202
5.5.2. Regroupement des syllabes en unités rythmiques.....	204
5.5.3. MOMEL & INTSINT	206
5.5.4. Aix-MARSEC : une base de données pluri-linéaire	207
5.6. MISE EN FORME DES DONNEES.....	209
5.6.1. Description des divers paramètres.....	209
5.6.2. Remarques sur les découpages syllabiques.....	213
5.7. RESUME ET CONCLUSION.....	214

Pour tester les hypothèses soulevées dans la première partie, il est nécessaire d'utiliser un corpus oral d'anglais britannique suffisamment important pour pouvoir tester les hypothèses et avec un minimum d'annotation et d'alignement. Or, peu de corpus sont actuellement facilement disponibles en terme d'accès ou de coût. De plus, peu d'entre eux possèdent un traitement phonétique suffisamment fin pour être exploités lors d'études statistiques ; par traitement fin, nous entendons une transcription et un alignement au niveau phonémique.

Dans la première partie de ce chapitre, nous approfondirons la définition du terme corpus et des divers styles de parole qu'il peut recouvrir ; nous aborderons ensuite la question de la disponibilité des corpus en anglais. Nous présenterons par la suite la base de données choisie dans cette étude, ainsi que tous les traitements effectués sur ce corpus ; il s'agit ici de la transcription phonétique de l'ensemble du corpus, de l'alignement automatique au niveau du phonème et enfin du regroupement automatique des phonèmes en unités de niveau supérieur. Nous verrons dans la dernière section la mise en forme de ces données sous forme de tableaux, nécessaire à leur exploitation statistique.

5.1. Quel(s) corpus ?

Le choix du corpus dans le cadre d'une étude expérimentale est très important, notamment si on considère la variété des corpus existants. Ils peuvent être de nature très diverse et le choix dépend de l'orientation de l'étude en cours. On peut s'intéresser à des corpus écrits et/ ou oraux selon l'objet d'observation et de l'objectif du travail : par exemple, un syntacticien et un phonéticien n'exploiteront pas les mêmes types de corpus ou bien ils utiliseront les mêmes corpus mais à des niveaux de représentation et

d'annotation différents. Dans cette description des différents types de corpus, nous nous intéressons uniquement aux corpus oraux, mais comme nous allons le voir, la nature des corpus oraux est elle-même très variée.

5.1.1. Types de corpus

Dans les études en phonologie et en phonétique, on distingue généralement deux types de corpus : les corpus de parole de laboratoire et les corpus de parole spontanée (nous reviendrons sur ce terme dans la section suivante). La parole de laboratoire, comme son nom l'indique, correspond à des enregistrements effectués en laboratoire, le plus souvent dans une chambre sourde afin que la qualité du signal sonore soit optimale. Les stimuli correspondent généralement à des listes de mots ou de phrases préparées au préalable par l'expérimentateur. Le premier avantage de ce type de corpus est la qualité des enregistrements, primordiale notamment dans le cadre de travaux en acoustique (comme sur la fréquence fondamentale ou l'intensité) ; le deuxième atout réside dans le choix des stimuli qui permettent à l'expérimentateur de tester spécifiquement certains phénomènes tout en contrôlant d'autres paramètres pour en neutraliser les effets. En revanche, le désavantage majeur souvent encouru avec ce type de corpus est le manque de naturel de la parole ainsi produite ; il ne s'agit pas de parole produite naturellement mais de parole contrainte à but non-communicatif et le résultat peut parfois être peu naturel, voire très artificiel.

La deuxième grande catégorie de corpus est constituée de parole authentique, produite dans des conditions naturelles de production, par opposition à de la parole contrainte de laboratoire. Ceci signifie d'une part que les enregistrements sont effectués dans des conditions naturelles de production et de communication (contrairement à la parole de laboratoire recueillie dans une chambre sourde) et d'autre part qu'ils ne contiennent pas de matériel préparé au préalable par l'expérimentateur. Les méthodes pour obtenir de tels corpus sont très variables ; on peut par exemple organiser des jeux de rôle ou des débats entre plusieurs personnes sur un thème choisi au préalable et enregistrer les sujets qui débattent sur la question. De tels enregistrements présentent l'avantage de pouvoir être enregistrés dans une chambre sourde et de recueillir par conséquent des données de bonne qualité. En revanche, le risque est le chevauchement de voix, difficilement exploitable par la suite. Il est ainsi recommandé d'utiliser des

micro-casques permettant d'obtenir chaque voix sur une piste séparée plutôt que de faire un enregistrement global des locuteurs ; toutefois, on se retrouve rapidement dans un environnement artificiel pour les locuteurs qui peut alors les gêner dans leur production. Malgré l'avantage évident du caractère naturel de la parole, travailler sur ce type de corpus pose d'autres problèmes d'ordre technique tels que la difficulté d'avoir un signal limpide dans la perspective de traitements acoustiques et de traitements automatiques.

Afin de recueillir des données orales authentiques, Brown, Anderson, Shillcock et Yule (1984) proposent la méthode du *Map Task* : deux sujets, se connaissant très bien ou pas (variation du degré de familiarité entre les sujets), sont placés face à face. Ils possèdent chacun une carte géographique mais avec des divergences (présence/ absence de certains repères, changements de noms, modification des distances, *etc.*). Sur l'une de ces cartes se trouve un trajet tracé et l'autre sujet doit reporter ce trajet sur sa propre carte. Les sujets sont ainsi amenés à débattre sur les trajets possibles sur leur carte (différente de l'autre), même s'ils sont conscients qu'ils sont enregistrés, ils retombent rapidement en situation de communication naturelle.

Un autre moyen d'obtenir des corpus oraux consiste à obtenir des enregistrements radiophoniques ; ceux-ci sont mixtes dans le sens où il peut s'agir d'interviews et donc se rapprocher d'un style parfaitement naturel. En revanche, certains corpus sont beaucoup moins naturels car ils se rapprochent plus de la lecture que de la production authentique. Entre ces deux extrêmes, on trouve des enregistrements qualifiés de naturels car ils sont perçus comme tels, c'est-à-dire qu'ils sont préparés mais non lus et la production résultante reste donc authentique.

Les études en phonétique, notamment en prosodie, se tournent de plus en plus vers des corpus de type naturel plutôt que de parole de laboratoire dans le but de faire ressortir les caractéristiques authentiques de la production langagière. Ainsi, ces études se tournent plutôt principalement vers la parole naturelle et de moins en moins vers la parole de laboratoire, jugée trop artificielle et non représentative des caractéristiques authentiques de la parole. Notons que malgré les différentes remarques que nous venons de faire au sujet de la parole de laboratoire, elle n'est en aucun cas discréditée ; elle est bien souvent la seule solution possible pour mettre en relief certains phénomènes. De plus, certaines études utilisent les deux types de données orales et combinent la lecture de mots (et/ ou de textes) et la conversation libre ; c'est notamment le cas du Projet du

Français Contemporain *PFC* (Durand, Laks et Lyche 2002) et du projet équivalent sur l'anglais *PAC* (Phonologie de l'Anglais Contemporain ; Durand et Pukli, à paraître), mais également de *IViE* (*Intonational Variation in English* ; Grabe et Post 2002). Le but de *PFC* et de *PAC* est de recueillir un ensemble d'enregistrements de variétés du français et de l'anglais. Les deux projets suivent le même protocole avec quatre types d'enregistrement : lecture de mots, lecture de texte (reprenant un maximum de mots de la liste), conversation guidée par l'expérimentateur et conversation libre. *IViE* regroupe environ 36 heures d'enregistrements de neuf régions des îles britanniques suivant cinq styles différents. Les enregistrements sont de trois types : des lectures de phrases et une lecture d'un texte (Cendrillon), de la parole semi-spontanée (reformulation du contenu du texte juste lu) et des interactions de type *Map Task*. Ces trois projets ont pour but de faciliter les études comparatives entre différentes variétés régionales de l'anglais et du français. Le but de tels corpus est d'offrir des données orales variées et complémentaires dans le but d'être exploitables par des expérimentateurs venant d'horizons différents.

Comme nous venons de le souligner, la nature des corpus oraux est très variée. C'est dans une perspective descriptive qu'Eskénazi (1993) distingue le 'type de discours' du 'style de parole', distinction terminologique importante dès lors que se pose la question du choix de corpus dans une étude expérimentale. Le premier correspond à la tâche discursive du locuteur lors de l'enregistrement du corpus, tels que la lecture, les monologues, les dialogues, *etc.* En revanche, le 'style de parole' est défini par la situation de communication, à savoir le contexte dans lequel se trouvent les interlocuteurs. Ce critère englobe tout un ensemble d'informations relatives au degré de familiarité, du contexte social, de la production spontanée ou non spontanée, *etc.*

Afin de représenter les divers styles de parole, Eskénazi regroupe ces différents paramètres autour de trois axes graduels que sont le degré d'intelligibilité, la familiarité entre les protagonistes et l'appartenance sociale (Figure 16 ci-après). Ces trois axes définissent un espace tridimensionnel permettant de représenter les différents styles de parole. Les styles de parole ne sont pas des entités discrètes et le passage graduel d'une catégorie à l'autre peut être représenté grâce à cet espace déterminé par trois continuums. Il permet donc de représenter les différents styles de parole et ainsi d'observer les différents effets phonétiques et phonologiques dus aux passages d'un style à un autre.

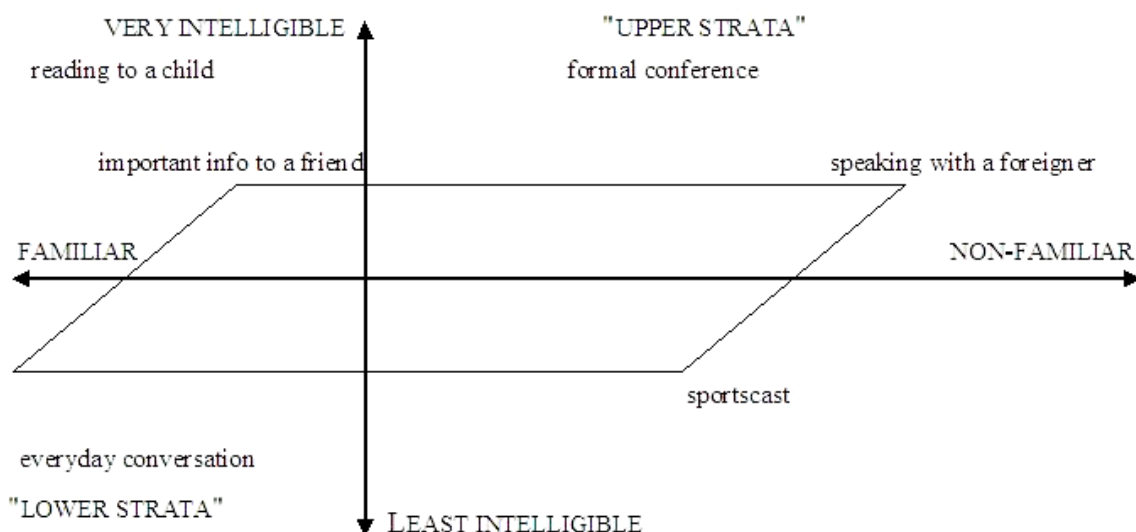


Figure 16. Représentation tridimensionnelle des divers styles de parole en fonction du degré d'intelligibilité, de la familiarité et de l'appartenance sociale (Eskénazi 1993: 503)

5.1.2. Qu'entend-on par parole "spontanée" ?

Cette brève présentation des catégories de corpus oraux et de styles de parole nous amène directement vers une question souvent épineuse, celle de la "spontanéité de la parole". Bon nombre d'études font référence à la parole spontanée pour décrire la parole naturelle et opposent souvent catégoriquement la parole spontanée à la parole de laboratoire, aboutissant donc sur une distinction binaire stricte. Or, plutôt que de les opposer de manière catégorique, il est préférable de parler d'échelle ou de degré de spontanéité constituant un continuum avec aux extrémités d'une part la parole spontanée et d'autre part la parole contrainte.

Le terme "spontané" est devenu trop imprécis et recouvre des styles très différents ; ce qui est considéré comme spontané pour certains ne le sera pas pour d'autres, et inversement. Nous utiliserons de préférence le terme 'authentique' ou 'naturel', à savoir un type de parole non contraint dont le but est de communiquer un message à une ou plusieurs autres personnes. La notion d'authenticité dépend ici de critères perceptifs puisque la parole est qualifiée d'authentique lorsqu'elle est *perçue* comme telle. On obtient par conséquent un continuum défini selon le critère de l'intention de communiquer un message donné.

On pourrait définir un deuxième continuum représentant la simultanéité de planification. On pourrait ici utiliser le terme de spontané et opposer la parole spontanée en tant que parole dont la planification est simultanée à celle d'oralisation et la parole non-spontanée en tant que parole non planifiée par le locuteur mais dont le support a été préparé par exemple par un expérimentateur comme dans le cas des corpus de parole lue. Notons que la différence entre la parole purement spontanée et non-spontanée (selon cette définition) est parfois délicate à faire, notamment dans le cas d'enregistrements de très bons acteurs qui récitent un texte pourtant non planifié en simultané mais qui est perçu comme étant naturel.

En suivant l'étude présentée par Eskénazi (1993), les différents types de corpus oraux à notre disposition peuvent être représentés grâce à deux continuums illustrés dans la Figure 17 ci-après. L'axe vertical illustre le continuum allant de la parole spontanée à la parole non spontanée, la différence étant fondée sur le délai de planification. L'axe horizontal définit lui le continuum allant de la parole naturelle à la parole contrainte de laboratoire en fonction de la situation de communication. Les deux axes définissent un espace représentant les différents types de corpus que l'on peut rencontrer ; notons que ces différents styles sont souvent englobés sous les termes de parole spontanée ou non-spontanée, obscurcissant ainsi certaines subtilités pourtant fondamentales dans le cadre d'études expérimentales.

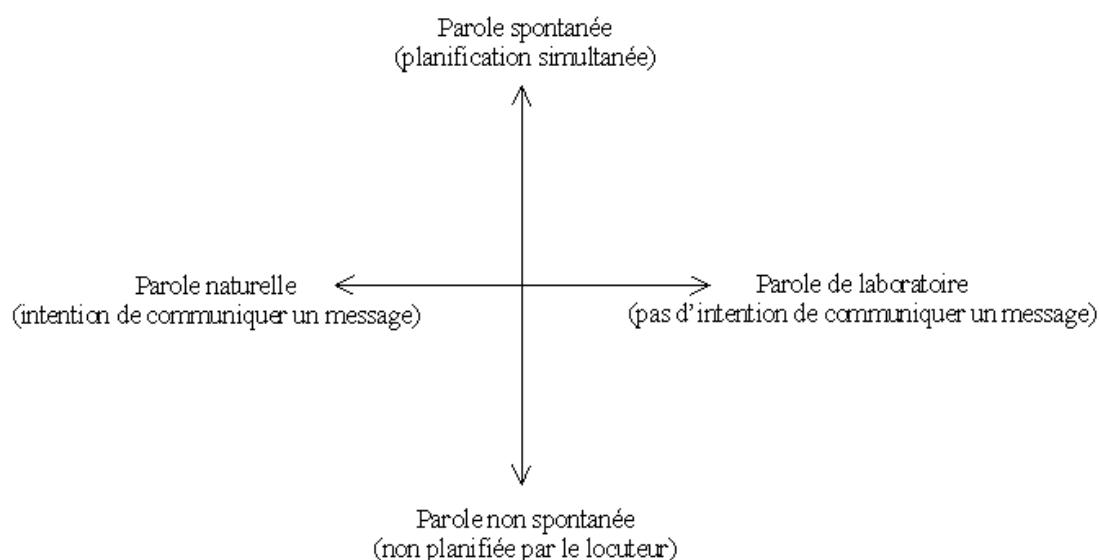


Figure 17. Représentation des différents corpus en fonction du type de parole

Notre étude est fondée sur l'observation d'un corpus oral d'enregistrements radiophoniques (cf. section 5.2 pour une description détaillée du corpus). Le corpus est considéré comme étant composé de parole authentique dans le sens où d'une part les enregistrements sont à but communicatif et d'autre part les enregistrements sont perçus comme étant naturels ; nous nous fions par conséquent au contexte de communication et à l'indice perceptif. De plus, la quantité de données orales (environ cinq heures et demie pour notre corpus) compense toutes les portions du corpus qui pourraient être perçues comme moins naturelles que d'autres. Nous ne nous situons pas au niveau extrême de spontanéité (sur la base du délai de planification) pour des raisons techniques d'exploitation ultérieure des données. En effet, notre corpus correspond à un signal sonore clair, aisément exploitable ; il n'en serait pas de même avec un corpus purement "spontané" tel que des discussions, principalement en raison des chevauchements de voix ou des signaux enregistrés dans de mauvaises conditions.

La définition des corpus en fonction des deux continuums définis par Eskénazi permet de classer notre corpus comme étant de la parole authentique avec un certain degré de spontanéité puisque la planification ne se fait pas en totale simultanéité.

5.1.3. Transcription et annotation de corpus oraux

Après avoir sélectionné un type spécifique de données orales, il convient de dire quelques mots sur les différents types de transcription des données. Delais-Roussarie distingue six catégories de transcription en fonction du niveau de représentation sur lequel elles portent (2002 : 7) :

- Transcription à partir des scripts d'enregistrement : les textes originaux en contexte de lecture.
- Transcription orthographique : encodage orthographique du contenu du signal.
- Transcription correspondant à une représentation phonémique de citation (*citation-phonemic representation*) : formes phonologiques sous-jacentes.
- Transcription phonétique large : transcription phonémique incluant les phénomènes propres à la production orale tels que la liaison, l'insertion du schwa, etc.
- Transcription phonétique étroite : transcription exacte du signal.

- Transcription acoustique-phonétique : analyse et segmentation du signal en terme de phases articulatoires (comme les phases de fermeture et d'explosion d'une plosive non-voisée).

Le choix de la transcription dépend donc directement de l'étude que l'on veut mener. Par exemple, dans un travail en syntaxe, on s'intéressera tout particulièrement au niveau orthographique ainsi qu'au niveau de regroupement des mots en unités syntaxiques plus larges. En revanche, les analyses dans le domaine de la phonétique nécessitent une transcription au niveau du segment, ce qui implique une transcription nettement plus fine (et souvent plus coûteuse en terme de temps de travail).

Pour obtenir cette transcription phonétique, on passe souvent par des étapes intermédiaires, notamment le codage orthographique souvent considéré comme le niveau fondamental de transcription. Il existe des points de discussion à tous les niveaux de transcription ; celui de la transcription orthographique pose le problème des conventions de notation telle que l'utilisation des symboles de ponctuation ou non (Delais-Roussarie 2002 : section 2.2) et le choix de transcription pure ou de transcription en fonction des réalisations effectives (par exemple *j'sais* en français au lieu de *je sais*).

Les mêmes problèmes seront soulevés lors de la phase de transcription phonétique (étroite ou large). De plus, se pose à ce niveau de transcription la question de l'alphabet adopté. On a longtemps utilisé l'Alphabet Phonétique International, mais celui-ci pose un problème de portabilité du fait qu'il requiert l'installation d'une police spécifique et qu'il n'est pas toujours reconnu lors d'utilisation de certains logiciels, notamment de traitement du signal. De ce fait, l'alphabet SAMPA (*Speech Assessment Methods Phonetic Alphabet*)¹⁸ est souvent préféré à l'IPA car il est constitué uniquement de symboles ASCII utilisables sur toutes les plates-formes et sans installation préalable de police (cf. liste des symboles SAMPA avec leur équivalent en IPA en Annexe 1). De plus, il est désormais aisé de convertir une transcription SAMPA en une transcription IPA. Pour ces raisons, l'alphabet SAMPA a par conséquent été préféré à l'alphabet IPA dans ce travail.

¹⁸ SAMPA : disponible à partir du site <http://www.phon.ucl.ac.uk/home/sampa/home.htm>

5.1.4. Disponibilité des corpus

Les corpus d'anglais sont nombreux et si on se réfère aux corpus donnés sur le site de l'université de Lancaster, pas moins de 20 corpus différents sont mis à la disposition des expérimentateurs (voir Auran, Bouzon et Hirst 2002 pour une liste détaillée de ces différents corpus) ; ces corpus portent principalement sur de l'anglais britannique ou américain. Toutefois, s'il est vrai que les corpus d'anglais sont nombreux, rares sont ceux qui sont disponibles et qui répondent de manière précise aux attentes spécifiques des chercheurs qui les exploitent. En effet, peu, sinon aucun, sont transcrits phonétiquement et possèdent un niveau d'analyse nécessaire à son exploitation. Point également important, peu d'entre eux sont disponibles gratuitement ou à faible coût.

Il existe toutefois le corpus d'anglais britannique IViE. Malgré un alignement au niveau orthographique et une annotation prosodique d'une partie du corpus (environ quatre heures sur les trente-six heures du corpus), IViE n'est pas entièrement annoté et n'offre pas la finesse d'analyse nécessaire pour des études expérimentales (transcription phonémique, voire alignement temporel). De plus, IViE est tout particulièrement intéressant dans le cadre d'études comparatives concernant les variétés régionales de l'anglais britannique ; or, il ne représente pas un corpus type quand on s'intéresse à l'anglais dit standard (plutôt qu'à l'anglais dans ses diversités régionales).

C'est dans cette perspective que nous proposons une méthodologie appliquée à la base de données MARSEC. Ce corpus possède les caractéristiques essentielles pour des études ultérieures mais nécessitent un ensemble de traitements supplémentaires pour être totalement exploitable. Cette méthodologie ainsi que l'ensemble des traitements relèvent du projet Aix-MARSEC¹⁹ dont le but est la constitution d'une base de données orales authentiques d'anglais britannique standard (en l'occurrence MARSEC) annotée et alignée au niveau phonétique dans le but d'être exploitée dans de nombreuses recherches sur l'anglais britannique (Auran, Bouzon et Hirst 2002 ; 2004a et b ; Auran et Bouzon 2003 ; Bouzon 2004 ; Bouzon, Auran et Hirst 2003 ; Auran, Bouzon, Hirst, Levy et Nocéra 2004).

La base de données Aix-MARSEC est développée en collaboration avec C. Auran et D. Hirst dans le cadre de EPGA²⁰ (*English Prosody Group of Aix*), groupe d'anglicistes

¹⁹ http://www.lpl.univ-aix.fr/~EPGA/fr_marsec.html

²⁰ <http://www.lpl.univ-aix.fr/~EPGA/index.html>

au Laboratoire Parole et Langage dont le but est de favoriser les études en prosodie en anglais britannique.

5.2. Description du corpus

5.2.1. Le corpus d'origine : du SEC au MARSEC

Le corpus utilisé dans ce projet est le corpus MARSEC (*MAchine Readable English Corpus*) correspondant à l'origine au corpus SEC (*Spoken English Corpus*). Le SEC est un corpus élaboré par des phonéticiens de l'université de Lancaster et de l'université de Leeds (G. Knowles et L. Taylor), ainsi que par des chercheurs travaillant sur la parole (principalement P. Alderson et B. Williams) appartenant au groupe IBM (UK Scientific Centre).

Les données orales du SEC correspondent à des enregistrements de la BBC datant des années 1980 (entre novembre 1984 et mars 1987). Les enregistrements sont de nature diverse et regroupent onze types différents d'enregistrements listés dans la Figure 18. Ces différents styles recouvrent un certain panel de catégories de discours et peuvent par exemple être exploités séparément lors d'études expérimentales. Il est donc intéressant de travailler à la fois sur la globalité du corpus en tant qu'échantillon de parole continue et sur des styles différents pour en tester leur pertinence dans la perspective de recherche en question. Le SEC rassemble 53 locuteurs différents, 17 locutrices et 36 locuteurs, ce qui représente une couverture intéressante des variations inter-individuelles. Aix-MARSEC correspond à de la parole authentique dans le sens où il est produit dans un but communicatif et qu'il n'est pas contraint par des tâches comme ce serait le cas avec de la parole de laboratoire.

<i>Groupe A.</i>	Commentaires	<i>Groupe G.</i>	Fiction
<i>Groupe B.</i>	Bulletin d'informations	<i>Groupe H.</i>	Poésie
<i>Groupe C.</i>	Parole publique de type I	<i>Groupe J.</i>	Dialogues
<i>Groupe D.</i>	Parole publique de type II	<i>Groupe K.</i>	Propagande
<i>Groupe E.</i>	Emissions religieuses	<i>Groupe M.</i>	Divers
<i>Groupe F.</i>	Reportages		

Figure 18. Liste des types d'enregistrement dans le SEC

Les enregistrements représentent une durée totale de 339 minutes et 18 secondes, le tout étant divisé en 408 fichiers d'environ 60 secondes dans le but de faciliter la manipulation informatique du corpus (cf. CD-ROM 1, dossier 'Audio'). Chaque fichier-son s'arrête à la pause la plus proche des soixante secondes moyennes de chaque enregistrement, tout en gardant une portion de chevauchement entre deux fichiers, ceci pour éviter que des syllabes ou phonèmes ne soient tronqués et donc inexploitable.

L'ensemble du corpus est transcrit orthographiquement (environ 55.000 mots) et annoté prosodiquement grâce à un système de marques tonétiques (*Tonetic Stress Marks* noté *TSM*). Les quatorze marques utilisées permettent de décrire le contour mélodique d'un énoncé ; comme l'explique Roach (1994), cet accent indique le mouvement de la voix : il débute sur la syllabe annotée et continue jusqu'à la syllabe accentuée suivante ou jusqu'à la frontière d'unité intonative. Les deux transpositeurs, G. Knowles et B. Williams, ont travaillé en parallèle tout en conservant 80 passages de chevauchement (soit 9% du corpus) dans le but de tester et de valider leur annotation. Les extensions GOK (pour G. Knowles) et BJW (pour B. Williams) sont ajoutées dans les noms de fichier pour pouvoir identifier le transpositeur relatif à chaque fichier.

Le SEC est ensuite devenu le corpus MARSEC qui, comme son nom l'indique, est rendu exploitable de manière informatique. Le passage de SEC à MARSEC correspond essentiellement à deux modifications. Tout d'abord, certaines marques tonétiques ont été remplacées dans le but d'obtenir une liste homogène constituée uniquement de symboles ASCII, utilisables plus facilement ; la liste de ces symboles est donnée dans la Figure 19 ci-après.

–	<i>low level</i>	,	<i>low rise</i>
~	<i>high level</i>	‘	<i>low fall</i>
<	<i>step-down</i>	,\	<i>(low rise-fall – not used)</i>
>	<i>step-up</i>	\,	<i>low fall-rise</i>
/’	<i>(high) rise-fall</i>	*	<i>stressed but unaccented</i>
‘/	<i>high fall-rise</i>		<i>minor intonation unit boundary</i>
/	<i>high rise</i>		<i>major intonation unit boundary</i>
\	<i>high fall</i>		

Figure 19. Liste des 14 symboles ASCII utilisés pour l'annotation prosodique du MARSEC (Roach 1994 : 94)

La deuxième modification apportée au corpus correspond à l'alignement manuel du signal en mots ; chaque mot du corpus est transcrit orthographiquement et chacun de ces intervalles représentés par un mot est temporellement aligné avec sa réalisation dans le signal sonore. Cet alignement permet de récupérer la durée de chacun des 55.000 mots du corpus.

5.2.2. Homogénéisation des données et pré-traitements

Par homogénéisation des données, nous entendons les traitements préliminaires effectués sur le corpus afin que celui-ci soit plus aisément manipulable. Le premier travail a été de passer en revue la totalité des fichiers du corpus dans le but de détecter tout problème dans les fichiers-son et dans les fichiers d'étiquettes correspondants. C'est au cours de cette vérification que nous nous sommes aperçus que certains fichiers étaient défectueux et ont donc été abandonnés. De plus, pour faciliter les traitements automatiques ultérieurs, nous avons supprimé les 9% de chevauchement du corpus, avec priorité alphabétique pour les transcriptions de B. Williams, à savoir les fichiers-étiquettes avec l'extension '.BJW'.

Nous avons ensuite vérifié et corrigé manuellement la totalité des 55.000 étiquettes de mot grâce au programme Praat utilisé tout au long de ce travail (Boersma et Weenink 1996). Nous avons apporté des corrections à l'étiquetage en cas de décalages entre les étiquettes et le contenu du signal supérieurs à 50ms ; les 82 fichiers avec des modifications d'étiquettes (environ 20% des fichiers) possèdent l'extension '.mod' afin de pouvoir les identifier aisément. Ce travail long était nécessaire pour travailler sur un alignement fiable au niveau des mots.

5.3. Transcription phonétique du corpus

La première grande partie de ce projet consiste à transcrire phonétiquement la totalité du corpus et ceci de manière automatique en utilisant des scripts en langage Perl (rédigés par C. Auran et D. Hirst et disponibles dans les ressources Aix-MARSEC sur le site du groupe EPGA²¹, ainsi que dans le CD-ROM 1, dossier 'scripts'). Nous ne détaillerons toutefois pas ici le fonctionnement précis de chacun des scripts.

²¹ http://www.lpl.univ-aix.fr/~EPGA/fr_ressources.html

Cette phonétisation se fait en deux phases : (1) la phonétisation dite brute du corpus puis (2) l'optimisation de cette phonétisation par application de règles d'élision de phonèmes.

5.3.1. Phonétisation brute

5.3.1.1. Méthodologie

La phonétisation brute du corpus s'effectue par extraction automatique à partir d'un lexique, autrement dit, chacun des mots du corpus transcrit orthographiquement est recherché dans un dictionnaire électronique de prononciation possédant en entrées les formes orthographiques et en sortie les formes phonétiques. Le dictionnaire utilisé pour cette conversion graphème-phonème est le *Advanced Learner's Dictionary* (publié chez *Oxford University Press*) ; il contient environ 71 000 mots, ainsi qu'un codage des catégories syntaxiques (cf. CD-ROM 1, dossier 'Dictionnaires', fichier 'ALD.txt'). Avant d'être utilisé, le dictionnaire a été au préalable converti en alphabet SAMPA.

Ce dictionnaire est à la base de la phonétisation du corpus, mais trois autres dictionnaires, ou listes complémentaires, sont également utilisés dans le cadre de la phonétisation brute. Lors de la phase de conversion graphèmes-phonèmes à partir du dictionnaire général, environ 700 mots présents dans le corpus étaient absents du dictionnaire, nous les avons par conséquent transcrits manuellement à partir du dictionnaire de prononciation de Wells (1990). Ces mots ainsi que leur transcription phonétique sont alors regroupés dans un dictionnaire supplémentaire de formes complémentaires (cf. CD-ROM 1, dossier 'Dictionnaires', fichier 'mots_compl.txt').

Dans la transcription orthographique du corpus, certaines formes "problématiques" sont présentes telles que la transcription des hésitations ou les mots tronqués lorsque le corpus a été divisé en fichiers d'environ une minute. Pour traiter ces cas, nous avons créé un troisième dictionnaire avec la transcription manuelle de chacune de ces formes problématiques (cf. CD-ROM 1, dossier 'Dictionnaires', fichier 'formes_pb.txt').

Le quatrième et dernier dictionnaire utilisé est celui regroupant l'ensemble des formes réduites spécifiques à l'anglais britannique (cf. CD-ROM 1, dossier 'Dictionnaires', fichier 'formes_réduites.txt'). En effet, la transcription des mots présents dans le dictionnaire général correspond aux formes de citation ; or, la parole, d'autant plus lorsqu'elle est authentique, est caractérisée par des phénomènes de

réduction. Les formes concernées sont listées d'après certaines études (O'Connor 1967 ; Faure 1975 ; Duchet 1994 ; Viel 1995 ; Lilly et Viel 1998) et transcrites avec leur forme réduite dans un dictionnaire spécifique. Le choix entre la forme de citation (dictionnaire général) et la forme réduite se fait en fonction de son statut mélodique : en présence d'une marque tonétique (TSM), le mot est transcrit avec sa forme pleine, en l'absence de TSM, le mot est transcrit avec sa forme réduite. Par exemple, la conjonction *and* non précédée d'une TSM est transcrite /ænd/ (dictionnaire de formes réduites) ; en revanche, 'and (précédée d'une TSM) est transcrit avec la forme de citation /ænd/ (dictionnaire général).

Notons que malgré des réalisations divergentes de certaines formes en fonction de leur position dans l'énoncé (Deschamps, Duchet, Fournier et O'Neil 2004 : 86), aucun traitement spécifique n'est ici réalisé pour tenir compte de la position de ces mots. En effet, les réalisations non réduites de certains mots, comme par exemple en position finale d'énoncé, sont généralement précédées d'une marque tonétique (et sont donc transcrits avec leur forme pleine).

En conséquence, la phonétisation brute du corpus s'effectue en fonction de quatre dictionnaires différents :

- le dictionnaire général
- le dictionnaire des mots complémentaires transcrits manuellement
- le dictionnaire des formes réduites
- le dictionnaire des formes problématiques

Nous avons volontairement créé des dictionnaires (ou listes) séparés dans le but de n'apporter aucun changement au dictionnaire général pour pouvoir éventuellement utiliser des versions mises à jour de ce dictionnaire sans perdre les modifications apportées manuellement. De plus, dans l'éventualité où cette phonétisation serait appliquée à d'autres corpus, on peut imaginer faire des dictionnaires aditifs spécifiques à chaque corpus utilisé.

5.3.1.2. Traitements spécifiques

Lors de la phonétisation brute, certains phénomènes ont requis un traitement spécifique, comme par exemple les chiffres non présents dans le dictionnaire général. Une fonction Perl (écrite par D. Hirst) nous permet tout d'abord de transcrire orthographiquement les

chiffres puis de rechercher ces formes dans le dictionnaire général. De manière similaire, une fonction Perl nous permet de traiter les acronymes de type *BBC* en séparant les majuscules qui, elles, sont présentes dans le dictionnaire. Les combinaisons de lettres et de chiffres (comme par exemple les codes postaux en Angleterre) sont traitées de manière identique en séparant les chiffres et les lettres pour obtenir des formes présentes dans le dictionnaire général.

Cette fonction nous permet également de décomposer les formes contractées en auxiliaire + contraction et les génitifs en substantif + forme du génitif, le tout étant présent dans le dictionnaire mais sous des formes séparées. Enfin, nous avons également traité automatiquement la réalisation du morphème ‘-s’ (/s/, /z/ ou /ɪz/) du pluriel et de la troisième personne du singulier au présent simple, ainsi que le morphème ‘-d’ (/d/, /t/ ou /ɪd/) du prétérit simple et du participe passé régulier en fonction du contexte phonémique.

Un traitement spécifique est également prévu pour la réalisation de l'article défini *the* ainsi que la présence du -r- de liaison.

– En ce qui concerne l'article défini, le dictionnaire général possède la forme de citation /ði:/ utilisée lorsque l'article est précédé d'une marque prosodique. Lorsqu'il n'est pas précédé d'une marque tonétique, le script choisit entre la forme /ðɪ/ devant un mot commençant par une voyelle et /ðə/ dans les autres cas. Nous sommes conscients que la forme correcte de l'article devant un mot commençant par une voyelle est la forme avec la voyelle tendue /i/ (Duchet 1994), mais nous nous sommes contentés d'une transcription large avec uniquement les phonèmes /ɪ/ et /i:/.

– Dans le cas du -r- de liaison, nous avons placé un ‘r*’ à la fin de chacun des mots possédant un -r- de liaison potentiel (dans l'ensemble des dictionnaires). Lors de la phase de phonétisation, si le mot suivant commence par une voyelle, le -r- de liaison apparaît dans la transcription ; si en revanche le mot suivant débute par une consonne, le -r- n'apparaît pas. Par exemple, pour la préposition *for*, les formes possibles sont /fɔ:r*/ (forme de citation dans le dictionnaire général, utilisée en présence d'une TSM) et /fər*/ (forme réduite utilisée en l'absence de toute TSM) ; pour chaque cas, le -r- de liaison apparaît uniquement lorsque le mot suivant commence par une voyelle.

Deux problèmes se posent lors de cette phase de phonétisation brute : les doublons et les dates. Par doublons, nous entendons les mots avec deux entrées dans le

dictionnaire général et donc deux réalisations possibles. On a par exemple *wind* qui peut être réalisé /wind/ dans le cas du substantif ou /waɪnd/ dans le cas du verbe. On rencontre également le cas des paires de mots identifiables uniquement en fonction du patron accentuel comme le mot *present* réalisé /'prezənt/ dans le cas du substantif mais /pri'zent/ dans le cas du verbe. Le choix entre les différentes formes est impossible à prédire puisque nous ne tenons pas compte des catégories syntaxiques ni de l'emplacement de l'accent lexical. Pour tous les cas de doublons, nous n'avons pour l'instant aucune solution automatique ; par conséquent, nous conservons par défaut le substantif, en supposant qu'il est le plus probable.

Le deuxième problème est posé par la réalisation des chiffres entre 1000 et 1999 qui diffère s'il s'agit d'une date ou d'un chiffre. Si on prend l'exemple de 1999, il peut être réalisé *nineteen ninety nine* dans le cas d'une date ou *one thousand nine hundred and ninety nine* dans le cas d'un chiffre. Ici aussi, il est impossible de prédire quelle forme sera produite ; nous avons choisi d'utiliser par défaut la date compte tenu de la nature du corpus.

Nous envisageons toutefois deux solutions pour traiter ces problèmes ; la première consiste à récupérer l'étiquetage syntaxique CLAWS appliqué au corpus SEC pour permettre le choix des formes correspondantes. De plus, une annotation syntaxique de la base de données est en projet au Laboratoire Parole et Langage. La deuxième solution envisageable est l'utilisation d'un système de reconnaissance automatique de la parole : celui-ci donnerait la forme la plus probable par rapport au contenu du signal sonore.

Au cours des traitements spécifiques, nous avons également fait une première prédiction de la durée des phonèmes du corpus à partir des durées de mot en s'inspirant de l'approche de Campbell (1992a). En effet, comme nous l'avons décrit dans le premier chapitre (section 1.3.2.1), son approche consiste à prédire tout d'abord la durée des syllabes puis la durée des phonèmes à l'intérieur de ces syllabes en fonction du principe d'élasticité ; celui-ci suppose que, lorsque la syllabe est allongée, l'allongement porte uniformément sur tous les phonèmes, et inversement lorsque la syllabe est raccourcie. On obtient alors un facteur d'allongement au niveau de la syllabe typiquement appelé *transformée z* en statistique ou *facteur z*.

Dans notre approche, nous appliquons ce même principe au niveau du mot en fonction d'une part de la durée des mots (obtenue grâce à l'alignement temporel en mots) et de la moyenne et de l'écart type de chacun des phonèmes qui le composent. On

obtient ainsi la transformée z au niveau du mot que l'on peut représenter grâce aux formules (1) et (2) présentées dans la Figure 20. La première formule correspond à la durée du mot, à savoir la somme de la durée moyenne de chacun des phonèmes et des écarts types de ces phonèmes multipliés par un facteur z . Pour calculer ce facteur z (formule 2), on soustrait à la durée du mot la somme de la durée moyenne de chacun des phonèmes qui le composent, le tout étant ensuite divisé par la somme des écarts types de ces mêmes phonèmes.

L'application de l'algorithme de Campbell permet de prédire la durée des phonèmes du corpus et d'obtenir la transformée z de la totalité des mots. Ces informations seront utilisées ultérieurement lors de l'optimisation de la phonétisation.

$$Duree_Mot = \sum_{i=1}^{Nb_phonemes} (m_{phoi} + z * sd_{phoi}) \quad (1)$$

$$z = \frac{Duree_Mot - \sum_{i=1}^{Nb_phonemes} (m_{phoi})}{\sum_{i=1}^{Nb_phonemes} (sd_{phoi})} \quad (2)$$

Figure 20. Formules permettant d'obtenir le facteur z (2) en fonction de la durée du mot (1) et de la moyenne et de l'écart type de chaque phonème

5.3.2. Optimisation de la phonétisation

5.3.2.1. Méthodologie

La première phase de phonétisation que nous avons qualifiée de "brute" représente une première estimation de la réalisation phonémique des différents locuteurs du corpus. Cependant, celle-ci ne rend pas compte de certains phénomènes de production caractéristiques de l'oral tels que les élisions de phonèmes, les réductions, les assimilations, les métathèses, *etc.* Or, pour qu'une transcription phonétique soit fiable, il faut tenir compte de ces phénomènes, notamment dans le cas de corpus de parole authentique.

C'est dans cette optique que nous proposons l'application de règles phonotactiques d'élision de phonèmes dans la base de données Aix-MARSEC. Ceci n'exclut pas la

nécessité de rendre également compte des autres phénomènes cités, il s'agit simplement d'une première étape dans l'amélioration du système de phonétisation automatique que nous venons de présenter. Il est nécessaire à ce stade de préciser que le but de cette approche est d'aboutir à une meilleure transcription phonétique, à savoir une meilleure correspondance entre la transcription et le contenu du signal, et ceci dans le but d'aboutir à un alignement automatique au niveau du phonème de meilleure qualité. Le même type d'approche a par exemple été appliqué à l'allemand par Kipp, Wesenick et Schiel (1996).

Pour cela, nous avons établi onze règles d'élision de phonèmes en anglais à partir de l'observation manuelle d'échantillons du corpus et de certains travaux présents dans la littérature rendant compte de ce phénomène (Jones 1990 ; Wells 1990 ; Gimson, réédité par Cruttenden 1997). Les règles découlent du consensus entre les élisions systématiquement observées dans les données et celles qui se trouvent dans ces ouvrages. Les onze règles sont ordonnées et appliquées automatiquement en fonction de trois paramètres, certains concernant uniquement certaines règles :

- PARAMETRE INTONATIF : certaines règles s'appliquent uniquement en l'absence de marque prosodique (TSM).
- PARAMETRES TEMPORELS : il en existe trois. Tout d'abord, nous avons fixé un seuil minimal de 5ms de durée phonémique au-dessous duquel on suppose que les phonèmes ne sont pas produits. Ainsi, tous les phonèmes présents dans la transcription dont la durée est estimée à moins de 5ms sont automatiquement supprimés. Deuxième type de contrainte temporelle, pour quatre phonèmes, nous avons établi à partir de Klatt (1979) un seuil de durée minimale au-delà duquel le phonème en question ne peut être éliminé. Ces seuils sont les suivants : 110ms pour /θ/ et 55ms pour /t/, /d/ et /ə/. Ces seuils nous permettent d'affiner les règles concernant l'élision de ces quatre phonèmes. La troisième et dernière contrainte temporelle concerne la transformée z (cf. Figure 20) : quel que soit le contexte phonémique, les règles s'appliquent uniquement si la transformée z est inférieure à un certain seuil (fixé ici à zéro). En effet, si celle-ci est inférieure à zéro, le mot est raccourci et on suppose qu'il y aura application des règles et donc élision de phonèmes ; en revanche, si elle est supérieure à zéro, le mot est allongé et on suppose qu'il n'y aura aucune élision. Notons que lorsque deux règles s'appliquent dans un même mot, la transformée z est recalculée avant que la deuxième règle ne soit appliquée.

- PARAMETRE PHONOTACTIQUE : certaines règles s'appliquent en fonction d'un contexte phonémique spécifique.

La prise en compte des deux premières contraintes permet de ne pas appliquer les règles de manière brute dès lors que le contexte phonémique s'y prête. Ainsi, l'élision des phonèmes se fait en tenant compte du contexte phonotactique mais également du contexte intonatif et temporel.

5.3.2.2. Règles d'élision de phonèmes

Les douze règles d'élision de phonèmes sont résumées dans le Tableau 2 ci-après (pour une description plus détaillée, voir Auran et Bouzon 2003).

Règles	Phonèmes	Contextes	Contraintes	Exemples
1			<5ms	
2	d	and	TSM	and <u>u</u> then
3	h	he('s/l/d) him his her	TSM	in <u>h</u> er case
4	t d	{[t][d]} # {[t][d]}	seuil - sauf '-ed'	I've got <u>t</u> to
5	t d	C ₁ + {[t][d]} # C ₂ - {[h][j]}	seuil	mustn't <u>t</u> lose
6	p k	nasale homorganique + {[p][k]} (#) C - {[r][l][j]}		gl <u>i</u> mpse
7	l	[ɔ:] + [l] (#) C		al <u>l</u> ways
8	θ	C + [θ] (#) [s]	seuil	twelf <u>th</u> s
9	ptk bdg	[s/z] + {[p b][t d][k g]} (#) [s/z]		tourist <u>s</u>
10	ə	[ə] + {[l][r]} (#) + voyelle réduite {[ɪ][ə]}	seuil - */r/	cam <u>e</u> ra
11	ə	# [kən] ('syll (syll _[0...n])) #	TSM - seuil	co <u>n</u> front
12	ə	{[k][p]} + [ə] + [n] #	seuil	ope <u>n</u>

Tableau 2. Présentation des onze principes d'élision de phonèmes

Dans la première colonne de ce tableau se trouve le numéro attribué à chacune des règles, la deuxième colonne précise le phonème concerné, la troisième présente le contexte phonotactique (lorsqu'il en est question), la quatrième détaille les diverses

contraintes (intonatives, temporelles et/ ou phonotactiques) propres à chaque règle et la cinquième illustre chaque règle avec un exemple (le phonème potentiellement élide est alors souligné). En ce qui concerne le contexte (troisième colonne), 'C' et 'V' représentent respectivement n'importe quelle consonne ou voyelle, '#' symbolise une frontière de mot (facultative si placée entre parenthèses) et '-' est un signe d'exclusion.

On peut répartir ces règles en deux catégories selon qu'elles s'appuient ou non sur des contraintes phonotactiques :

- Règles non-phonotactiques (de deux types) : élision des phonèmes dont la durée prédite est inférieure ou égale à 5ms (règle 1) ; élision fondée sur la nature morphologique des unités (règles 2 et 3).
- Règles phonotactiques : elles précisent l'inventaire des contextes spécifiques "autorisant" l'élision d'un phonème donné.

Chacune des règles s'applique à la condition stricte que la durée prédite du phonème en question soit supérieure à 5ms et que le facteur z soit inférieur à zéro.

Outre les trois types de contraintes, certaines règles sont caractérisées par des restrictions supplémentaires (cf. tableau). Par exemple, dans la règle n°4, l'élision du /d/ s'applique sauf lorsqu'il s'agit du morphème '-ed' réalisé /ɪd/ comme par exemple dans *unexpected turn*. La règle n°10 possède une restriction supplémentaire : elle ne s'applique pas lorsqu'il s'agit de la séquence /rəl/ suivie d'une voyelle réduite (comme dans *necessarily*) et ceci pour deux raisons ; l'attaque /r/ n'est pas licite en anglais et le phonème /r/ ne peut pas être en position finale de syllabe en anglais britannique standard. L'élision est donc impossible.

Lorsque l'élision d'un phonème selon les règles présentées a pour conséquence la suite de deux phonèmes identiques adjacents, un traitement spécifique fait en sorte que les deux phonèmes soient regroupés en un phonème unique. Tel est le cas pour les règles n°9 et n°10 : dans l'exemple *tourists*, s'il y a élision de /t/, on aura la forme /tʊəɾɪss/ ensuite réduite à /tʊəɾɪs/. Cette réduction ainsi que la décision d'élider les consonnes /t/ et /d/ dans les contextes spécifiés dans les règles 4 et 5 ont pour but d'améliorer l'alignement. En effet, dans les cas concernés, la durée de la tenue du /t/ ou /d/ restant est double mais on ne retrouve pas pour autant les deux phonèmes distinctement dans le signal, autrement dit, on ne retrouve pas deux gestes articulatoires complets pour chaque phonème. Or, si deux phonèmes sont présents dans la

transcription, l'aligneur automatique va les rechercher sans succès et va par conséquent engendrer des erreurs.

5.3.2.3. Evaluation de l'efficacité des règles d'élision

L'efficacité de l'application de ces règles d'élision a ensuite été évaluée en comparant la phonétisation optimisée (après application automatique des règles) et une transcription manuelle d'un ensemble de fichiers. Pour cette évaluation, nous avons utilisé les mesures de rappel, précision, silence et bruit ainsi que la notion de F-mesure (Tableau 3 ci-dessous).

MESURES	
RAPPEL	50,51 %
PRECISION	74,44 %
SILENCE	49,49 %
BRUIT	25,56 %
F-MESURE	60,18 %

Tableau 3. Résultats de l'évaluation des élisions

L'évaluation présentée dans le Tableau 3 révèle que l'application des onze règles permet de prédire environ 50% des élisions observées dans le corpus (mesure de rappel). Elle montre également que, parmi les contextes concernés par les règles, notre prédiction est correcte dans plus de 74% des cas (précision). La notion de F-mesure est tout particulièrement intéressante puisqu'elle représente la moyenne harmonique entre les mesures de rappel et de précision ; pour qu'un modèle soit performant, il faut que cette mesure soit la plus importante possible. Ceci signifie par conséquent que la qualité de la prédiction des cas d'élision est évaluée à environ 60% (F-mesure).

60% est un résultat satisfaisant mais nous avons en projet d'améliorer la performance de nos règles afin qu'elles rendent compte d'une plus grande proportion des élisions et ceci de trois façons :

- l'affinement des contraintes, principalement temporelles, permettrait de rendre compte de davantage de cas d'élisions ;

- la création de nouvelles règles d'élision : nous avons notamment remarqué l'élision quasi-systématique de la consonne /ð/ dans l'article défini *the*, ceci représentant la moitié des élisions non prédites par nos onze règles ;
- l'utilisation du système d'alignement automatique : nous reviendrons sur ce point ultérieurement.

5.4. Alignement automatique des phonèmes

5.4.1. Méthodologie

Nous avons ensuite procédé à l'alignement automatique à partir de la transcription phonétique ainsi obtenue. Nous entendons par alignement une mise en correspondance temporelle entre la transcription et le contenu du signal sonore. C'est sur cet alignement phonémique que va se fonder l'alignement des unités de niveau supérieur. L'alignement automatique est effectué par Christophe Lévy et Pascal Nocéra au Laboratoire Informatique d'Avignon (LIA²²) en utilisant un système de Force Viterbi (Nocéra, Linares, Massoné et Lefort 2000).

Ce système utilise des Modèles de Markov Cachés (HMMs) entraînés sur le corpus TIMIT. L'alignement se fait par application de l'algorithme Viterbi contraint par la phonétisation, autrement dit, on présente au système une chaîne de phonèmes (résultat de la phonétisation) ainsi que le signal correspondant et celui-ci recherche dans le signal la totalité de ces phonèmes et place des bornes temporelles au début et à la fin de chacun d'entre eux. On parle de Force Viterbi puisqu'on force le système de reconnaissance à détecter la chaîne de phonèmes donnée en entrée, d'où la nécessité d'avoir la meilleure correspondance signal-transcription possible. On obtient alors un alignement temporel entre la chaîne de phonèmes de la transcription et le signal.

Lorsque l'aligneur ne détecte pas un des phonèmes en entrée, il fixe automatiquement sa durée à 10ms. Notons que si ce seuil s'avérait systématique et fiable, on pourrait imaginer une suppression des phonèmes dont la durée est fixée à 10ms par l'aligneur et affiner par conséquent la transcription. Il serait donc intéressant d'évaluer ces seuils afin de peut-être pouvoir les exploiter dans le cadre de la prédiction des élisions de phonèmes.

²² <http://www.lia.univ-avignon.fr/php/accueil.php>

5.4.2. *Evaluation et perspectives d'amélioration de l'alignement automatique*

Nous avons ensuite procédé à l'évaluation de l'alignement automatique en comparant les résultats de l'alignement automatique avec les mêmes fichiers alignés manuellement (environ quatre fichiers d'une minute chacun). L'évaluation consiste à estimer la quantité de décalages entre les alignements automatique et manuel, ainsi que l'étendue des décalages. La mesure des décalages a été effectuée de manière automatique à l'aide de scripts en langage Perl, elle est résumée dans le Tableau 4.

Les seuils allant de 5 à 64ms utilisés pour l'évaluation sont ceux cités par Di Cristo (P.) et Hirst (1997). Ces résultats signifient que, par exemple, environ 68% des étiquettes ont un décalage inférieur à 20ms. Ce score est satisfaisant mais pourrait être amélioré par exemple en utilisant un modèle plus adapté au type de corpus que nous utilisons, notamment un modèle phonétique de l'anglais britannique puisque le système a été entraîné sur de l'anglais américain (corpus TIMIT). Cette amélioration est actuellement en projet car l'alignement d'environ une heure et demi du corpus a été corrigé manuellement par des étudiants en DEA²³ du Laboratoire Parole et Langage et ces informations vont nous permettre, en collaboration avec C. Lévy (LIA), de développer un modèle pour l'anglais britannique. Ainsi, le nouvel aligneur automatique devrait permettre des résultats bien meilleurs.

Nous avons également en perspective d'appliquer un algorithme de programmation dynamique DTW (*Dynamic Time Warping*). Cet algorithme permet la comparaison d'un signal et de sa synthèse jusqu'à correspondance temporelle entre les deux signaux (Di Cristo (P.) et Hirst 1997). De plus, il est évident que plus la transcription est proche du signal sonore, meilleur sera l'alignement ; il convient donc d'améliorer en tout premier lieu la phase de phonétisation grâce à l'affinement des règles d'élision ainsi que la prise en compte d'autres phénomènes de langue tels que les assimilations.

²³ Nous remercions à ce sujet S. Ali, C. De Looze, K. Edensor, A. Foucart, I. Løvstrøm, L.M. Macarlotti, M. Meghriche, S. Mortreux et A. Tortel.

Seuils	% de décalages inférieurs au seuil
64 ms	93.25 %
32 ms	82.02 %
20 ms	68.37 %
16 ms	59.97 %
15 ms	57.40 %
10 ms	42.43 %
5 ms	23.72 %

Tableau 4. Résultats de l'évaluation de l'alignement automatique

5.5. Autres niveaux d'analyse et d'alignement

Après ces phases de transcription et d'alignement temporel au niveau du phonème, les phonèmes sont automatiquement regroupés en unités supérieures, à savoir en syllabes, en constituants syllabiques (attaque, noyau et coda), en pieds accentuels et ictus silencieux (modèle d'Abercrombie), en unités rythmiques étroites et anacrouses (modèle de Jassem), en mots et en unités intonatives mineures et majeures. Dans le but de rendre Aix-MARSEC exploitable dans un cadre plus large de recherche, nous avons également appliqué l'algorithme MOMEL et le codage des points cibles en sortie avec l'alphabet INTSINT.

Ces différents regroupements se situent à des niveaux distincts et sont représentés comme tels dans le logiciel Praat puisque chaque regroupement se situe à un niveau (ou *tier*) différent dans le textgrid (cf. Figure 23 pour une illustration Praat de ces différents niveaux).

5.5.1. Regroupement en syllabes et en constituants syllabiques

Une fois les phonèmes alignés avec le signal, ceux-ci sont regroupés en syllabes à l'intérieur des mots en suivant trois phases différentes. Un premier regroupement des phonèmes en syllabes se fait à l'intérieur du mot en fonction du principe d'attaque maximale (Pulgram 1970) : chaque voyelle est considérée comme noyau vocalique

Pour effectuer un tel regroupement, nous avons dû dans un deuxième temps procéder à un traitement spécifique des consonnes syllabiques /m - n - l/. En effet, ces trois consonnes deviennent syllabiques lorsqu'elles apparaissent dans le contexte suivant :

C + {[m] [n] [l]} (+ {[z] [d]}) frontière syllabique

En conséquence, après un regroupement des phonèmes en fonction du principe d'attaque maximale, les consonnes syllabiques sont automatiquement détectées et ensuite considérées comme des noyaux syllabiques.

Une troisième phase de resyllabification s'opère alors en suivant le même principe d'attaque maximale mais en tenant compte des consonnes syllabiques. Pour résumer, dans un mot comme *information*, la syllabification lors de la première phase est /ɪn.fə.meɪfɪn/ où les points représentent les frontières syllabiques. Dans ce mot est présent le contexte C + /n/ + #, le /n/ est donc considéré comme syllabique. On a alors une troisième phase de resyllabification qui donne /ɪn.fə.meɪ.ɪfɪn/, /n/ est désormais considéré comme noyau syllabique, /f/ devient alors l'attaque de la quatrième syllabe.

Le découpage en constituants syllabiques (attaque, noyau et coda) de la chaîne segmentale découle du regroupement des phonèmes en syllabes. Lors de cette phase, nous attribuons les étiquettes d'attaque, de noyau et de coda aux phonèmes en fonction du principe d'attaque maximale ; ces étiquettes sont ensuite récupérées dans le niveau "constituants syllabiques". Les regroupements en syllabes et en constituants syllabiques (attaque, noyau et coda) constituent par conséquent deux niveaux distincts.

5.5.2. Regroupement des syllabes en unités rythmiques

De la même manière que les phonèmes sont regroupés en syllabes, les syllabes sont regroupées en unités rythmiques selon d'une part la définition d'Abercrombie et d'autre part celle de Jassem (cf. Chapitre 3).

Avant de procéder à un tel regroupement, nous avons dû prédire l'emplacement de l'accent lexical puisque, jusqu'ici, nous ne possédons pas une telle information dans nos données. Nous possédons l'annotation prosodique de Aix-MARSEC mais ces marques tonétiques précèdent uniquement les syllabes proéminentes et lors d'une observation d'échantillons du corpus, nous nous sommes aperçue qu'un regroupement en unités

rythmiques selon cette annotation génèrerait des unités bien trop longues, ne reflétant donc pas la structuration rythmique authentique.

Pour prédire automatiquement l'emplacement des accents lexicaux dans le corpus, nous avons supposé qu'un mot dissyllabique possède nécessairement un accent lexical. Le patron lexical de chacun des mots est alors récupéré dans le dictionnaire général, incluant à la fois les accents primaires et les accents secondaires. En revanche, savoir si les mots monosyllabiques sont accentués pose une tout autre question puisque l'accent lexical n'est pas noté pour les mots monosyllabiques. Pour ces cas, nous avons utilisé une liste de mots (fonctionnels) monosyllabiques toujours inaccentués à partir de Faure (1975) et de Jassem (1994) et ceci en fonction de leur position initiale, médiale ou finale dans l'unité intonative (cf. CD-ROM 1, dossier 'Dictionnaires', fichier 'mots_inacc.txt'). Cette liste comporte principalement des prépositions, des pronoms et des conjonctions, ainsi qu'une triple distinction selon que le mot en question est en position initiale, médiale ou finale d'unité. Tous les mots monosyllabiques absents de ces trois listes sont automatiquement précédés d'un accent lexical.

Les énoncés ci-dessous (fichier A0101B) illustrent la prédiction de l'accent lexical résultant de l'approche adoptée ; toutes les syllabes accentuées sont précédées du symbole ' pour l'accent primaire et , pour l'accent secondaire.

*'Good 'morning 'More 'news a 'bout the 'reverend 'Sun 'Myung 'Moon 'founder of the
,unifi'cation 'church 'who's 'currently in 'jail for 'tax e'vasion*

Le résultat obtenu est jugé bon et fiable.

En ce qui concerne les différents degrés d'accentuation, nous suivons ici le dictionnaire général, à savoir :

- aucun accent postérieur à l'accent primaire noté '
- les différents accents précédant l'accent primaire (aucune distinction n'est faite entre les accents secondaires et tertiaires) sont notés ,. Si on prend l'exemple du mot *incomprehensibility*, la transcription en IPA est donc /ɪn,kɒmpri,hensə'bɪlɪtɪ/.

Les syllabes sont ensuite regroupées en unités rythmiques en fonction de ces patrons accentuels mais également en fonction des marques prosodiques. En effet, l'annotation prosodique manuelle reflète les phénomènes de déplacements d'accent dus à des conflits accentuels que nous avons décrits dans le troisième chapitre. En cas de non

correspondance entre l'accent lexical et la marque tonétique, nous tenons compte de cette dernière pour le regroupement rythmique. Notons qu'il serait intéressant de comparer l'emplacement des accents lexicaux avec l'annotation prosodique afin de répertorier les contextes systématiques (ou non) de déplacements accentuels dans Aix-MARSEC.

Ainsi, les syllabes sont tout d'abord regroupées en pieds accentuels (codés F) selon la définition d'Abercrombie ; les syllabes inaccentuées en début d'unités intonatives sont codées SI pour *silent ictus*. Sur un niveau distinct, les syllabes sont regroupées selon le modèle de Jassem, à savoir en unités rythmiques étroites (codées NRU) et en anacrouses (codées ANA). Les débuts de pieds et de NRU sont déterminés par la présence d'un accent primaire ou secondaire.

5.5.3. *Mots et unités intonatives*

Ces deux unités représentent deux niveaux supplémentaires d'alignement déjà présents dans MARSEC. La transcription orthographique du niveau du mot possède également l'annotation prosodique (TSM). Le regroupement en unités intonatives mineures et majeures se fait sur le même niveau, la distinction se faisant au travers de la notation puisque le symbole | représente une frontière mineure et || une frontière majeure.

En ce qui concerne le niveau du mot, Aix-MARSEC possède toujours le niveau du mot avec annotation prosodique, mais nous avons ajouté un niveau supplémentaire dans lequel nous avons conservé uniquement la transcription orthographique, sans l'annotation prosodique. Ainsi, il est possible de travailler sur la transcription orthographique sans les TSM.

Par conséquent, les niveaux du mot et de l'unité intonative (mineure et majeure) représentent trois niveaux d'alignement supplémentaires dans Aix-MARSEC (niveaux 6, 7 et 8 dans la Figure 23).

5.5.4. *MOMEL & INTSINT*

Nous avons également appliqué l'algorithme MOMEL (MODélisation de la MELodie) développé au Laboratoire Parole et Langage (Hirst et Espesser 1993 ; Hirst, Di Cristo et Espesser 2000) dont le but est de modéliser la courbe de fréquence fondamentale à

l'aide d'un ensemble de points-cibles. Ces points-cibles sont ensuite codés grâce à l'alphabet INTSINT (International Transcription System for INTonation).

Ces deux étapes nous permettent d'avoir un deuxième type de codage de l'intonation et nous pouvons envisager une comparaison possible de ce codage avec le système de TSM appliqué au corpus. Le codage INTSINT des données représente deux niveaux supplémentaires d'annotation de la base de données : un premier niveau constitué du codage INTSINT et un deuxième niveau avec les valeurs en Hertz correspondantes (niveaux 9 et 10 dans la Figure 23).

Ce codage de l'intonation fait partie des différents traitements appliqués au corpus mais il ne sera pas exploité dans ce travail.

5.5.5. Aix-MARSEC : une base de données pluri-linéaire

Aix-MARSEC est par conséquent un corpus produit par 53 locuteurs différents, regroupant onze styles d'enregistrement, d'une durée d'environ cinq heures et demie de parole authentique, à savoir 195.687 phonèmes, 88.794 syllabes et 54.083 mots. Aix-MARSEC, c'est également une méthodologie englobant un système de phonétisation, d'alignement temporel au niveau phonémique et de regroupement automatique à divers niveaux de représentation.

Le corpus est aligné aux niveaux du phonème, du constituant syllabique (attaque, noyau et coda), de la syllabe, du pied et de l'ictus silencieux, de l'unité rythmique étroite et de l'anacrouse, du mot (avec TSM), du mot seul et de l'unité intonative mineure et majeure. Il est également codé grâce à l'alphabet INTSINT dont les valeurs en Hertz apparaissent sur un niveau distinct. Tous les niveaux d'alignement représentent des niveaux séparés et sont disponibles en format d'étiquette textgrid propre à Praat. Chaque fichier possède par conséquent un fichier-étiquette avec l'alignement temporel de chacun des niveaux cités.

Aix-MARSEC se présente donc comme une base de données pluri-linéaire, comme nous pouvons le voir dans l'illustration d'un extrait du corpus (Figure 23). Sur cette figure extraite directement du programme Praat, on peut voir les différents niveaux de représentation, à savoir le niveau (1) des phonèmes, (2) des constituants syllabiques, (3) des syllabes, (4) des pieds accentuels et ictus silencieux, (5) des unités rythmiques et des anacrouses, (6) des mots, (7) des mots avec l'annotation prosodique, (8) des unités

intonatives mineures et majeures, (9) du codage INTSINT et (10) des valeurs en Hertz correspondantes.

Aix-MARSEC constitue par conséquent une base de données exploitable dans des recherches expérimentales de divers horizons. Elle a servi dans divers travaux, notamment dans des travaux portant sur l'influence d'un ensemble de facteurs prosodiques sur la durée des mots (Bouzon 2002 ; Bouzon et Hirst 2002), ainsi que pour une étude prosodico-pragmatique de l'anaphore (Auran & Hirst 2004). Aix-MARSEC est désormais disponible à prix coûtant sur le site de EPGA²⁴.

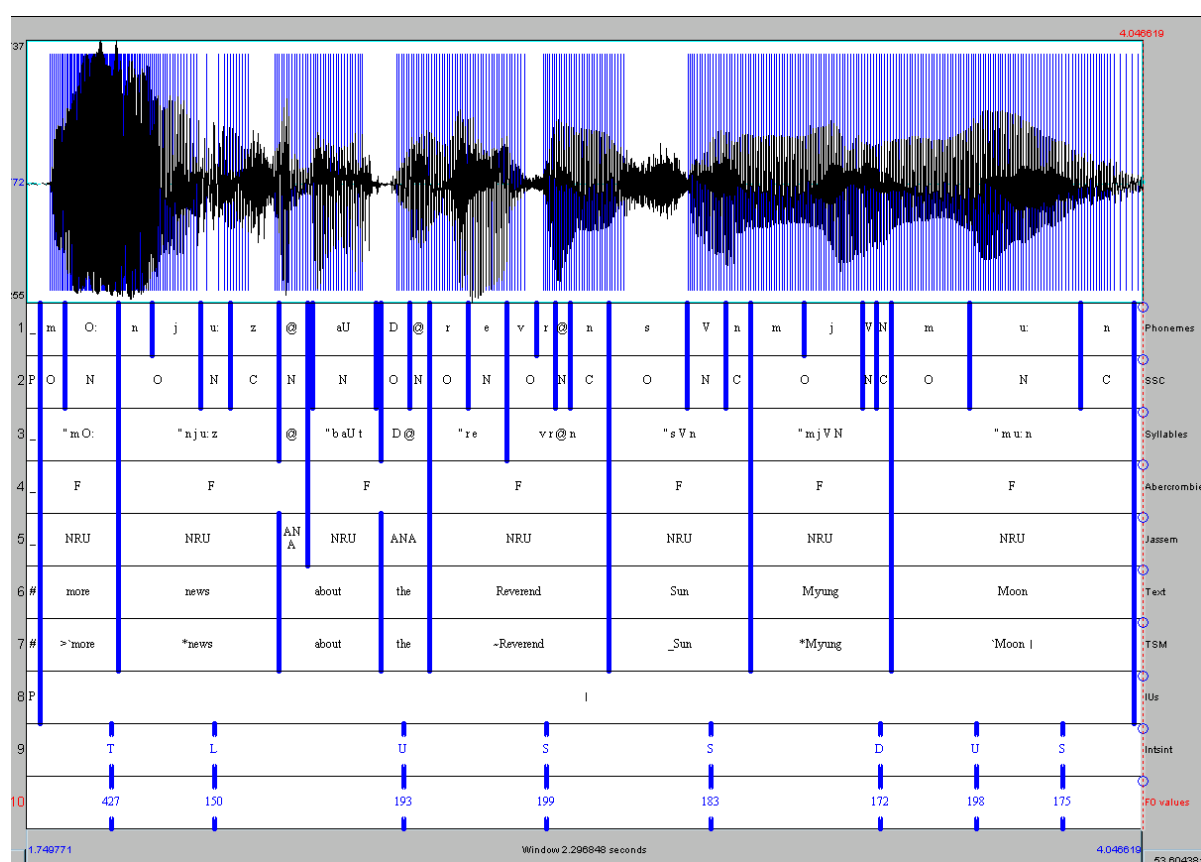


Figure 23. Illustration d'un extrait du corpus Aix-MARSEC (dans Praat) avec les neuf niveaux d'alignement

²⁴ Aix-MARSEC est disponible à l'adresse suivante : http://www.lpl.univ-aix.fr/~EPGA/fr_marsec.html

5.6. Mise en forme des données

5.6.1. Description des divers paramètres

La base de données que nous venons de présenter est celle que nous utiliserons dans ce travail, et ceci pour diverses raisons. Tout d'abord, c'est une base de données orale variée regroupant un ensemble de locuteurs différents, dans des contextes différents de communication. Le style de parole est neutre dans le sens où il est naturel et rarement emphatique. De plus, Aix-MARSEC offre un alignement temporel à divers niveaux de représentation nécessaires dans cette étude. En effet, un tel alignement nous permet de tester les hypothèses posées dans les chapitres précédents puisque nous pouvons observer la durée des phonèmes, des constituants syllabiques, des syllabes, des pieds accentuels, des unités rythmiques étroites (NRU), des anacrouses, des mots et des unités intonatives, et ceci dans divers contextes.

Pour pouvoir exploiter ces données, il nous faut tout d'abord regrouper toutes les informations qu'elles contiennent sous la forme d'un tableau ou plus exactement sous la forme de données tabulaires exploitables dans un logiciel de statistique. La mise en place de ce tableau est effectuée automatiquement grâce à un script Perl écrit par C. Auran (CD-ROM 1, dossier 'scripts', fichier 'mise-en-forme.pl'). Le tableau de données est organisé autour de l'unité minimale qu'est le phonème dans le sens où il possède une entrée (ligne) par phonème et un ensemble d'informations est fourni pour chacune de ces entrées. Tous les codages se présentent sous forme d'étiquettes abrégatives en anglais ; en effet, tous les traitements et annotations des données sont en anglais puisque notre but est de distribuer gratuitement le corpus à la communauté linguistique. La totalité des données est présente sur le CD-ROM 1 (fichier 'données.txt').

La première catégorie d'information concerne le fichier ('*file*'), puis le phonème dont il est question ('*phon*'), ainsi que sa durée brute exprimée en ms ('*dp*') et normalisée ('*zp*') selon l'approche de Campbell (cf. section 5.3.1.1 pour les détails concernant le calcul de la transformée *z*) et le mot auquel le phonème appartient ('*word*'). La deuxième catégorie d'information concerne les unités de rang supérieur auxquelles le phonème appartient. On a ainsi la durée brute de la syllabe ('*ds*'), du pied accentuel ('*df*'), du mot ('*dw*'), de l'unité rythmique ('*dr*') et de l'unité intonative ('*di*') que l'on obtient grâce à l'alignement temporel entre les différents niveaux de représentation et le signal.

Dans le but de tester les différentes hypothèses relatives à l'isochronie, nous avons besoin des informations relatives au nombre de constituants inférieurs compris dans une unité donnée, comme par exemple le nombre de phonèmes dans une syllabe ou de syllabes dans un pied. Nous utiliserons le terme général de 'complexité' pour faire référence à cette catégorie d'information. Nous possédons par conséquent le nombre de phonèmes dans la syllabe ('*nps*'), le nombre de phonèmes ('*npf*') et de syllabes ('*nsf*') dans le pied accentuel, le nombre de phonèmes ('*npw*') et de syllabes ('*nsw*') dans le mot, le nombre de phonèmes ('*npi*'), de syllabes ('*nsi*'), de pieds ('*nfi*') et de mots ('*nwi*') dans l'unité intonative et le nombre de phonèmes ('*npr*') et de syllabes ('*nsr*') dans l'unité rythmique (NRU ou anacrouse). Pour la complexité de l'anacrouse et de l'ictus silencieux, nous utilisons respectivement la complexité de la NRU et du pied accentuel puisque les données ont été codées ainsi. Toutes ces informations sont par conséquent réunies dans le but de tester les effets de la complexité sur la durée des différentes unités.

Une troisième catégorie regroupe des informations relatives à la nature du phonème ('*pho*'), à la structure syllabique ('*ss*'), au statut accentuel du phonème ('*str*'), à la présence et la nature de la marque tonétique ('*tsm*') et à la nature de l'unité rythmique ('*ru*'). Dans la colonne '*pho*', les différents phonèmes sont catégorisés selon qu'il s'agit d'une consonne ('*C*'), d'une voyelle longue ('*V:*'), d'une diphtongue ('*VD*') ou de toute autre voyelle ('*V*'). La colonne relative à la structure syllabique permet de savoir à quel constituant syllabique le phonème appartient, à savoir s'il s'agit de l'attaque ('*O*'), du noyau ('*N*') ou de la coda ('*C*'). Dans la colonne '*str*', chaque phonème est codé comme accentué ('*S*') ou inaccentué ('*U*'). Dans la colonne '*tsm*', en l'absence de TSM, le phonème est codé '0' ; dans le cas contraire, on trouve le symbole ASCII correspondant à l'annotation du mouvement mélodique dont il est question. Enfin, la colonne '*ru*' nous permet de distinguer les NRU ('*R*'), les anacrouses ('*A*') et les ictus silencieux ('*SI*'), ces derniers correspondant aux anacrouses uniquement en position initiale d'unité intonative.

Une quatrième catégorie d'informations concerne le découpage syllabique en fonction de la notion d'ambisyllabité. En effet, dans le Chapitre 2, nous avons soulevé la problématique de l'ambisyllabité en anglais et afin de pouvoir tester les différentes hypothèses relatives à cette problématique, il est nécessaire de retrouver une telle information dans nos données. Ainsi, grâce aux colonnes '*dsa*' et '*npsa*', nous

possédons respectivement la durée brute des syllabes lorsque le regroupement en syllabes tient compte des consonnes ambisyllabiques et le nombre de phonèmes dans cette syllabe. De la même manière, la colonne '*ssa*' détaille la structure syllabique en fonction du découpage fondé sur l'ambisyllabité avec comme catégorie supplémentaire '*A*' pour coder les consonnes ambisyllabiques.

Une dernière catégorie d'informations concerne l'indice de position de chaque phonème dans les différents constituants auxquels il appartient. Cet indice est calculé automatiquement grâce à un script Perl écrit par Daniel Hirst (cf. CD-ROM 1, dossier '*scripts*', fichier '*addindices.pl*') pour la position du phonème dans la syllabe ('*ips*'), dans l'unité rythmique ('*ipr*'), dans le mot ('*ipw*') et dans l'unité intonative ('*ipi*'). Rappelons que les données tabulaires sont organisées en fonction des phonèmes ; or, cette organisation présente un biais lorsque nous effectuons des tests sur les durées de constituants plus grands que le phonème. Si on s'intéresse par exemple à '*ds*' (durée de la syllabe), on prend en compte la durée d'une syllabe donnée autant de fois qu'elle contient de phonèmes. Le même problème se pose avec la durée de l'unité rythmique, du pied, du mot et de l'unité intonative. C'est dans le but de remédier à ce problème que nous avons codé tous les phonèmes en fonction de leur position dans ces constituants. Au moment des tests statistiques, nous créons des sous-fichiers pour les syllabes, les unités rythmiques, les mots et les unités intonatives (en ne prenant que le premier phonème de chaque unité). Notons que l'indice de position du phonème dans le pied n'est pas nécessaire puisque le pied et l'unité rythmique commencent par le même phonème.

Ainsi, lorsqu'on prend en compte la durée des syllabes, on donne comme condition que '*ips*' soit égal à 1, on ne prend donc en compte qu'une seule fois la durée de chaque syllabe. Pour faciliter les tests statistiques dans R, nous avons créé des sous-fichiers pour chaque constituant. Nous avons créé un sous-fichier pour la syllabe avec la condition que '*ips*' soit égal à 1. De la même manière, en cumulant les conditions, nous avons créé des sous-fichiers pour la NRU, l'anacrouse (Jassem), l'anacrouse initiale (anacrouses en position initiale d'unité intonative, correspondant à l'ictus silencieux pour Abercrombie), le mot et l'unité intonative.

Les pauses entre les unités intonatives sont conservées dans les données tabulaires, elles sont codées '*P*' dans toutes les colonnes, sauf pour '*dp*' et '*zp*' dans lesquelles on trouve respectivement la durée brute et la durée normalisée de ces pauses. L'ensemble

de toutes ces informations est récapitulé dans le Tableau 5 ci-après avec dans la première colonne un numéro indicatif pour chacun des paramètres, dans la deuxième colonne l'étiquette de chaque information et dans la troisième colonne un bref rappel de la signification de ces paramètres.

1	<i>file</i>	Nom du fichier auquel le phonème appartient
2	<i>word</i>	Mot auquel le phonème appartient
3	<i>phon</i>	Phonème (SAMPA)
4	<i>dp</i>	Durée brute du phonème
5	<i>zp</i>	Durée normalisée du phonème
6	<i>ds</i>	Durée syllabique brute
7	<i>nps</i>	Nombre de phonèmes dans la syllabe
8	<i>npsa</i>	Nombre de phonèmes dans la syllabe (ambisyllabité)
9	<i>dsa</i>	Durée des syllabes (ambisyllabité)
10	<i>df</i>	Durée du pied accentuel
11	<i>npf</i>	Nombre de phonèmes dans le pied
12	<i>nsf</i>	Nombre de syllabes dans le pied
13	<i>dw</i>	Durée du mot
14	<i>npw</i>	Nombre de phonèmes dans le mot
15	<i>nsw</i>	Nombre de syllabes dans le mot
16	<i>di</i>	Durée de l'unité intonative
17	<i>npi</i>	Nombre de phonèmes dans l'unité intonative
18	<i>nsi</i>	Nombre de syllabes dans l'unité intonative
19	<i>nfi</i>	Nombre de pieds dans l'unité intonative
20	<i>nwi</i>	Nombre de mots dans l'unité intonative
21	<i>dr</i>	Durée de l'unité rythmique (Jassem)
22	<i>npr</i>	Nombre de phonèmes dans l'unité rythmique
23	<i>nsr</i>	Nombre de syllabes dans l'unité rythmique
24	<i>pho</i>	Catégorie de phonème
25	<i>ss</i>	Structure syllabique
26	<i>ssa</i>	Structure syllabique (ambisyllabité)
27	<i>str</i>	Statut accentuel du phonème

28	<i>t_{sm}</i>	Marque tonétique
29	<i>r_u</i>	Nature de l'unité rythmique
30	<i>i_{ps}</i>	Indice du phonème dans la syllabe
31	<i>i_{pr}</i>	Indice du phonème dans l'unité rythmique
32	<i>i_{pw}</i>	Indice du phonème dans le mot
33	<i>i_{pi}</i>	Indice du phonème dans l'unité intonative

Tableau 5. Énumération des différentes informations présentes dans le tableau de données

5.6.2. Remarques sur les découpages syllabiques

En ce qui concerne le regroupement syllabique des phonèmes selon le principe d'ambisyllabité ('*ssa*'), il convient de donner quelques explications quant à l'approche adoptée. Notons tout d'abord que ce regroupement syllabique en tenant compte du phénomène d'ambisyllabité n'apparaît pas comme un niveau supplémentaire d'annotation et d'alignement dans les fichiers-étiquettes, mais uniquement dans les données tabulaires.

Le codage des syllabes en fonction des consonnes ambisyllabiques ('*ssa*') présent dans les données est également effectué de manière automatique grâce à un script Perl. Ainsi, toute consonne intervocalique simple à l'intérieur d'un mot est systématiquement considéré comme ambisyllabique, à l'exception de la consonne /ŋ/ qui ne peut se trouver en position d'attaque de syllabe. Dans un mot tel que *city*, le phonème /t/ est codé '*A*' dans la colonne '*ssa*' et le nombre de phonèmes ('*n_{psa}*') est respectivement de trois dans la première syllabe (/sɪt/) et de deux dans la deuxième syllabe (/tɪ/).

Le découpage est plus délicat en cas de groupes consonantiques complexes en position intervocalique. Quand deux ou plus de deux consonnes se trouvent en position intervocalique, l'affiliation de ces consonnes est déterminée en fonction du caractère licite des attaques (d'après Gimson, réédité par Cruttenden 1997 ; cf. Figure 21 et Figure 22 de ce chapitre). Par exemple, dans *degree*, /gr/ est une attaque licite, c'est par conséquent le phonème /g/ qui est ambisyllabique. En revanche, le groupe consonantique /ŋkw/ dans le mot *inquisition* (/ɪŋkwɪzɪʃn/) ne représente pas une attaque licite en anglais contrairement à /kw/, c'est donc /k/ qui est ambisyllabique. La représentation pour ces deux mots est illustrée dans la Figure 24.

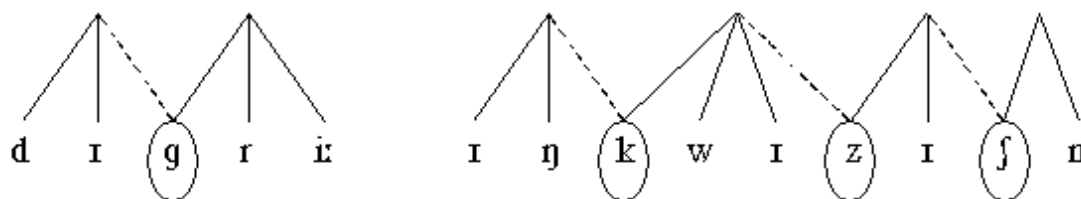


Figure 24. Regroupement des phonèmes en syllabes en tenant compte du principe d'ambisyllabité pour les mots *degree* et *inquisition*

Toutes les consonnes intervocaliques codées ambisyllabiques se trouvent à l'intérieur des mots. On ne s'intéresse pas ici, du moins pas dans ce travail, au phénomène de resyllabification des consonnes intervocaliques aux frontières de mots.

5.7. Résumé et conclusion

Dans ce chapitre, nous avons dans un premier temps présenté les différents types de corpus et explicité la notion de spontanéité de la parole. Nous avons choisi de travailler ici sur la base de données Aix-MARSEC car il s'agit d'une base de données comportant plus de cinq heures de parole authentique variée, à savoir de la parole à but communicatif et perçue comme naturelle.

Aix-MARSEC est également une base de données alignée temporellement aux niveaux des phonèmes, des constituants syllabiques, des syllabes, des mots (seuls et avec TSM) et des unités intonatives mineures et majeures. Grâce à un traitement automatique supplémentaire, il existe également un alignement en pieds accentuels et ictus silencieux selon la définition d'Abercrombie et Halliday et en unités rythmiques étroites et anacrouses selon le modèle de Jassem. C'est à partir de cet alignement que nous allons tester la première série d'hypothèses concernant la structuration rythmique.

Les constituants syllabiques sont doublement codés : en fonction du principe d'attaque maximale et en fonction du principe d'ambisyllabité. Ainsi, la base de données Aix-MARSEC nous permettra de tester la problématique d'ambisyllabité sur de la parole authentique.

Dans les deux chapitres suivants, nous approfondirons et testerons dans Aix-MARSEC les hypothèses soulevées au cours des quatre premiers chapitres concernant la structuration rythmique (Chapitre 6) et l'ambisyllabité (Chapitre 7) en anglais britannique.

CHAPITRE 6.
COMPARAISON ET EVALUATION
DES DEUX MODELES RYTHMIQUES

6.1. HYPOTHESES ET OUTILS STATISTIQUES.....	217
6.1.1. Récapitulatif	217
6.1.2. Hypothèses	218
6.1.2.1. Syllabe.....	219
6.1.2.2. Pied accentuel.....	219
6.1.2.3. Unité rythmique et anacrouse.....	220
6.1.2.4. Structure prosodique et accentuation	220
6.1.3. Données et outil statistique	220
6.1.3.1. Homogénéisation des données	221
6.1.3.2. Outil statistique	223
6.2. RESULTATS ET INTERPRETATION	225
6.2.1. Isochronie stricte	225
6.2.1.1. Passage C01.....	225
6.2.1.2. Totalité des données	228
6.2.2. Isochronie faible.....	230
6.2.2.1. Passage C01.....	230
6.2.2.2. Totalité des données	238
6.2.3. Effet du paramètre accentuel.....	243
6.2.3.1. Passage C01.....	243
6.2.3.2. Totalité des données	251
6.3. RESUME ET CONCLUSION.....	257

Nous allons nous intéresser dans ce chapitre à la première problématique soulevée dans ce travail, à savoir celle de l'organisation rythmique de l'anglais. Nous avons comparé les deux modèles différents proposés par Jassem (1952) et par Abercrombie (1967)/ Halliday (1970). La différence fondamentale entre les deux modèles réside dans la définition des unités rythmiques, la NRU et l'anacrouse pour Jassem et le pied accentuel pour Abercrombie/ Halliday. Notre but est donc de déterminer lequel de ces deux modèles représente le mieux la structure rythmique de l'anglais britannique.

La première section de ce chapitre est consacrée à la récapitulation et la formulation des hypothèses concernant la structuration rythmique, ainsi qu'à la description des outils statistiques choisis afin de tester les divers phénomènes. La deuxième section présente ensuite les différents résultats organisés par hypothèse et leur interprétation.

6.1. Hypothèses et outils statistiques

6.1.1. Récapitulatif

Nous avons présenté dans le Chapitre 3 deux modèles différents proposés pour rendre compte de l'organisation rythmique de l'anglais britannique. On a tout d'abord vu le modèle de Jassem (1952) dont la particularité est de découper les énoncés en NRU (regroupant une syllabe accentuée et les syllabes inaccentuées suivantes jusqu'à la fin du mot) et en anacrouses (les syllabes inaccentuées restantes). Le deuxième modèle est celui d'Abercrombie (1964) et de Halliday (1967), celui-ci étant généralement présenté comme le modèle supra-syllabique classique de l'anglais. Les énoncés sont ici découpés en pieds accentuels commençant par une syllabe accentuée et englobant toutes les syllabes inaccentuées suivantes jusqu'à la syllabe accentuée suivante mais sans l'inclure ou jusqu'à la pause suivante. En cas de syllabes inaccentuées en début d'unité

intonative, celles-ci sont précédées d'un ictus ou accent silencieux, permettant de maintenir un certain rythme.

Selon les auteurs respectifs, les NRU et les pieds accentuels sont caractérisés par la notion d'isochronie faible (Chapitre 4). Ceci signifie que la durée des sous-constituants (phonèmes/ syllabes) a tendance à diminuer plus ils sont nombreux dans le pied ou la NRU ; on a donc une compression de la durée des sous-constituants pour s'adapter à la durée du constituant auquel ils appartiennent.

La deuxième caractéristique de ces deux modèles est la durée similaire (ou quasi similaire) des syllabes à l'intérieur de la NRU et du pied : celles-ci se répartissent également la durée de l'unité rythmique quel que soit leur statut accentuel. Ainsi, il n'y a pas de différence de durée entre les syllabes accentuées et inaccentuées dans la NRU et dans le pied accentuel, ce qui va à l'encontre de l'acceptation générale selon laquelle l'accent affecte la durée des phonèmes et des syllabes (Klatt 1987 par exemple).

6.1.2. Hypothèses

Les différentes hypothèses dans cette première partie portent d'abord sur l'effet des différents constituants prosodiques sur la durée du phonème, de la syllabe, de l'unité rythmique et du pied, mais également du mot et de l'unité intonative. Ces informations sont obtenues à partir de l'alignement automatique aux niveaux de la NRU et de l'anacrouse, du pied accentuel, du mot et de l'unité intonative (Chapitre 5). Nous ferons référence aux pieds commençant par un ictus silencieux par le terme 'anacrouse initiale' puisque ce type de pied est une anacrouse (dans la signification attribuée par Jassem) en position initiale d'unité intonative.

En conséquence, nous allons dans un premier temps aborder la question de l'effet de la complexité (nombre de sous-constituants) sur la durée des constituants. L'hypothèse sous-jacente à cette première partie concerne l'isochronie que nous considérons sous deux facettes selon qu'il s'agit d'une application "stricte" ou "faible". La première suppose que la durée des constituants est identique quel que soit le nombre de sous-constituants ; la deuxième suppose qu'il existe un certain degré de compression des sous-constituants lorsque leur nombre augmente dans un constituant donné.

Nous considérons ici l'isochronie dans sa définition globale dans le sens où elle peut s'appliquer à tous les niveaux de constituance et pas seulement aux niveaux du pied accentuel et de la NRU. Les diverses analyses porteront par conséquent sur les différents

niveaux hiérarchiques de constituance, à savoir la syllabe, l'anacrouse initiale et le pied accentuel (modèle d'Abercrombie/ Halliday), l'anacrouse et la NRU (modèle de Jassem), et l'unité intonative. Nous ajouterons également le niveau du mot. En testant les hypothèses d'isochronie stricte et d'isochronie faible, nous allons déterminer lequel des modèles de Jassem ou d'Abercrombie correspond le mieux à la structuration rythmique de l'anglais. Nous espérons par conséquent que nos résultats nous permettront de valider un des deux modèles.

La deuxième hypothèse concerne l'effet du paramètre accentuel. Les deux modèles (Jassem et Abercrombie/ Halliday) supposent qu'il n'existe aucun effet intrinsèque de l'accent : les syllabes se répartissent également la durée des NRU ou des pieds accentuels selon le modèle en question. Nous testerons dans un deuxième temps l'effet du paramètre accentuel de manière globale puis dans les deux unités rythmiques.

Les hypothèses que nous testons dans la deuxième section de ce chapitre sont résumées dans les sous-sections suivantes.

6.1.2.1. Syllabe

Comme nous l'avons remarqué précédemment, l'hypothèse de l'isochronie est à considérer sous deux facettes, celle de l'isochronie stricte et celle de l'isochronie faible. Ainsi, l'hypothèse de l'isochronie stricte au niveau de la syllabe suppose que la complexité de la syllabe n'a aucune influence sur sa durée ; la durée de la syllabe est constante quel que soit le nombre de sous-constituants (phonèmes).

En revanche, selon l'hypothèse de l'isochronie faible, la syllabe affecte la durée de ses sous-constituants qui ont tendance à être plus courts lorsqu'ils sont plus nombreux. On s'attend donc à un effet négatif significatif de la complexité de la syllabe sur la durée des phonèmes.

6.1.2.2. Pied accentuel

Le même type de raisonnement s'applique au niveau du pied accentuel. L'hypothèse d'isochronie stricte suppose que la complexité du pied n'a aucune influence sur sa durée. Selon l'hypothèse de l'isochronie faible, le pied en tant que constituant phonologique affecte la durée de ses sous-constituants (syllabes et phonèmes) qui ont tendance à devenir plus courts lorsqu'ils sont plus nombreux. On attend par conséquent

un effet négatif significatif de la complexité du pied sur la durée de ses sous-constituants (syllabes et phonèmes).

6.1.2.3. Unité rythmique et anacrouse

Les hypothèses soulevées concernant la NRU et l'anacrouse sont identiques à celles pour la syllabe et le pied accentuel ; il est toutefois important de préciser que la NRU est comprise dans le pied, on s'attend par conséquent à ce que les effets observés soient moins marqués que pour le pied. Ainsi, en ce qui concerne l'isochronie stricte, on suppose que la durée de la NRU n'est pas influencée par sa complexité : sa durée est constante quel que soit le nombre de sous-constituants. En revanche, l'hypothèse de l'isochronie faible implique que la NRU affecte la durée de ses sous-constituants, ceci se traduisant par un effet négatif de la complexité de la NRU sur la durée des sous-constituants.

D'après le modèle de Jassem, les phonèmes compris dans l'anacrouse sont prononcés aussi rapidement que possible, leur durée ne devrait donc pas être affectée par la complexité de l'anacrouse.

6.1.2.4. Structure prosodique et accentuation

Dans les deux modèles, la présence d'un accent n'affecte pas la durée des syllabes (et de ses sous-constituants) puisque les syllabes se partagent équitablement la durée globale du pied ou de la NRU. Comme nous l'avons montré dans le Chapitre 3, cette caractéristique de leurs modèles est en contraste avec l'hypothèse défendue habituellement dans la littérature selon laquelle les syllabes accentuées (ainsi que les phonèmes dont elles sont composées) sont plus longues que les syllabes inaccentuées (Klatt 1987 ; Fant, Kruckenberg et Nord 1991a).

6.1.3. Données et outils statistiques

Toutes les analyses sont effectuées sur la base de données Aix-MARSEC. Au-delà des différents traitements détaillés dans le Chapitre 5, certaines décisions ont dû être prises pour s'assurer que les données, et donc les résultats, sont fiables. Ces décisions sont détaillées dans la section suivante. Nous verrons ensuite les tests statistiques utilisés pour tester les hypothèses que nous venons de soulever.

6.1.3.1. Homogénéisation des données

La durée des phonèmes est fortement influencée par l'identité des phonèmes (Klatt 1987) ; c'est pour cette raison que nous utilisons la durée normalisée des phonèmes plutôt que leur durée brute en tant que variable dépendante. La durée normalisée des phonèmes est obtenue en soustrayant la durée moyenne du phonème à la valeur brute et en divisant le résultat par l'écart type du phonème en question²⁵. Ainsi, les différences intrinsèques entre les phonèmes sont neutralisées et il est donc possible de travailler sur la durée normalisée globale des phonèmes sans tenir compte de la catégorie ou de l'identité des phonèmes.

Toutefois, il convient de préciser que la moyenne et l'écart type de chaque phonème sont obtenus sur l'ensemble de la base de données Aix-MARSEC. Il se peut donc que la durée normalisée calculée à partir de ces deux valeurs soit biaisée car elle ne tient pas compte des différences entre les 53 locuteurs. En raison de cette possibilité de biais, toutes les analyses sont faites sur les durées normalisées mais également sur les durées brutes. Si les résultats sont similaires avec les deux types de valeurs, nous serons en mesure de conclure que les résultats sont fiables.

De plus, pour s'assurer que les durées normalisées de phonèmes ne sont pas biaisées, les différentes analyses sont dans un premier temps effectuées sur une partie du corpus produite par un seul locuteur. Pour cela, nous avons recherché dans la base de données le locuteur qui s'exprimait le plus longtemps ; il s'agit du passage C01 qui représente environ 30 minutes et qui est réalisé par un locuteur unique, David Henderson. Nous avons par conséquent effectué les analyses sur ce passage en recalculant au préalable les moyennes et écarts types de chaque phonème de ce locuteur et ensuite les durées normalisées des phonèmes. Une fois ces analyses effectuées et en fonction des résultats, nous faisons les mêmes tests sur l'ensemble des données du corpus. Si les résultats sont similaires, on peut considérer que les tests effectués sur le corpus entier ne sont pas biaisés par l'influence des 53 différents locuteurs (pour le

²⁵ La durée normalisée z_d des phonèmes est obtenue grâce à la formule suivante :

$$z_d = (d - \text{mean}_d) / \text{sd}_d$$

où d est la durée du phonème, mean_d sa durée moyenne et sd_d son écart type.

calcul des durées normalisées). Par conséquent, tous les tests sont effectués dans un premier temps sur le passage, puis sur la totalité des données.

Les deux étapes de ces analyses présentent un autre avantage, au niveau statistique cette fois. Nous avons remarqué que travailler sur une quantité importante de données (comme c'est le cas avec Aix-MARSEC) pose pour les analyses statistiques le problème du « trop de données ». En effet, du fait du nombre important de données, tous les effets testés deviennent très significatifs même si leurs effets sont très petits. Ainsi, grâce aux tests intermédiaires sur le passage C01 comportant moins de données, nous avons la possibilité de comparer les niveaux de significativité de chaque test et de valider, ou d'invalidier, l'effet du facteur en question.

Contrairement aux phonèmes, nous utiliserons les valeurs brutes pour la durée des constituants de niveau supérieur, à savoir la syllabe, le pied accentuel, la NRU, l'anacrouse, l'anacrouse initiale, le mot et l'unité intonative. Il n'était en effet pas dans nos moyens de calculer les valeurs normalisées pour la durée de ces unités. On suppose toutefois que la durée de ces constituants est beaucoup moins variable que celle des phonèmes, la prise en compte des durées brutes ne devrait par conséquent pas altérer les résultats.

Avant de commencer à tester les diverses hypothèses concernant l'isochronie, il est nécessaire d'effectuer un premier travail d'inspection des données afin de déterminer le ou les tests statistiques les plus appropriés. La première remarque est que les données n'ont pas une distribution normale. Si on observe la répartition des durées segmentales brutes dans les données, on s'aperçoit que certains cas se situent au-delà de 500ms (119 valeurs sur près de 200 000 phonèmes) et on suppose qu'il s'agit d'erreurs d'alignement, dues par exemple à une mauvaise détection de la pause qui suit. Tous les phonèmes dont la durée excède 500ms sont par conséquent exclus des analyses.

Nous avons adopté la même approche pour les phonèmes dont la durée est inférieure à 15ms puisqu'il s'agit probablement également d'erreurs d'alignement dues à la méthode de segmentation utilisée. Rappelons que lorsque l'aligneur ne détecte pas un phonème dans le signal, il lui attribue une durée fixe de 10ms. Il est donc raisonnable de supprimer systématiquement ces phonèmes. Nous avons étendu le filtrage des données à 15ms pour supprimer une partie des données aberrantes (certainement dues à des erreurs d'alignement) pour qu'elles ne biaisent pas les résultats. Par conséquent, seules les durées phonémiques entre 15 et 500ms ont été utilisées pour les différents tests

statistiques. La répartition des durées segmentales brutes ainsi obtenue est présentée dans la Figure 25 ci-après.

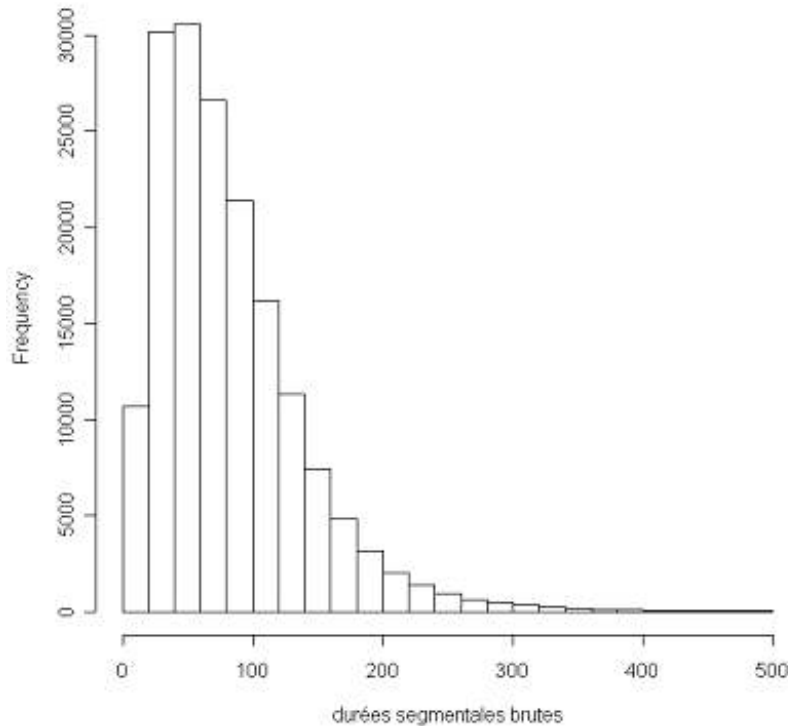


Figure 25. Répartition des durées segmentales brutes dans la base de données Aix-MARSEC

6.1.3.2. Outils statistiques

Pour toutes les analyses statistiques, nous utilisons le logiciel R²⁶, un langage et un environnement pour les analyses statistiques et leurs représentations graphiques. Il présente les avantages suivants :

- Il est multi-plateforme : on peut l'utiliser sous Windows, Mac ou Unix
- Il est puissant dans le sens où il permet de travailler sur un nombre important de données (contrairement à d'autres logiciels tels que Statview)
- Il ne pose pas le problème de caractères inconnus : certains logiciels ne reconnaissent pas certains caractères, comme les symboles phonétiques ou d'autres symboles utilisés pour des annotations (comme les systèmes tournant sous MS-DOS).

²⁶ Téléchargeable gratuitement à l'adresse suivante : <http://www.r-project.org/>.

- Il permet de conserver des données et des analyses en mémoire : ceci rend son utilisation plus facile et plus rapide puisque les données sont stockées et il est possible de sauvegarder toutes les modifications d'une utilisation à l'autre.

Il se présente sous la forme d'une console et les diverses commandes se font selon un langage spécifique à R. Les données utilisées doivent être présentées sous la forme de données tabulaires avec ici à chaque ligne un phonème suivi d'un ensemble d'informations. Chacune de ces informations devient alors une variable dans R.

Les analyses statistiques utilisées pour tester les différentes hypothèses sont des analyses de corrélation et de régression linéaire et des analyses de (co)variance, le tout dans le logiciel de statistique R. L'analyse de corrélation est principalement utilisée pour tester l'hypothèse d'isochronie ; elle permet de tester le lien existant entre une variable indépendante et la variable dépendante qu'elle est supposée conditionner. Une corrélation forte signifie par conséquent que la variable indépendante conditionne la variable dépendante. L'analyse de (co)variance est utilisée pour tester l'effet de l'accent ; elle permet en effet de tester l'influence d'un facteur catégoriel donné (analyse de variance) ou de l'influence couplée de deux facteurs, l'un catégoriel, l'autre numérique (analyse de covariance) sur la durée des constituants.

En ce qui concerne les analyses de corrélation, nous utilisons plus précisément une analyse de régression linéaire qui se décompose en deux mesures : le coefficient de détermination (R^2) et le coefficient de régression linéaire (ou pente). Dans le contexte de ce travail, ces deux mesures nous permettent d'obtenir les informations suivantes :

- Le *coefficient de détermination* nous indique la part de variabilité de la durée d'un constituant déterminée par sa complexité. Prenons l'exemple d'une analyse de régression linéaire sur la durée de la syllabe en fonction du nombre de phonèmes qu'elle contient : un R^2 de 0.2885 indique que 28,85% de la variabilité de la durée de la syllabe sont déterminés par le nombre de phonèmes qu'elle contient.
- La *pente* permet de savoir de combien augmente ou diminue la durée des constituants pour chaque sous-constituant ajouté. Si on reprend l'exemple cité plus haut et que la pente est par exemple de 69.77, ceci signifie que pour chaque phonème ajouté dans une syllabe, on a une augmentation moyenne de 69.77ms.

Pour toutes les analyses effectuées, les résultats complets obtenus dans R sont présentés en annexe (Annexe 2) dans l'ordre dans lequel ils sont présentés dans le texte.

Les différentes hypothèses de corrélation concernant l'isochronie sont résumées dans le Tableau 6 ci-après. L'isochronie stricte suppose premièrement qu'il n'y a pas de corrélation (notée '0' dans le tableau) entre la durée des constituants et leur complexité et deuxièmement qu'il existe une corrélation négative (notée '-') entre la durée des sous-constituants et la complexité du constituant auquel ils appartiennent. L'isochronie faible suppose qu'il y a une corrélation positive (notée '+') entre la durée des constituants et leur complexité et qu'il existe une corrélation négative entre la durée des sous-constituants et la complexité du constituant auquel ils appartiennent. Enfin, s'il n'y a aucune isochronie, nous devrions avoir une corrélation positive entre la durée des constituants et leur complexité et aucune corrélation entre la durée des sous-constituants et la complexité du constituant auquel ils appartiennent.

	Durée des constituants	Durée des sous- constituants
Isochronie stricte	0	-
Isochronie faible	+	-
Aucune isochronie	+	0

Tableau 6. Résumé des différentes prédictions selon les hypothèses de corrélation concernant l'isochronie

6.2. Résultats et interprétation

Chaque hypothèse (isochronie stricte, isochronie faible et effet de l'accent) est testée dans un premier temps sur le passage C01, puis sur la totalité des données Aix-MARSEC afin de voir si les tendances sont les mêmes dans les deux cas.

6.2.1. Isochronie stricte

6.2.1.1. Passage C01

L'isochronie stricte suppose que la durée des constituants n'est pas influencée par le nombre de sous-constituants (complexité). Pour tester cette hypothèse, nous utilisons l'analyse de régression linéaire portant sur la durée des différents constituants (syllabes, pieds, NRU, anacrouses, anacrouses initiales, mots et unités intonatives) en fonction de

la complexité en terme de sous-constituants (phonèmes, syllabes, pieds accentuels ou mots selon les cas).

Les résultats concernant l'hypothèse d'isochronie stricte sur le passage C01 sont présentés dans le Tableau 7 ci-après (cf. Annexe 2.1.1 pour les résultats détaillés) avec pour chaque analyse le coefficient de détermination R^2 sur la première ligne, le coefficient de régression linéaire (pente) sur la deuxième ligne et la valeur p sur la troisième ligne. Le tableau se lit de la manière suivante : on effectue une analyse de régression en testant l'effet de la complexité (première ligne : nombre de phonèmes, syllabes, pieds ou mots) sur la durée des constituants (première colonne : syllabe, pied, NRU, anacrouse, anacrouse initiale, mot et unité intonative). Par exemple, le premier résultat de l'analyse de régression porte sur la durée de la syllabe en fonction du nombre de phonèmes dans la syllabe, cette analyse donnant comme résultat un R^2 de 0,427, une pente de 74,726 et une valeur p inférieure à $2.2e-16$.

Toutes les analyses sont très significatives ($p < 2.2e-16$), cette valeur étant la plus petite valeur possible dans R. Pour toutes les analyses, le coefficient de régression linéaire est positif, ce qui signifie que la durée des différents constituants augmente en fonction de leur complexité, il n'y a donc pas d'isochronie stricte.

Par exemple, le pied accentuel augmente d'environ 54ms pour chaque phonème ajouté et la NRU d'environ 51ms ; le pied accentuel augmente d'environ 115ms pour chaque syllabe ajoutée et la NRU d'environ 100ms. La complexité en terme de phonèmes et de syllabes explique respectivement environ 64% ($R^2=0.6462$) et 50% ($R^2=0.506$) de la variabilité de la durée des pieds accentuels. De même, la complexité en terme de phonèmes et de syllabes explique respectivement environ 51% ($R^2=0.515$) et 31% ($R^2=0.312$) de la variabilité de la durée des NRU. Il est par conséquent évident que la durée de ces constituants, tout comme celle des autres constituants, n'est pas constante quel que soit le nombre de phonèmes/ syllabes qu'ils contiennent.

		Nombre de phonèmes	Nombre de syllabes	Nombre de pieds	Nombre de mots
Syllabe	R ²	0,427			
	pente	74,726	-	-	-
	p	<2.2e-16			
Pied	R ²	0,6462	0,506		
	pente	54,636	115,411	-	-
	p	<2.2e-16	<2.2e-16		
NRU	R ²	0,515	0,312		
	pente	51,527	100,097	-	-
	p	<2.2e-16	<2.2e-16		
ANA	R ²	0,664	0,566		
	pente	54,491	119,204	-	-
	p	<2.2e-16	<2.2e-16		
ANA initiale	R ²	0,7136	0,637		
	pente	54,456	123,447	-	-
	p	<2.2e-16	<2.2e-16		
Mot	R ²	0,757	0,633		
	pente	66,651	156,709	-	-
	p	<2.2e-16	<2.2e-16		
Unité Intonative	R ²	0,5242	0,5107	0,495	0,435
	pente	65,099	156,46	367,05	248,229
	p	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16

Tableau 7. Résultats des analyses de régression linéaire concernant l'hypothèse d'isochronie stricte pour le passage C01. Les lignes correspondent aux variables dépendantes (durée des constituants) et les colonnes aux variables indépendantes (complexité)

Le résultat concernant la durée de la syllabe en fonction de sa complexité en terme de nombre de phonèmes signifie qu'environ 42% de la variabilité de la durée syllabique est expliquée par le nombre de phonèmes dans la syllabe. La pente est ici particulièrement importante puisque pour chaque phonème supplémentaire dans une syllabe, celle-ci augmente environ de 74ms, la durée moyenne des phonèmes étant de

81ms. Par conséquent, la durée de la syllabe dépend fortement de sa complexité. Ce résultat est illustré par la Figure 26 ci-après représentant la durée des syllabes en fonction du nombre de phonèmes qu'elles contiennent.

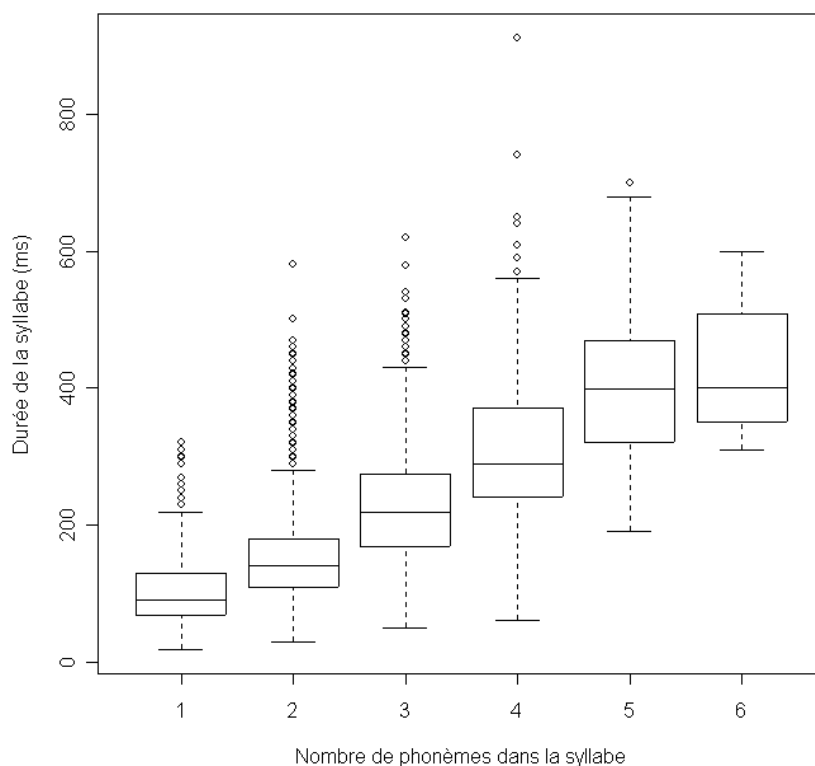


Figure 26. Box plot de la durée des syllabes en fonction du nombre de phonèmes qu'elles contiennent (C01)

Il ressort de cette figure que la durée des syllabes augmente quasi linéairement en fonction de leur complexité. Même si cette figure ne tient pas compte de l'effet de l'accent, elle donne une première indication concernant l'effet de la complexité sur la durée des syllabes.

6.2.1.2. Totalité des données

Les mêmes analyses concernant l'hypothèse d'isochronie stricte sont effectuées cette fois sur la totalité des données (soit environ 168 000 phonèmes). Les résultats des analyses de régression linéaire sont présentés dans le Tableau 8 ci-après (cf. Annexe 2.1.2 pour les résultats détaillés). Tous les résultats de la régression linéaire dans le cadre de cette hypothèse d'isochronie stricte sont très significatifs ($p < 2.2e-16$). Pour

toutes les analyses, le coefficient de régression linéaire est positif ; ceci signifie que la durée des différents constituants prosodiques augmente en fonction du nombre de sous-constituants qu'ils contiennent. Il n'existe par conséquent pas d'isochronie stricte, et ceci quel que soit le niveau de constituance.

		Nombre de phonèmes	Nombre de syllabes	Nombre de pieds	Nombre de mots
Syllabe	R ²	0,297			
	pente	71,203	-	-	-
	p	<2,2e-16			
Pied accentuel	R ²	0,451	0,368		
	pente	51,927	110,253	-	-
	p	<2,2e-16	<2,2e-16		
NRU	R ²	0,363	0,24		
	pente	51,455	103,032	-	-
	p	<2,2e-16	<2,2e-16		
ANA	R ²	0,465	0,384		
	pente	53,63	119,554	-	-
	p	<2,2e-16	<2,2e-16		
ANA initiale	R ²	0,392	0,341		
	pente	52,918	119,862	-	-
	p	<2,2e-16	<2,2e-16		
Mot	R ²	0,604	0,48		
	pente	68,31	158,081	-	-
	p	<2,2e-16	<2,2e-16		
Unité intonative	R ²	0,576	0,562	0,545	0,515
	pente	67,758	163,962	366,33	232,406
	p	<2,2e-16	<2,2e-16	<2,2e-16	<2,2e-16

Tableau 8. Résultats des analyses de régression linéaire pour l'hypothèse d'isochronie stricte sur la totalité des données avec comme variable dépendante la durée des constituants et comme variable indépendante leur complexité

Les résultats obtenus sur la totalité des données sont comparables à ceux obtenus pour le passage C01 ; il est toutefois important de remarquer que cette similitude vient

certainement de la variabilité moins marquée de la durée des constituants supérieurs, en comparaison avec celle des phonèmes. Malgré cette proximité des résultats, on remarque que, à l'exception de l'unité intonative, les R^2 sont moins élevés pour la totalité des données que pour le passage C01, les pentes étant, elles, similaires. Par exemple, la complexité de la syllabe permet de rendre compte d'environ 30% de la variabilité de la durée des syllabes dans la totalité des données, contre environ 42% dans le passage C01, les pentes sont, elles, voisines (71.203 et 74.726 respectivement). Ces divergences s'expliquent par le rajout d'une variable supplémentaire, celle du locuteur puisque le passage C01 est produit par un seul locuteur, alors que la totalité des données englobe 53 locuteurs différents.

6.2.2. Isochronie faible

6.2.2.1. Passage C01

Pour tester l'hypothèse de l'isochronie faible, nous avons également utilisé des analyses de régression linéaire (cf. Annexe 2.2.1 pour les résultats détaillés). Dans ces analyses, nous cherchons à voir si la durée des sous-constituants (phonèmes, syllabes et pieds) est affectée par la complexité des constituants (nombre de phonèmes, de syllabes et de pieds selon les cas).

Les résultats obtenus concernant l'effet de la complexité en terme de phonèmes sur la durée des phonèmes sont présentés dans le Tableau 9 ci-après (cf. Annexe 2.2.1.1). Les résultats non-significatifs sont représentés en gras, il s'agit notamment de l'effet de la complexité des syllabes sur la durée normalisée des phonèmes.

		Durée des phonèmes	
		zp	dp
Nombre de phonèmes dans la syllabe	R ²	0,000	0,000
	pente	0,013	0,893
	p	0,1244	0,021
Nombre de phonèmes dans le pied	R ²	0,009	0,019
	pente	-0,034	-2,346
	p	<2,2e-16	<2,2e-16
Nombre de phonèmes dans la NRU	R ²	0,018	0,044
	pente	-0,067	-5,019
	p	<2,2e-16	<2,2e-16
Nombre de phonèmes dans l'ANA	R ²	0,005	0,006
	pente	-0,039	-1,808
	p	6,39e-05	7,02e-06
Nombre de phonèmes dans l'ANA initiale	R ²	0,003	0,005
	pente	-0,029	-1,518
	p	0,023	0,006
Nombre de phonèmes dans le mot	R ²	0,005	0,01
	pente	-0,024	-1,608
	p	<2,2e-16	<2,2e-16
Nombre de phonèmes dans l'UI	R ²	0,007	0,009
	pente	-0,009	-0,498
	p	<2,2e-16	<2,2e-16

Tableau 9. Résultats des analyses de régression linéaire en ce qui concerne l'hypothèse d'isochronie faible dans le passage C01. Les lignes représentent les variables indépendantes (complexité des différents constituants en terme phonèmes) et les colonnes les variables dépendantes (durée des phonèmes). 'zp' et 'dp' sont respectivement les durées normalisée et brute des phonèmes.

Remarques concernant les résultats

On remarque en premier lieu que les R² et les pentes sont globalement très faibles. Cette particularité des résultats s'explique en partie par le nombre important de facteurs influençant la durée des phonèmes. Nous testons ici uniquement l'effet de la complexité sur la durée des phonèmes, mais de nombreux autres paramètres influencent ces durées

et ceux-ci ne sont pas pris en compte dans cette analyse. C'est probablement en raison de cette multitude de facteurs externes que les R^2 et les pentes sont si faibles, même si les valeurs p restent toutefois significatives dans la plupart des cas.

Devant des valeurs aussi faibles (R^2 et pentes), nous nous sommes posée un certain nombre de questions concernant la méthode adoptée. Pour s'assurer que l'erreur ne venait pas du filtrage des données (exclusion de tous les phonèmes dont la durée se situe en dessous de 15ms et en dessus de 500ms) ou du choix des outils statistiques, nous avons effectué quelques analyses supplémentaires.

Pour déterminer si les résultats ne sont pas biaisés par l'effet des valeurs aberrantes (notamment les erreurs d'alignement), nous avons effectué les mêmes analyses sur des données situées entre 20 et 200ms (au lieu de 15/ 500ms). Les résultats sont similaires, on peut donc conclure que les résultats ne sont pas, ou très peu, influencés par les valeurs aberrantes.

Pour nous assurer que notre test statistique est bien approprié, nous avons eu recours grâce à l'aide de R. Espesser à une analyse de régression avec un modèle mixte à partir des durées brutes. Les modèles classiques supposent l'indépendance des observations. En revanche, les modèles à effets mixtes (dits aussi modèles mixtes) s'appliquent à des données structurées par groupe ou blocs, et donc caractérisées par la présence de corrélations entre observations à l'intérieur d'un même groupe (Pinheiro et Bates 2000). Par exemple, les analyses utilisant un modèle mixte nous permettent de faire une analyse de corrélation avec les facteurs utilisés précédemment utilisés (durée brute des phonèmes et complexité) mais également avec le facteur d'identité des phonèmes. Les résultats obtenus avec le modèle mixte sont tout à fait comparables à ceux obtenus avec la durée normalisée des phonèmes. Nous en concluons par conséquent que les résultats sur les durées normalisées sont fiables et qu'il n'est pas nécessaire d'utiliser une méthode différente. Sans ignorer les résultats obtenus pour la durée brute, nous nous fierons plutôt à ceux obtenus sur la durée normalisée.

La troisième possibilité de biais réside dans la qualité de l'alignement automatique. En effet, il existe de nombreuses erreurs d'alignement et ces erreurs peuvent influencer grandement les résultats. Nous supposons qu'étant donné la quantité de données prise en compte, les erreurs se compensent et qu'elles n'influencent pas les résultats. Il conviendrait toutefois de le vérifier, par exemple sur de l'alignement manuel. En effet, nous possédons depuis peu environ 90 minutes du corpus alignées manuellement. Il est

par conséquent en projet d'effectuer les tests statistiques sur cette portion alignée manuellement afin de comparer les résultats avec ceux obtenus avec l'alignement automatique.

Toutefois, malgré des résultats peu marqués (moins que ceux sur l'isochronie stricte par exemple), les résultats restent tout à fait interprétables.

Interprétation des résultats

Tous les résultats sont significatifs, sauf pour l'effet du nombre de phonèmes dans la syllabe sur la durée normalisée des phonèmes ($p=0.1244$). Nous retiendrons le résultat obtenu pour la durée normalisée, mais on peut noter que la valeur p pour la durée brute est assez élevée, surtout en comparaison avec la plupart des valeurs p suivantes. Ce résultat signifie que la complexité de la syllabe n'a aucun effet sur la durée des phonèmes. Il semblerait par conséquent qu'il n'y ait aucune compression des phonèmes quelle que soit la taille de la syllabe, ce qui rejoint le résultat obtenu dans le cadre du test de l'hypothèse d'isochronie stricte.

Si on observe la représentation des durées normalisées moyennes de phonèmes en fonction de la complexité de la syllabe (Figure 27), on s'aperçoit que celles-ci sont relativement constantes quelle que soit la complexité de la syllabe. Il ne semble donc pas y avoir de compression des phonèmes et la durée de la syllabe augmente linéairement en fonction du nombre de phonèmes qu'elle contient. Ce résultat confirme les remarques de Hill, Jassem et Witten (1978a) ; en revanche, il contredit directement le modèle de Campbell (1992) qui décrit un phénomène d'ajustement des durées phonémiques dans le cadre de la syllabe. Or, il est évident dans les résultats obtenus que la durée des phonèmes reste constante et que, par conséquent, la durée des syllabes augmente quasi linéairement en fonction de leur complexité.

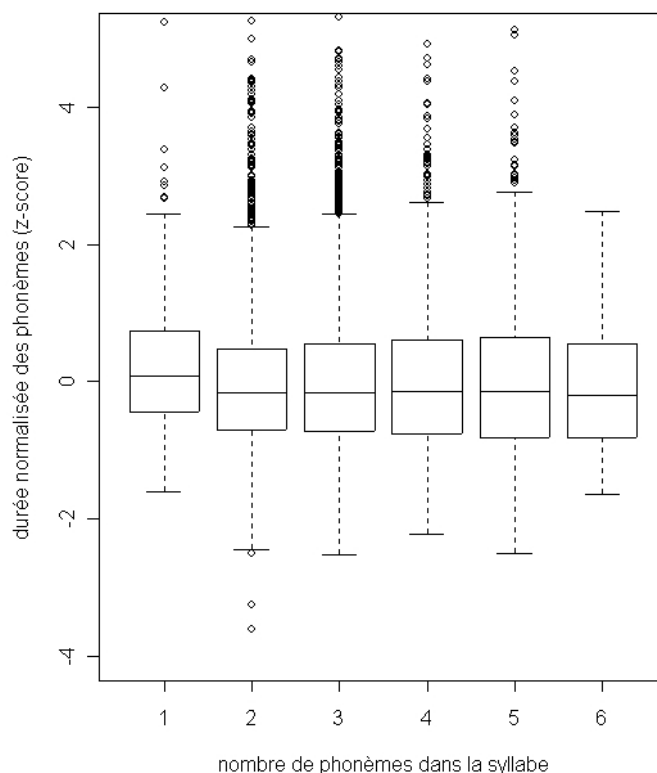


Figure 27. *Durée normalisée des phonèmes en fonction du nombre de phonèmes dans la syllabe (passage C01)*

Bien que les effets soient très faiblement marqués, il est intéressant de comparer les résultats obtenus pour le pied accentuel et pour la NRU (très significatifs pour les durées normalisées et brutes). On s'aperçoit que la pente est presque deux fois plus importante dans la NRU que dans le pied, le R^2 étant lui aussi plus important. Il y a donc plus de compression dans la NRU que dans le pied, ceci allant dans le sens du modèle décrit par Jassem. Ces compressions sont faibles puisque les pentes sont de -0.034 pour le pied et de -0.067 pour la NRU (durée normalisée), ceci représentant une compression d'environ -2.346ms pour le pied et -5.019ms pour la NRU pour chaque phonème ajouté. On peut d'ailleurs se demander si une compression d'environ 2 ou 5ms par phonème ajouté est perceptible ; il est toutefois possible que cette compression, si petite soit-elle, contribue par exemple au caractère naturel de la parole synthétique (Campbell 1992a).

En ce qui concerne l'anacrouse et l'anacrouse initiale, on peut voir qu'il existe un faible degré de compression qui s'avère être similaire pour les deux unités (-0.039 pour l'anacrouse et -0.029 pour l'anacrouse initiale, les deux étant en terme de durée normalisée). Il n'y a donc quasiment pas de compression dans les deux types d'anacrouse. La même conclusion peut s'appliquer au mot et à l'unité intonative.

Le Tableau 10 et le Tableau 11 ci-après présentent les résultats des analyses de régression considérant cette fois l'effet sur la durée des phonèmes du nombre de syllabes (Tableau 10) et de pieds (Tableau 11) dans les divers constituants (cf. respectivement Annexes 2.2.1.2 et 2.2.1.3 pour les résultats détaillés).

Les résultats obtenus dans ces tableaux suivent les mêmes tendances, à savoir que les R^2 et les pentes sont très faibles pour toutes les analyses. La compression de la durée (brute et normalisée) des phonèmes est supérieure dans la NRU, en comparaison avec celle dans le pied accentuel. L'effet est non significatif dans l'anacrouse initiale (il n'y a donc pas de compression) et la valeur p dans l'anacrouse est nettement plus importante que dans les autres constituants.

		Durée des phonèmes	
		zp	dp
Nombre de syllabes dans le pied	R^2	0,008	0,018
	pente	-0,08	-5,55
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans la NRU	R^2	0,015	0,039
	pente	-0,152	-11,676
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans l'ANA	R^2	0,002	0,003
	pente	-0,063	-2,999
	p	0,007	0,002
Nombre de syllabes dans l'ANA initiale	R^2	-0,001	0,000
	pente	-0,016	-1,261
	p	0,599	0,332
Nombre de syllabes dans le mot	R^2	0,004	0,008
	pente	-0,051	-3,497
	p	4,65e-15	<2,2e-16
Nombre de syllabes dans l'UI	R^2	0,006	0,007
	pente	-0,022	-1,131
	p	<2,2e-16	<2,2e-16

Tableau 10. Résultats des analyses de régression (C01) concernant la durée des phonèmes en fonction du nombre de syllabes dans les constituants supérieurs

		Durée des phonèmes	
		zp	dp
Nombre de pieds dans l'UI	R ²	0,003	0,001
	pente	-0,036	-1.236
	p	1,93e-12	3,23e-07

Tableau 11. Résultats des analyses de régression (C01) concernant la durée des phonèmes en fonction du nombre de pieds dans l'unité intonative

Nous avons également effectué des analyses de régression pour tester l'effet de la complexité en terme de syllabes sur la durée des syllabes. Les résultats obtenus pour la durée brute et logarithmique des syllabes sont présentés dans le Tableau 12 ci-après (Annexe 2.2.1.4). A défaut d'avoir la durée normalisée des syllabes, nous avons également utilisé la durée logarithmique des syllabes qui permet de se rapprocher d'une distribution normale des durées syllabiques (notamment en compensant le tassement des données limitées à gauche). Les résultats obtenus avec les durées logarithmiques sont globalement similaires à ceux obtenus avec les durées brutes, probablement parce que la durée des syllabes est beaucoup moins variable (en comparaison avec des durées phonémiques par exemple).

Les effets sont non significatifs pour l'anacrouse et l'anacrouse initiale, il n'y a donc pas de compression, ce qui confirme les résultats obtenus pour la durée des phonèmes. Il est intéressant de voir que les résultats pour le pied accentuel et la NRU sont ici très différents ; que ce soit pour la durée brute ou pour la durée logarithmique, le R² est nettement plus élevé pour la NRU (R²=0.199/0.191) que pour le pied (R²=0.074/0.062). Par conséquent, dans les NRU, la durée des syllabes diminue en fonction du nombre de syllabes (diminution d'environ 56ms pour chaque syllabe ajoutée), cet effet étant très significatif. En revanche, une telle tendance ne s'observe pas dans le pied accentuel : on observe une certaine compression mais nettement plus faible (pente= -24.182/-0.115) et la complexité du pied ne rend compte que d'environ 7% de la variabilité des durées syllabiques (environ 6% pour les durées logarithmiques).

Les résultats concernant le mot et l'unité intonative sont également peu marqués (R² et pentes faibles), avec pour les deux unités des pentes négatives pour les durées syllabiques brutes et logarithmiques. Les pentes sont plus importantes pour le mot que

pour l'unité intonative, mais les R^2 sont, eux, très proches. Il existe par conséquent un phénomène de compression faiblement marqué au sein du mot et de l'unité intonative.

		Durée des syllabes	
		ds	log(ds)
Nombre de syllabes dans le pied	R^2	0,074	0,062
	pente	-24,182	-0,115
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans la NRU	R^2	0,199	0,191
	pente	-56,611	-0,261
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans l'ANA	R^2	0,000	0,002
	pente	-2,692	-0,03
	p	0,197	0,065
Nombre de syllabes dans l'ANA initiale	R^2	-0,001	-0,001
	pente	-2,088	-0,018
	p	0,474	0,445
Nombre de syllabes dans le mot	R^2	0,021	0,013
	pente	-12,213	-0,05
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans l'UI	R^2	0,019	0,014
	pente	-3,977	-0,018
	p	<2,2e-16	<2,2e-16

Tableau 12. Résultats des analyses de régression (C01) concernant la durée des syllabes en fonction de la complexité des constituants supérieurs en terme de nombre de syllabes

Il ressort de ces diverses analyses qu'il n'existe aucune compression des phonèmes dans la syllabe : celle-ci augmente linéairement en fonction du nombre de phonèmes qu'elle contient. Le deuxième résultat intéressant concerne le pied et la NRU puisqu'on observe plus de compression des phonèmes dans la NRU que dans le pied (en fonction de leur complexité en terme de phonèmes et de syllabes). Si on considère les durées syllabiques, cette tendance s'accroît puisque la compression des durées syllabiques dans les NRU est environ deux fois plus importante et la complexité des NRU en terme de

syllabes rend compte d'environ 20% des durées syllabiques (pour environ 7% dans le cas des pieds accentuels). Les mêmes analyses sont ensuite effectuées dans la totalité des données.

6.2.2.2. Totalité des données

Les résultats des analyses de régression pour la totalité des données sont présentés dans le Tableau 13 ci-après (cf. Annexe 2.2.2.1).

		Durée des phonèmes	
		zp	dp
Nombre de phonèmes dans la syllabe	R ²	0,000	0,000
	pente	0,018	-0,622
	p	1,95e-12	5,45e-05
Nombre de phonèmes dans le pied	R ²	0,01	0,02
	pente	-0,036	-3,137
	p	<2,2e-16	<2,2e-16
Nombre de phonèmes dans la NRU	R ²	0,014	0,034
	pente	-0,059	-5,772
	p	<2,2e-16	<2,2e-16
Nombre de phonèmes dans l'ANA	R ²	0,007	0,009
	pente	-0,044	-2,752
	p	<2,2e-16	<2,2e-16
Nombre de phonèmes dans l'ANA initiale	R ²	0,002	0,005
	pente	-0,018	-1,641
	p	1,20e-07	<2,2e-16
Nombre de phonèmes dans le mot	R ²	0,003	0,005
	pente	-0,02	-1,621
	p	<2,2e-16	<2,2e-16
Nombre de phonèmes dans l'UI	R ²	0,007	0,008
	pente	-0,008	-0,485
	p	<2,2e-16	<2,2e-16

Tableau 13. Résultats des analyses de régression linéaire pour l'hypothèse d'isochronie faible : durée des phonèmes en fonction de la complexité des constituants en terme de phonèmes (totalité des données)

Toutes les analyses sont très significatives, au moins à $p < 5.445e-05$, contrairement aux résultats obtenus sur le passage C01 dans lesquels l'effet de la complexité de la syllabe n'avait pas un effet significatif sur la durée normalisée des phonèmes. De plus, les valeurs p sont nettement plus faibles dans la totalité des données, et cette très haute significativité des résultats relève certainement de la quantité importante de données.

Nous sommes ici confrontée au même problème que dans les résultats pour le passage C01, à savoir des R^2 très faibles, signifiant que le facteur de la complexité ne permet pas d'expliquer beaucoup de variabilité des durées phonémiques. Ce phénomène est très probablement dû à la variabilité de la durée des phonèmes et au nombre important de facteurs influençant leur durée, mais il conviendrait d'effectuer les mêmes tests sur une portion alignée manuellement.

Toutes les corrélations sauf une (durée phonémique normalisée en fonction du nombre de phonèmes dans la syllabe) sont négatives, ce qui implique un certain effet de compression aux niveaux de la syllabe (uniquement pour la durée brute), du pied, de la NRU, de l'anacrouse, de l'anacrouse initiale, du mot et de l'unité intonative.

Les résultats concernant la syllabe sont partiellement surprenants, notamment dans le cas de la durée normalisée. Le R^2 est de 0,000 ce qui signifie que la durée des phonèmes n'est pas influencée par la complexité de la syllabe, résultat identique à celui obtenu pour le passage C01. En revanche, ce qui est surprenant, c'est la pente positive pour les durées normalisées. Il est en effet peu probable que la durée des phonèmes augmente lorsque la syllabe se complexifie. Etant donné que cet effet était non significatif pour C01, on suppose que la valeur p atteint un niveau de significativité du fait de la quantité importante de données. De plus, la pente est très faible et elle peut quasiment être assimilée à une pente nulle, ce qui correspond nettement mieux aux résultats, à savoir que la durée de la syllabe a tendance à croître linéairement lorsque le nombre de phonèmes augmente.

Le R^2 ainsi que le coefficient de pente de régression sont plus importants pour la NRU (respectivement 0.014/-0.059 pour z_p et 0.034/-5.772 pour d_p) que pour le pied (0.010/-0.036 pour z_p et 0.020/-3.137 pour d_p) indiquant une compression légèrement plus importante des phonèmes dans la NRU. On retrouve par conséquent la tendance observée dans le passage C01 bien que très peu marquée. On remarque ici aussi une certaine compression des phonèmes au niveau de l'anacrouse (-0.044 pour z_p et -2.752

pour dp) alors que le modèle de Jassem n'en prédit aucune, cette compression étant légèrement plus importante que celle qui s'opère dans les anacrouses initiales (-0.018 pour zp et -1.641 pour dp).

Les pentes concernant la complexité du mot et de l'unité intonative sont très faibles, il y a donc très peu, voire aucune, compression des phonèmes dans ces unités. Ce résultat rejoint celui obtenu pour C01.

Les résultats des analyses de régression concernant l'effet de la complexité en terme de syllabes sur la durée brute et normalisée des phonèmes (Tableau 14, Annexe 2.2.2.2) sont également similaires à ceux obtenus dans C01.

		Durée des phonèmes	
		zp	dp
Nombre de syllabes dans le pied	R ²	0,011	0,019
	pente	-0,091	-7,219
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans la NRU	R ²	0,012	0,024
	pente	-0,134	-11,741
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans l'ANA	R ²	0,005	0,005
	pente	-0,093	-5,05
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans l'ANA initiale	R ²	0,002	0,003
	pente	-0,048	-3,14
	p	5,42e-09	3,74e-12
Nombre de syllabes dans le mot	R ²	0,003	0,004
	pente	-0,048	-3,372
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans l'UI	R ²	0,007	0,007
	pente	-0,019	-1,143
	p	<2,2e-16	<2,2e-16

Tableau 14. Résultats des analyses de régression linéaire avec comme variable dépendante la durée des phonèmes et comme variable indépendante la complexité des constituants supérieurs en terme de syllabes (totalité des données)

Toutefois, contrairement à C01, l'effet dans l'anacrouse initiale est très significatif, on suppose que ceci est dû à la quantité importante de données.

Toutes les pentes sont négatives, confirmant la tendance vers un certain degré de compression des phonèmes lorsque la complexité des constituants en terme de syllabes augmente. Les pentes sont environ deux fois plus importantes que les pentes obtenues en fonction de la complexité en terme de phonèmes, ceci étant inférieur au nombre moyen de phonèmes par syllabe respectif pour chacun des constituants. Cette différence confirme que le phénomène de compression est globalement faiblement marqué.

Les résultats sur l'effet de la complexité de l'unité intonative en terme de nombre de pieds sur la durée des phonèmes (Tableau 15, Annexe 2.2.2.3) suivent la même tendance que les autres résultats. Cet effet est très significatif, les pentes sont négatives, ce qui suggère une compression des phonèmes mais les R^2 restent très faibles, ce qui signifie que ce facteur détermine très peu de variabilité de la durée des phonèmes.

		Durée des phonèmes	
		zp	dp
Nombre de pieds dans l'UI	R^2	0,004	0,003
	pente	-0,033	-1,665
	p	<2,2e-16	<2,2e-16

Tableau 15. Résultats des analyses de régression avec comme variable dépendante la durée des phonèmes et comme variable indépendante la complexité de l'unité intonative en terme de nombre de pieds (totalité des données)

En ce qui concerne l'effet de la complexité des constituants en terme de syllabes sur la durée des syllabes exprimée en durée brute et en durée logarithmique (Tableau 16, Annexe 2.2.2.4), les résultats sont similaires à ceux obtenus pour C01 sauf pour les deux types d'anacrouse. En effet, l'effet de la complexité en terme de syllabes est ici très significatif dans l'anacrouse ($p=5.23e-16$ pour les durées brutes et $p<2.2e-16$ pour les durées logarithmiques) et l'anacrouse initiale ($p=0.001$ pour les durées brutes et $p=0.003$ pour les durées logarithmiques). La compression des syllabes dans ces deux unités est très faible (comme dans C01), ce qui rejoint l'hypothèse de Jassem.

		Durée des syllabes	
		ds	log(ds)
Nombre de syllabes dans le pied	R ²	0,077	0,067
	pente	-29,222	-0,135
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans la NRU	R ²	0,165	0,171
	pente	-59,524	-0,273
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans l'ANA	R ²	0,005	0,006
	pente	-7,493	-0,054
	p	5,23e-16	<2,2e-16
Nombre de syllabes dans l'ANA initiale	R ²	0,001	0,001
	pente	-4,424	-0,022
	p	0,001	0,003
Nombre de syllabes dans le mot	R ²	0,016	0,01
	pente	-13,436	-0,052
	p	<2,2e-16	<2,2e-16
Nombre de syllabes dans l'UI	R ²	0,019	0,015
	pente	-3,682	-0,016
	p	<2,2e-16	<2,2e-16

Tableau 16. Résultats de l'analyse de régression avec comme variable dépendante la durée des syllabes (brute et exprimée en log) et comme variable indépendante la complexité des constituants supérieurs en terme de nombre de syllabes qu'ils contiennent (totalité des données)

Les autres effets sont tous très significatifs ($p < 2.2e-16$). Comme dans C01, le R^2 est nettement plus élevé pour la NRU ($R^2=0.165$ pour les durées brutes et $R^2=0.171$ pour les durées logarithmiques) que pour le pied ($R^2=0.077$ pour les durées brutes et $R^2=0.067$ pour les durées logarithmiques). La pente est négative pour les deux unités mais celle de la NRU est deux fois plus importante que celle du pied. Les résultats confirment donc la compression plus importante de la durée des syllabes dans la NRU en comparaison avec le pied.

En ce qui concerne la durée des syllabes en fonction de la complexité des mots et des unités intonatives, les résultats sont comparables à ceux obtenus dans C01, à savoir

quasiment aucune compression dans l'unité intonative (pentes de -3.682 pour les durées brutes et de -0.016 pour les durées logarithmiques pour chaque syllabe ajoutée). On trouve un peu plus de compression des syllabes dans le mot, mais celle-ci reste inférieure à celle que l'on trouve par exemple dans le pied.

6.2.3. Effet du paramètre accentuel

6.2.3.1. Passage C01

Pour tester l'effet du paramètre accentuel, nous utilisons des analyses de (co)variance. Celles-ci révèlent dans le passage C01 que l'effet du paramètre accentuel est très significatif sur la durée brute ($p < 2.2e-16$, $F(1,17126)=336.30$) et normalisée ($p=1.525e-12$, $F(1,17126)=50.091$) des phonèmes, ainsi que sur la durée brute ($p < 2.2e-16$, $F(1,6785)=630.87$) et logarithmique ($p < 2.2e-16$, $F(1,6785)=621.72$) des syllabes (cf. Annexe 2.3.1.1). Ces résultats confirment l'influence très marquée de l'accent sur la durée de ces constituants (Klatt 1987), ils semblent donc radicalement contredire l'hypothèse de Jassem et d'Abercrombie selon laquelle les syllabes se divisent équitablement la durée de la NRU ou du pied quel que soit leur statut accentuel.

Pour tester plus en détail l'effet du paramètre accentuel sur les durées phonémiques dans le cadre des deux modèles testés, nous avons utilisé une analyse de covariance en couplant le facteur accentuel au facteur de complexité de la NRU et du pied accentuel, mais également de la syllabe, du mot et de l'unité intonative. Les résultats détaillés concernant la durée normalisée et brute des phonèmes sont présentés en Annexe 2.3.1.2.

Dans C01, l'effet de l'interaction de la complexité de la syllabe et de l'accent a un effet non significatif sur la durée brute des phonèmes ($p=0.0898$) et un effet significatif sur la durée normalisée ($p=0.00816$). Ce résultat, non significatif dans un cas et peu significatif dans l'autre²⁷, n'est pas surprenant puisqu'on sait, de par les résultats obtenus dans les analyses de régression, que la durée des phonèmes n'est pas influencée par la complexité de la syllabe.

Ces analyses de covariance sont ensuite effectuées dans les autres constituants, à savoir le pied accentuel, la NRU, le mot et l'unité intonative. En ce qui concerne la NRU, les effets individuels et leurs interactions sont très significatifs pour les durées brutes

²⁷ en comparaison avec les autres niveaux de significativité de l'ordre de 0.0000000001.

($p < 2.2e-16$). En revanche, l'effet de l'accent est non significatif sur la durée normalisée des phonèmes ($p = 0.24909$), l'interaction des deux facteurs étant significative ($p = 0.00164$).

L'effet non significatif de l'accent dans la NRU est inattendu ; si on observe la Figure 28 ci-après, représentant la durée normalisée des phonèmes en fonction de la complexité de la NRU et du paramètre accentuel, on s'aperçoit que la courbe représentant la durée des phonèmes accentués n'est pas systématiquement au-dessus de la courbe représentant les phonèmes inaccentués. En effet, les courbes se croisent et celle des phonèmes inaccentués a même tendance à être plus souvent au-dessus de celles des accentués. On remarque que la distinction n'est plus claire au-delà de trois phonèmes, à savoir environ une syllabe puisque le nombre moyen de phonèmes par syllabe dans la NRU est de 2.94. Si on effectue la même analyse de covariance (sur les durées normalisées) dans les NRU de plus de trois phonèmes, le paramètre accentuel n'est pas significatif ($p = 0.0563$) et il en est de même pour l'interaction entre l'effet combiné de l'accent et la complexité de la NRU ($p = 0.9339$).

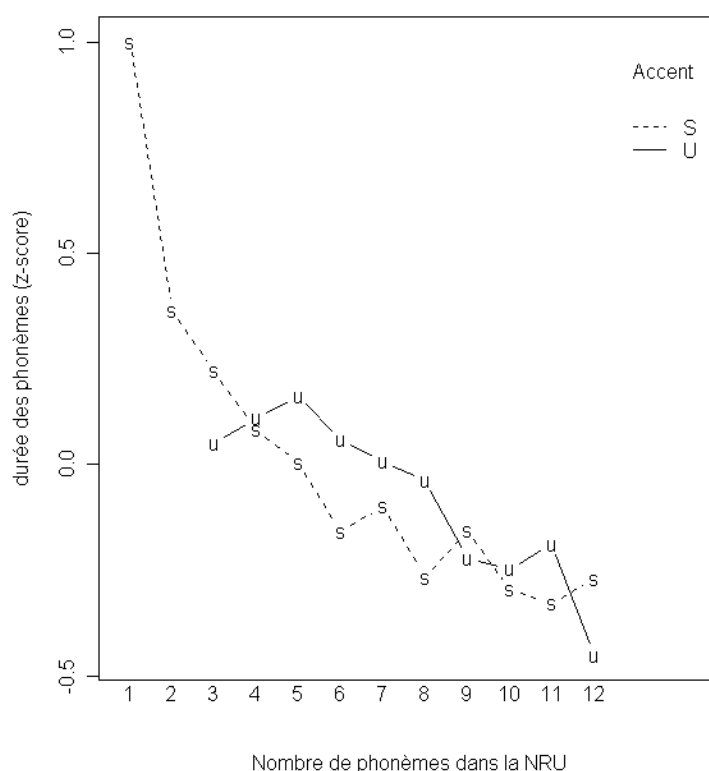


Figure 28. Durée normalisée des phonèmes en fonction de la complexité de la NRU et du statut accentuel (passage C01)

Il est quand même surprenant que les phonèmes inaccentués semblent être la plupart du temps plus longs que les phonèmes accentués (Figure 28). Ceci peut être dû à l'allongement final des phonèmes inaccentués en fin d'unité intonative. Nous avons par conséquent refait les mêmes analyses de covariance mais en excluant les deux derniers phonèmes des NRU en position finale d'unité intonative, en supposant que ces deux derniers phonèmes portent l'effet de l'allongement final. L'effet de l'accent n'est alors plus significatif sur les durées normalisées ($p=0.396$) et il en est de même pour l'interaction des deux facteurs ($p=0.568$). Les durées normalisées moyennes ainsi obtenues sont illustrées dans la Figure 29 ci-après. Par conséquent, la différence de durée entre les phonèmes accentués et les phonèmes inaccentués n'est pas significative dans les NRU dont on a supprimé les deux phonèmes en fin d'unité intonative supposés porter l'allongement final. Ce résultat est très surprenant mais il confirme les observations de Jassem selon lesquelles les phonèmes accentués ne sont pas plus longs que les phonèmes inaccentués à l'intérieur de la NRU.

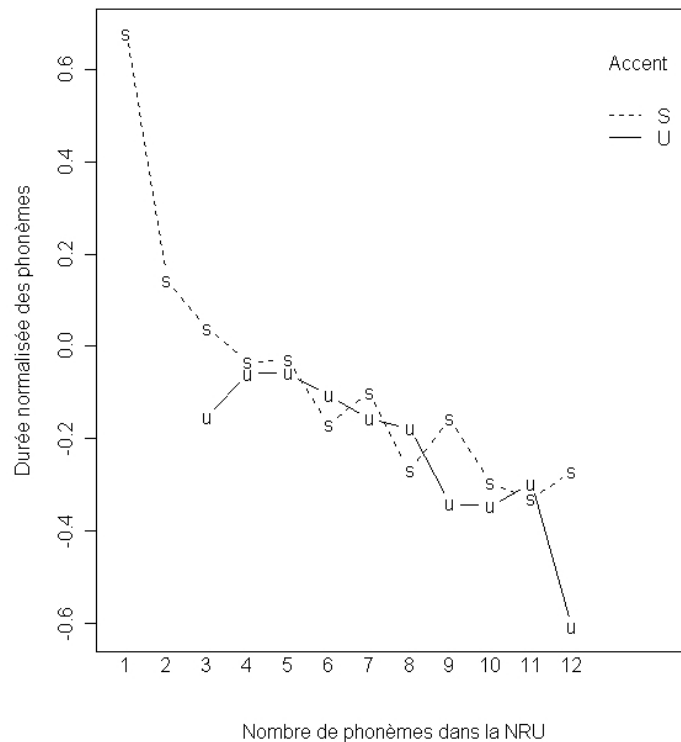


Figure 29. Durée normalisée moyenne des phonèmes dans la NRU (dont on a enlevé les deux phonèmes en fin d'unité intonative) en fonction de la complexité en phonèmes et de l'accent

Si on s'intéresse maintenant à la durée moyenne des phonèmes dans le pied en fonction de sa complexité et du paramètre accentuel (Figure 30 ci-après), les effets individuels de l'accent ($p=1.25e-14$ pour les durées normalisées et $p<2e-16$ pour les durées brutes) et leurs interactions avec la complexité ($p=2.81e-11$ pour les durées normalisées et $p<2e-16$ pour les durées brutes) sont tous très significatifs dans le pied accentuel. En observant la Figure 30, on s'aperçoit qu'à partir de cinq phonèmes (environ deux syllabes puisque le nombre moyen de phonèmes dans la syllabe à l'intérieur du pied est d'environ 2,8), la courbe des phonèmes accentués n'est pas systématiquement au-dessus de celle des phonèmes inaccentués. Si on effectue l'analyse de covariance sur les phonèmes dans les pieds de plus de quatre phonèmes, l'effet de l'accent n'est plus significatif ($p=0.123526$), de même que l'interaction n'est plus significative ($p=0.251947$).

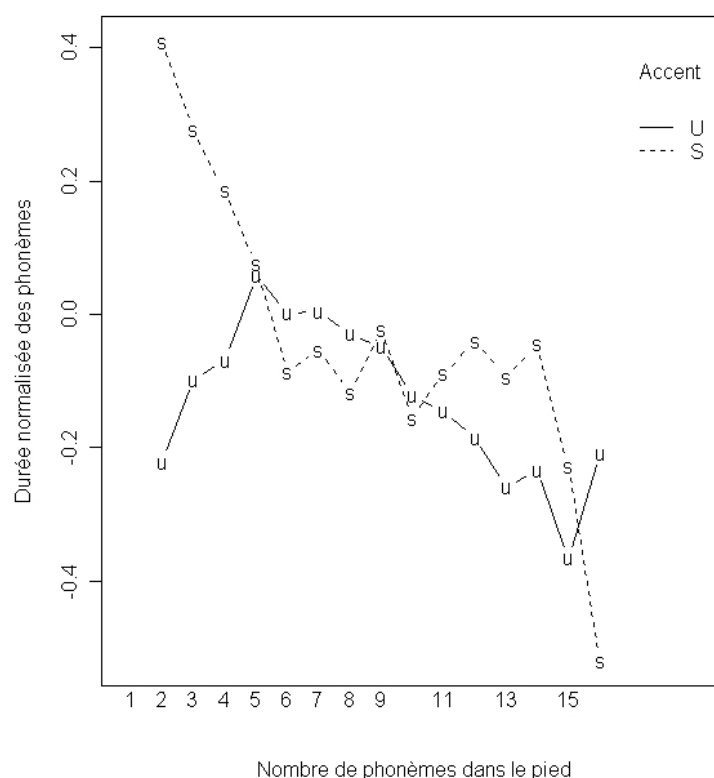


Figure 30. Représentation des durées normalisées des phonèmes en fonction du nombre de phonèmes dans le pied et du statut accentuel (C01)

Les différentes analyses de variance et covariance révèlent que l'effet du paramètre accentuel sur la durée des phonèmes est globalement très significatif. En revanche, cet effet est beaucoup moins marqué à l'intérieur du pied et il n'est plus significatif dans la

NRU. De plus, si on supprime les deux derniers phonèmes de la NRU en position finale d'unité intonative, l'effet simple et combiné de l'accent n'est plus significatif. Il semblerait donc que ce soit l'appartenance à une NRU, plutôt que l'accent, qui influence la durée des phonèmes (Jassem). En ce qui concerne les pieds, l'effet de l'accent est neutralisé dans les pieds de plus de quatre phonèmes.

Les mêmes tests sont ensuite effectués sur la durée des syllabes (cf. Annexe 2.3.1.3). Les effets simples du paramètre accentuel et de la complexité de la syllabe et leur interaction sont significatifs pour la durée brute et logarithmique des syllabes. Il est intéressant de remarquer que ces deux paramètres permettent de rendre compte d'environ 46% ($R^2=0.4619$) de la variabilité de la durée des syllabes. Si on observe la durée moyenne des syllabes en fonction de l'accent et de la complexité de la syllabe (Figure 31), on s'aperçoit que la durée de la syllabe augmente quasi linéairement en fonction de sa complexité, quel que soit le statut accentuel des phonèmes. Ces résultats rejoignent ceux obtenus pour l'hypothèse d'isochronie faible au niveau de la syllabe puisqu'il y a très peu, voire aucune compression des phonèmes au niveau de la syllabe. Il est également intéressant de voir que ce phénomène est présent autant pour les syllabes accentuées que les syllabes inaccentuées, de même que la différence entre les deux semble être relativement constante.

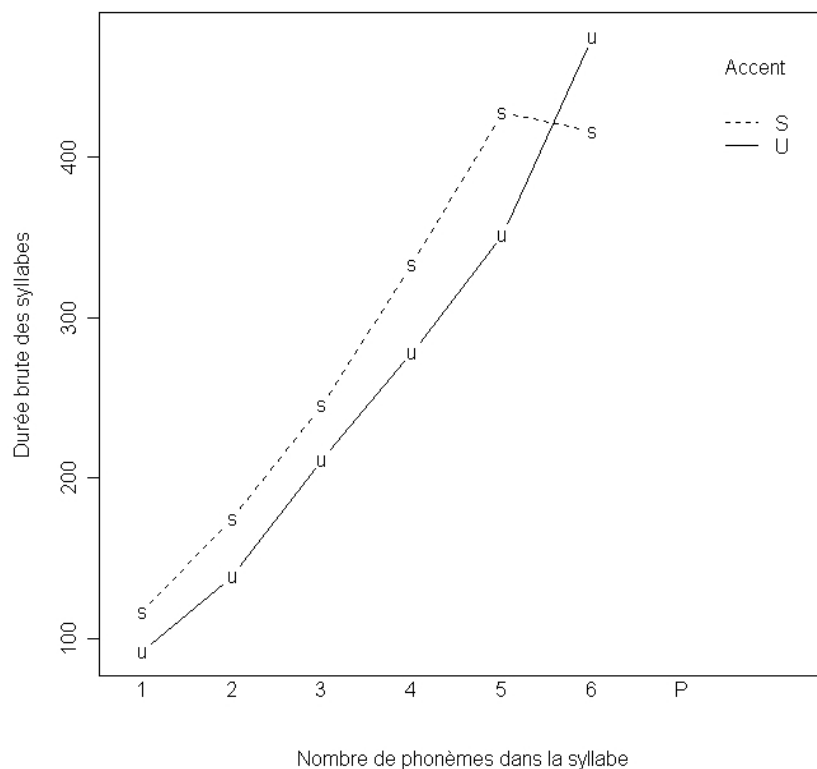


Figure 31. Durée brute moyenne des syllabes en fonction du statut accentuel et de la complexité de la syllabe

Le paramètre accentuel et le nombre de syllabes dans le pied ont également un effet très significatif sur la durée brute ($p < 2.2e-16$) et logarithmique ($p < 2.2e-16$) des syllabes. On observe la même tendance dans la NRU puisque les effets simples et combinés sont très significatifs à au moins $p = 5.26e-16$ (durée brute et logarithmique). La Figure 32 et la Figure 33 ci-après représentent respectivement les durées syllabiques brutes dans la NRU et dans le pied en fonction des deux paramètres ici testés dans le passage C01.

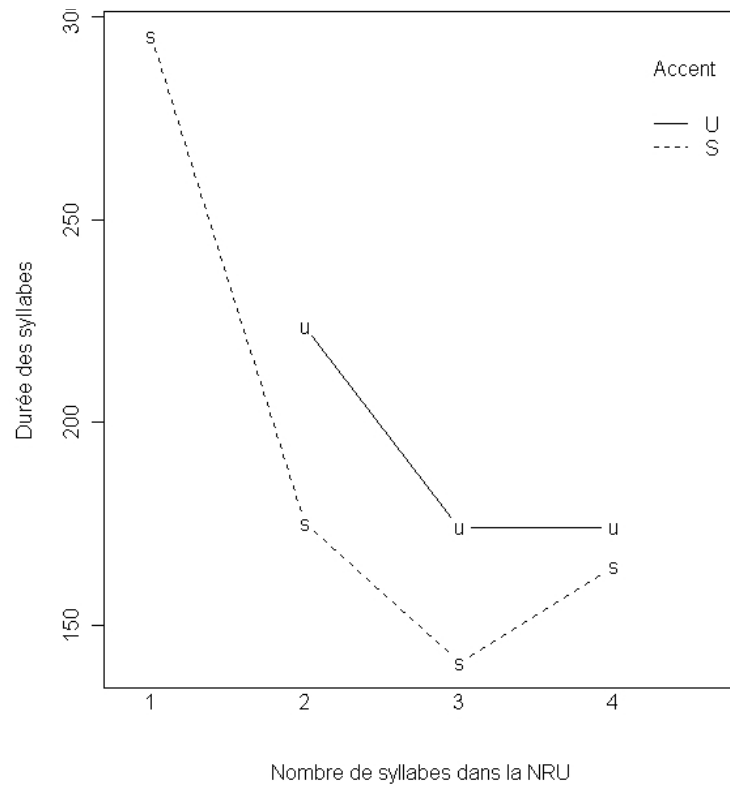


Figure 32. Représentation des durées syllabiques brutes en fonction du paramètre accentuel et du nombre de syllabes dans la NRU (passage C01)

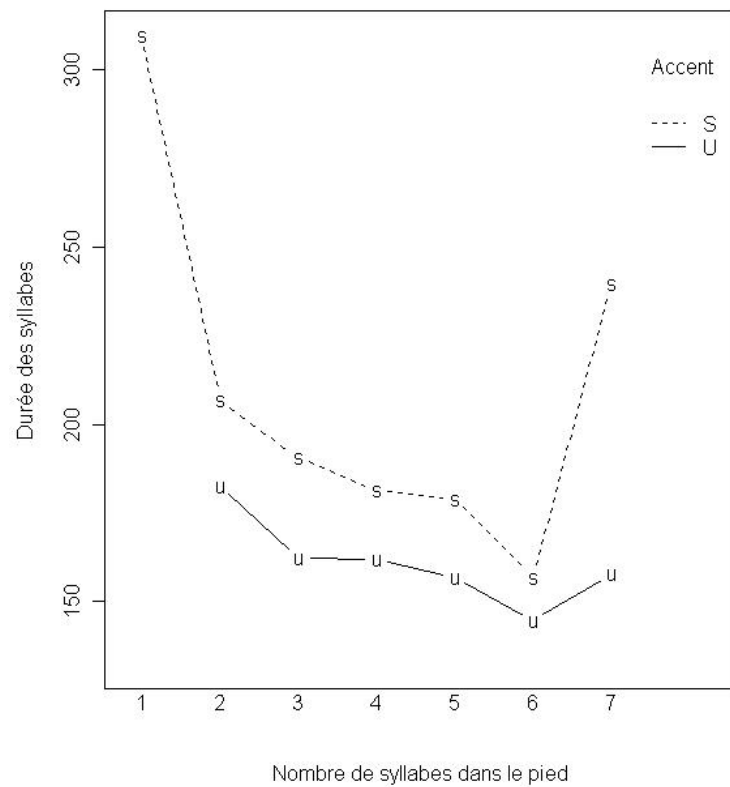


Figure 33. Représentation des durées syllabiques brutes en fonction du paramètre accentuel et du nombre de syllabes dans le pied accentuel (passage C01)

Il est intéressant, et surprenant, de voir que les syllabes inaccentuées dans la NRU sont relativement plus longues que les syllabes accentuées. Si on effectue des analyses ANOVA pour tester l'effet du paramètre accentuel dans les NRU de deux, trois et quatre syllabes, on s'aperçoit que l'effet est très significatif dans les NRU de deux ($p < 2.2 \times 10^{-16}$, $F(1,2186) = 197.74$) et trois syllabes ($p = 4.675 \times 10^{-13}$, $F(1,1101) = 53.626$) et dans les deux cas, la durée moyenne des syllabes inaccentuées est supérieure à celle des syllabes accentuées. Il est possible que cette différence soit due à un allongement des syllabes en position finale d'unité intonative. Il conviendrait de refaire les analyses de variance et de covariance en supprimant la dernière syllabe de la NRU en position finale d'unité intonative. En revanche, l'effet n'est pas significatif dans les NRU de quatre syllabes ($p = 0.3481$, $F(1,196) = 0.8847$), mais il y a nettement moins de cas dans les données.

Contrairement à la NRU, il ressort de la Figure 33 que la durée des syllabes accentuées dans le pied est supérieure à celles des syllabes inaccentuées. Il est intéressant de remarquer que pour les deux graphiques, les deux courbes représentant la durée des syllabes accentuées et des syllabes inaccentuées suivent approximativement les mêmes degrés de diminution et d'augmentation (à l'exception des valeurs les plus élevées, mais celles-ci représentent généralement peu de données). Par conséquent, il semblerait que la compression des syllabes dans la NRU et dans le pied se fait de façon similaire pour toutes les syllabes, quel que soit leur statut accentuel.

Dans le mot et dans l'unité intonative, les effets simples et combinés sur la durée (brute et logarithmique) des syllabes de l'accent et de la complexité en terme de nombre de syllabes sont tous très significatifs (au moins à $p = 7.91 \times 10^{-12}$).

Les différentes analyses concernant l'effet de l'accent révèlent que celui-ci est beaucoup moins marqué (voire pas du tout) si on le considère dans le cadre de la NRU et du pied accentuel. Plus précisément, son effet simple et combiné avec la complexité n'est plus significatif dans la NRU si on exclut de l'analyse les deux derniers phonèmes de la NRU en position finale d'unité intonative car ils sont supposés être allongés de par leur position finale. Après suppression de ces deux phonèmes, il n'y a donc plus de différence entre la durée des phonèmes accentués et des phonèmes inaccentués. En revanche, l'effet de l'accent dans le pied est très significatif, sauf dans les pieds de plus de quatre phonèmes mais ceux-ci ne représentent qu'une minorité.

On suppose que la différence d'effet de l'accent dans les deux unités rythmiques est due à la distinction entre la NRU et l'anacrouse (et donc entre les syllabes inaccentuées post- et pré-accentuelles), les deux unités n'étant pas distinguées dans le pied.

6.2.3.2. Totalité des données

Globalement, sur la totalité des données, l'effet du paramètre accentuel est très significatif sur la durée brute ($p < 2.2e-16$, $F(1,168569)=4088$) et normalisée ($p < 2.2e-16$, $F(1,168569)=946.25$) des phonèmes et sur la durée brute ($p < 2.2e-16$, $F(1,66538)=9148.4$) et logarithmique ($p < 2.2e-16$, $F(1, 66538)=10537$) des syllabes (cf. Annexe 2.3.2.1). Ces résultats sont identiques à ceux obtenus dans le passage C01, mais, tout comme nous l'avons fait pour le passage C01, il convient de tester l'effet du paramètre accentuel dans chacune des unités rythmiques et en le couplant avec la complexité grâce à une analyse de covariance.

L'interaction entre la complexité de la syllabe et le paramètre accentuel a un effet très significatif sur la durée brute ($p < 2.2e-16$) et normalisée ($p < 2.44e-14$) des phonèmes (cf. Annexe 2.3.2.2 pour les résultats concernant la durée des phonèmes). Ces résultats diffèrent quelque peu de ceux obtenus dans C01, notamment l'effet combiné sur la durée brute des phonèmes qui n'était alors pas significatif, et nous supposons que cette significativité est une conséquence directe de la quantité de données prises en compte dans l'analyse de covariance.

En ce qui concerne les autres constituants, les résultats sont similaires. Dans le pied, les effets simples et combinés sont tous très significatifs, qu'il s'agisse de la durée brute ou normalisée des phonèmes. De manière similaire, les effets simples et combinés sont très significatifs dans la NRU, ce qui diffère légèrement des résultats obtenus dans C01 puisque l'effet de l'accent avait un effet non significatif sur la durée normalisée des phonèmes (cf. Annexe 2.3.2.2).

La Figure 34 illustre les durées normalisées des phonèmes en fonction de la complexité de la NRU et du paramètre accentuel.

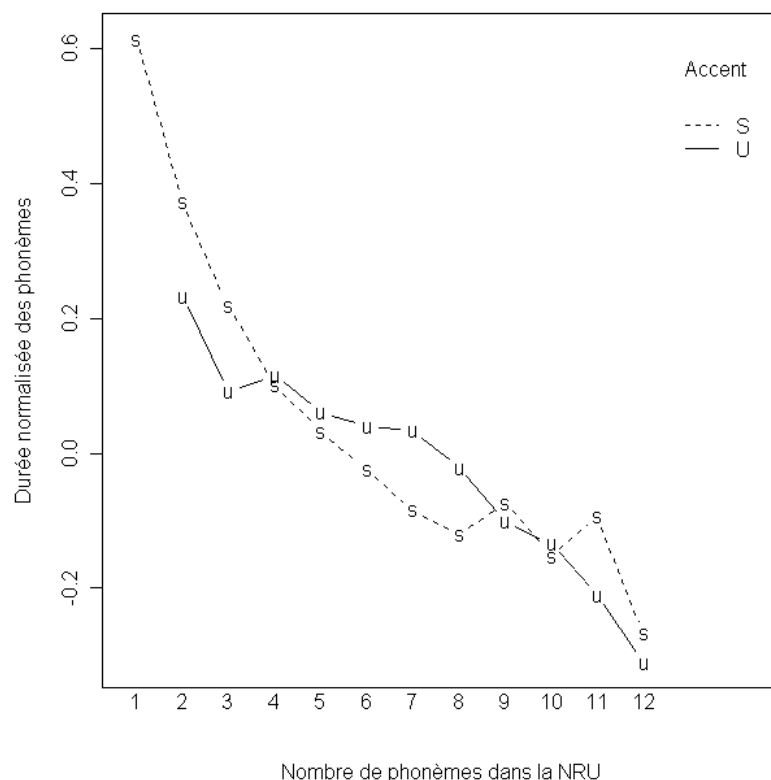


Figure 34. *Durée normalisée des phonèmes en fonction de la complexité de la NRU et du paramètre accentuel (totalité des données)*

Dans le passage C01, nous avons testé les mêmes effets dans les NRU de plus de trois phonèmes, l'effet simple de l'accent ainsi que l'effet combiné des deux paramètres étaient non significatifs. Nous avons donc conclu que l'effet de l'accent n'était plus significatif dans les NRU de plus de trois phonèmes (en moyenne une syllabe). Ce résultat se retrouve partiellement dans la totalité des données puisque l'effet de l'accent dans les NRU de plus de trois phonèmes n'est plus significatif sur les durées normalisées ($p=0.1539$) ; l'effet combiné reste, lui, très significatif ($p=0.0005$). En revanche, dans les NRU de plus de quatre phonèmes, l'effet simple de l'accent ($p=0.387$) et l'effet combiné ($p=0.880$) ne sont plus significatifs. Cette tendance se retrouve dans la Figure 34.

L'analyse de covariance sur les mêmes données mais en excluant les deux derniers phonèmes de la NRU en position finale d'unité intonative révèle ici des effets simples et combinés très significatifs sur les durées normalisées (respectivement $p=3.37e-12$ et $p=7.90e-05$). On ne retrouve pas ici le même effet que dans C01 et on suppose que la divergence est due au calcul différent des durées normalisées dans les deux cas. Les résultats obtenus dans C01 sont donc plus fiables puisque les durées normalisées ont été

recalculées spécifiquement pour ce locuteur. Pour s'assurer que cet effet non significatif de l'accent dans la NRU ne soit pas un artefact dû au locuteur du passage C01, nous avons fait la même analyse de covariance sur le locuteur (unique) du passage A01 en recalculant également les durées normalisées. Les résultats sont similaires à ceux obtenus pour C01 : les effets simple et combiné ne sont pas significatifs (respectivement $p=0.335$ et $p=0.525$).

La Figure 35 ci-après illustre les durées normalisées moyennes des phonèmes dans la NRU en fonction de l'accent et de la complexité en ayant au préalable supprimé les deux derniers phonèmes des NRU en position finale d'unité intonative.

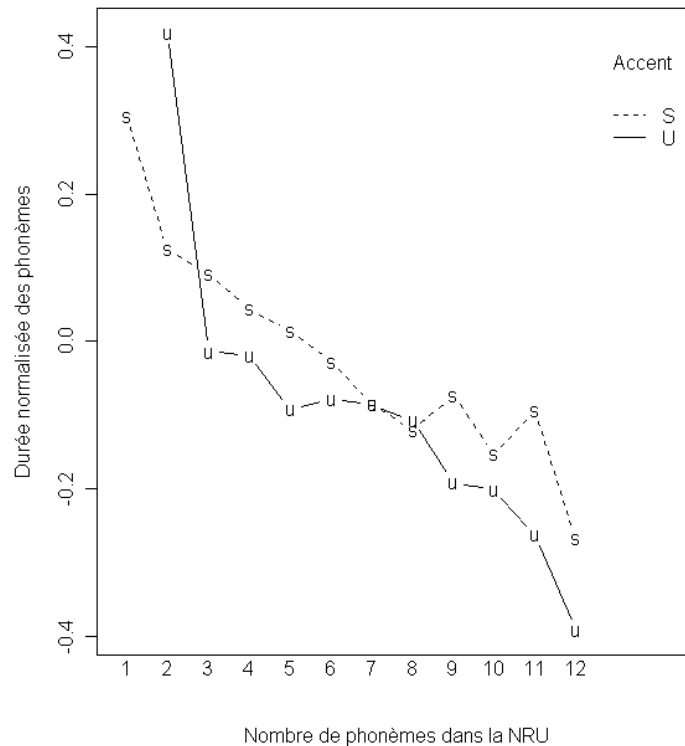


Figure 35. Durées normalisées moyennes des phonèmes dans la NRU en fonction de l'accent et de la complexité en ayant au préalable supprimé les deux derniers phonèmes des NRU en position finale d'unité intonative

En ce qui concerne les pieds accentuels (Figure 36), les analyses de covariance effectuées dans C01 révélaient que les effets simples et combinés n'étaient plus significatifs au-delà de trois phonèmes. Dans la totalité des données, l'interaction n'est plus significative au-delà de trois phonèmes ($p=0.11489$), mais l'effet (simple) de l'accent reste significatif ($p=0.00119$). C'est ici uniquement dans les pieds de plus de

sept phonèmes que les effets simples et combinés ne sont plus significatifs. En observant la Figure 36, il semble pourtant que la différence de durée entre les phonèmes accentués et inaccentués n'est plus marquée dans les pieds de plus de quatre phonèmes.

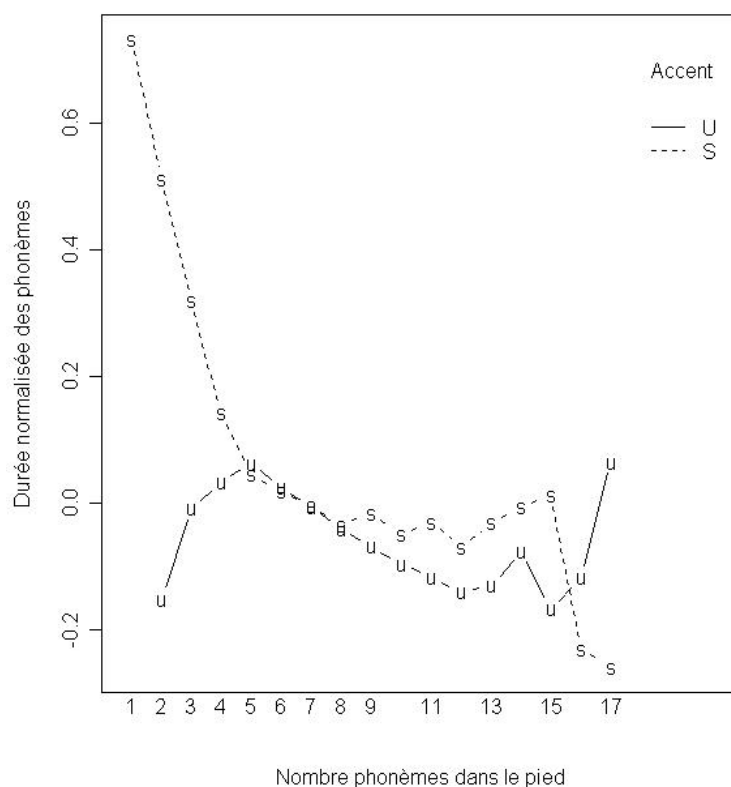


Figure 36. Représentation des durées normalisées des phonèmes en fonction de la complexité du pied et du paramètre accentuel (totalité des données)

Dans le mot et dans l'unité intonative, les effets simples et combinés ont tous un effet très significatif sur la durée normalisée et brute des phonèmes.

Nous avons ensuite testé les effets des mêmes facteurs sur la durée des syllabes grâce également à une série d'analyses de covariance (cf. Annexe 2.3.2.3). Les effets simples et combinés du paramètre accentuel et de la complexité de la syllabe ont un effet très significatif sur la durée brute et logarithmique de la syllabe. Si on observe la représentation de la durée syllabique en fonction de ces deux paramètres (Figure 37), la tendance est similaire à celle observée dans le passage C01 : la durée de la syllabe augmente quasi linéairement en fonction de sa complexité, ceci quel que soit le statut accentuel des phonèmes. On remarque toutefois une légère chute pour les phonèmes

inaccentués dans les syllabes de sept phonèmes, mais ils ne représentent que très peu de données.

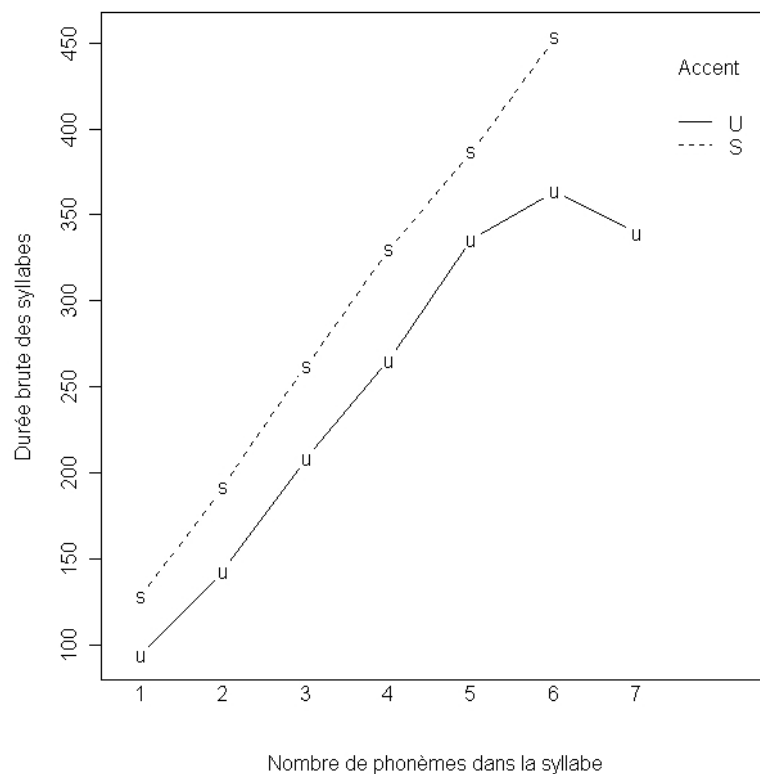


Figure 37. Représentation des durées syllabiques en fonction du paramètre accentuel et de la complexité des syllabes (totalité des données)

Les effets simples et combinés des deux paramètres (accent et complexité) ici testés ont ici aussi un effet significatif sur la durée brute et logarithmique des syllabes dans les NRU et dans les pieds. Les durées syllabiques brutes dans la NRU et dans le pied en fonction des deux paramètres sont illustrées respectivement dans la Figure 38 et la Figure 39 ci-après. On observe dans ces figures les mêmes tendances que pour le passage C01, à savoir que la durée des syllabes inaccentuées dans les NRU a tendance à être supérieure à celle des syllabes accentuées, alors que le phénomène inverse apparaît dans le pied accentuel. De plus, il semblerait que la compression des syllabes se fasse de manière identique quel que soit le statut accentuel, ceci étant plus marqué dans le pied.

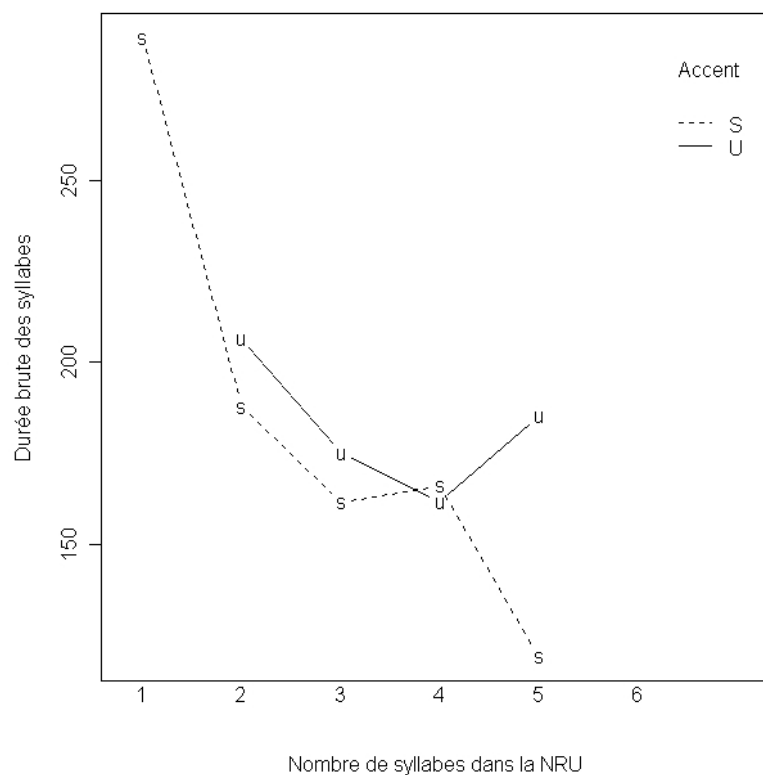


Figure 38. Représentation des durées syllabiques dans la NRU en fonction du paramètre accentuel et de la complexité de la NRU (totalité des données)

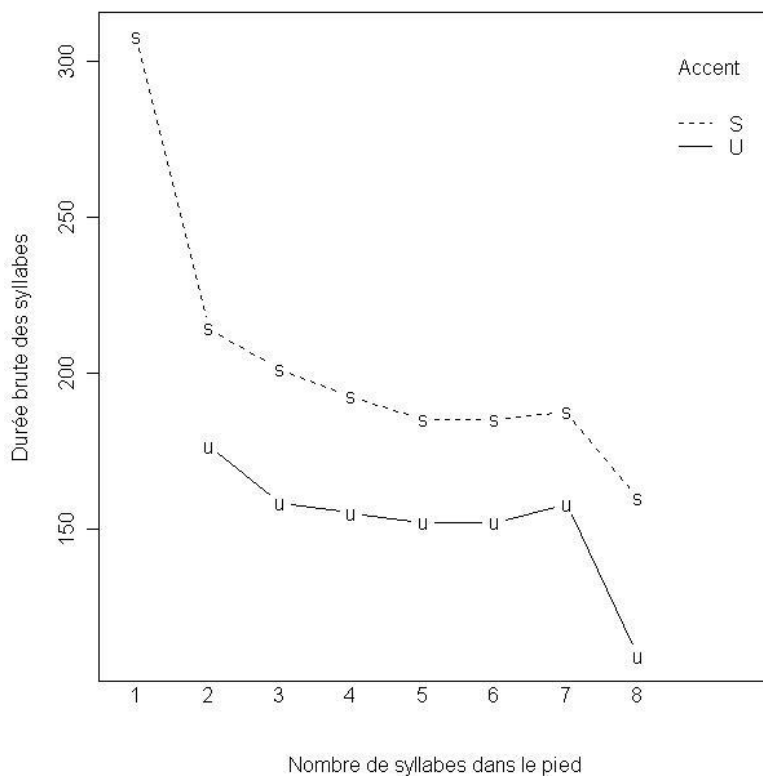


Figure 39. Représentation des durées syllabiques dans le pied accentuel en fonction du paramètre accentuel et de la complexité du pied (totalité des données)

6.3. Résumé et conclusion

Les résultats obtenus grâce aux analyses de régression et de (co)variance semblent valider la supériorité du modèle de Jassem en comparaison avec celui d'Abercrombie/ Halliday (Bouzon et Hirst 2004a ; Bouzon et Hirst, soumis).

Nous avons tout d'abord vu qu'il n'y avait pas en anglais britannique d'isochronie stricte aux niveaux de la syllabe, du pied, de la NRU, de l'anacrouse, de l'anacrouse initiale, du mot et de l'unité intonative. La durée de chacun de ces constituants augmente lorsque le nombre de sous-constituants croît.

Nous avons dans un deuxième temps observé en anglais britannique un certain degré de compression (plus ou moins marqué) des phonèmes et des syllabes aux niveaux de la NRU, du pied, des deux types d'anacrouse, du mot et de l'unité intonative, mais pas au niveau de la syllabe. La durée des syllabes a tendance à augmenter linéairement en fonction de leur complexité et les phonèmes qui les composent ont tendance à conserver la même durée quelle que soit la complexité de la syllabe. Ce résultat confirme les travaux de Hill, Jassem et Witten (1978a) mais contredisent ceux de Campbell (1992a). En ce qui concerne les unités rythmiques, les résultats sur l'isochronie faible montrent que le modèle de Jassem rend mieux compte de la structuration rythmique de l'anglais britannique. Il existe en effet plus de compression des phonèmes et des syllabes dans la NRU par rapport à l'anacrouse, ce qui semble confirmer la nécessité de distinguer ces deux unités au niveau rythmique, contrairement au découpage décrit dans le modèle d'Abercrombie. La compression dans le pied est proche de celle dans l'anacrouse, à savoir environ deux fois moins importante que celle de la NRU. La distinction entre NRU et anacrouse semble par conséquent être nécessaire.

Nous avons également testé l'effet du paramètre accentuel et nous avons observé que son effet est non significatif dans les NRU si on exclut les deux derniers phonèmes en position finale d'unité intonative. Ce résultat est remarquable et il confirme l'absence de différence de durée entre les phonèmes accentués et inaccentués dans les NRU comme Jassem l'avait décrit dans son modèle (si l'effet de l'allongement final est neutralisé). L'effet de l'accent est très significatif dans le pied, ce qui tend à montrer que l'effet de l'accent dans le pied est entièrement causé par la distinction entre l'anacrouse et la NRU.

Les résultats obtenus ici sont très intéressants puisqu'ils permettent de montrer que le modèle de Jassem permet une meilleure représentation de l'organisation rythmique de l'anglais britannique. De plus, les résultats concernant l'effet du paramètre accentuel sont particulièrement importants puisqu'ils montrent que c'est l'appartenance à la NRU plutôt que l'effet de l'accent qui influe sur la durée des phonèmes.

Les résultats que nous avons obtenus semblent par conséquent confirmer la supériorité du modèle de Jassem. Rappelons que ce découpage en ANA + NRU (= TRU) est propre à la structure rythmique de l'anglais et Jassem propose une unité différente pour rendre compte de la structure mélodique de l'anglais, il s'agit de l'unité tonale (section 3.2.2.1). Or, cette unité tonale regroupe un NRU avec l'anacrouse suivante, ce qui correspond parfaitement au pied accentuel. Il est par conséquent important de rappeler que pour Jassem l'organisation rythmique et l'organisation tonale ne se situent pas au sein de la même unité (Hirst 1998), contrairement à Abercrombie (entre autres) qui utilise une seule et unique unité pour rendre compte des deux types de phénomène. A partir des résultats obtenus concernant la structuration rythmique, on peut supposer que la distinction entre les deux structures, la TRU et l'unité tonale, est nécessaire pour une meilleure représentation des caractéristiques prosodiques de l'anglais.

CHAPITRE 7.
SYLLABIFICATION EN ANGLAIS BRITANNIQUE

7.1. HYPOTHESES ET OUTILS STATISTIQUES.....	262
7.1.1. Hypothèses	262
7.1.2. Outil statistique	264
7.1.3. Expérience en production.....	264
7.2. TEST SUR LA BASE DE DONNEES AIX-MARSEC	265
7.2.1. Résultats dans le passage C01	265
7.2.2. Résultats dans la totalité des données Aix-MARSEC	267
7.2.3. Conclusion.....	267
7.3. EXPERIENCE EN PRODUCTION.....	269
7.3.1. Elaboration des consignes	270
7.3.1.1. Phase d'entraînement (sujets francophones).....	271
7.3.1.2. Phase d'entraînement (sujet anglophone)	272
7.3.1.3. Ajout d'une deuxième partie	273
7.3.2. Elaboration de l'expérience.....	274
7.3.2.1. Stimuli	274
7.3.2.2. Aspects techniques	275
7.3.2.3. Sujets	276
7.3.3. Mise en forme des données	276
7.3.4. Interprétation des résultats	281
7.3.4.1. Répartition des différents types de découpages syllabiques	281
7.3.4.2. Occurrences de consonnes ambisyllabiques	285
7.3.4.3. Regroupement en position codaïque	289
7.3.4.4. Regroupement selon le principe d'attaque maximale	292
7.3.4.5. Comparaison des découpages des Phases 3 et 4	294
7.3.4.6. Bilan de l'expérience.....	296
7.4. RESUME ET CONCLUSION.....	298

Le but des modèles syllabiques est de prédire la durée des syllabes pour ensuite déterminer la durée des phonèmes qui les composent. Ce type d'approche pose la question des limites de l'unité syllabique ou plus globalement de la (re)syllabification en anglais. Il existe plusieurs façons de syllabifier un même mot polysyllabique : selon le principe d'attaque maximale, selon le principe de rime branchante ou selon le principe d'ambisyllabité (Chapitre 2). Nous avons cité un certain nombre d'études, un premier type de nature plus descriptive et un second décrivant des expériences menées dans un cadre psycholinguistique.

Notre approche est double dans le sens où nous abordons la problématique de l'ambisyllabité sous deux angles différents. Dans un premier temps, nous cherchons à déterminer à partir de la base de données Aix-MARSEC s'il existe une différence temporelle significative entre les consonnes potentiellement ambisyllabiques et les consonnes non ambisyllabiques (typiquement en position d'attaque) ; le même type de test est également réalisé sur la durée des voyelles précédant ces consonnes.

En complément de ces résultats, nous effectuons dans un deuxième temps une expérience de production pour explorer plus en détail les stratégies de syllabification de sujets anglophones natifs. Ce test nous permet également de tester l'influence des facteurs observés par Treiman et Danis (1988) comme influençant l'apparition des consonnes ambisyllabiques (cf. Chapitre 2).

Ce travail n'a pas pour prétention de déterminer de manière tranchée si l'ambisyllabité fait partie du système de l'anglais ou pas, mais plutôt d'explorer certaines pistes qui nous semblent pertinentes, avec les outils dont nous disposons. Il nous semble intéressant de tester cette hypothèse dans la base de données Aix-MARSEC, à savoir sur un corpus important de parole authentique, puis de compléter ces résultats avec une expérience en production (parole contrainte).

La première section est consacrée à la récapitulations des hypothèses concernant l'ambisyllabité et à la description de la méthodologie adoptée dans ce travail. Dans la deuxième section, nous présentons les résultats concernant les tests dans la base de données Aix-MARSEC. La troisième section porte sur la description de l'expérience en production ainsi que sur l'interprétation des résultats.

7.1. Hypothèses et outils statistiques

7.1.1. Hypothèses

Nous allons dans un premier temps tester s'il existe une différence statistiquement différente entre la durée des consonnes ambisyllabiques et des consonnes non ambisyllabiques. Une différence significative suppose un comportement distinct des deux types de consonnes, ceci allant dans le sens du principe d'ambisyllabité. En revanche, une absence de différence significative irait dans le sens de l'hypothèse selon laquelle les deux types de consonnes n'ont pas lieu d'être distingués.

Le même test est effectué sur la durée des voyelles précédentes dans le but de déterminer s'il existe une différence significative de durée entre les voyelles précédant les consonnes ambisyllabiques et celles précédant les consonnes non ambisyllabiques. On suppose ici que s'il existe une différence entre les deux types de consonnes, celle-ci va se refléter sur la durée de la voyelle précédente.

Les tests sont effectués dans la base de données orales authentiques Aix-MARSEC dans laquelle nous avons effectué un double codage des consonnes intervocaliques (section 5.6.2). Ce codage est illustré dans la Figure 40 ci-après.

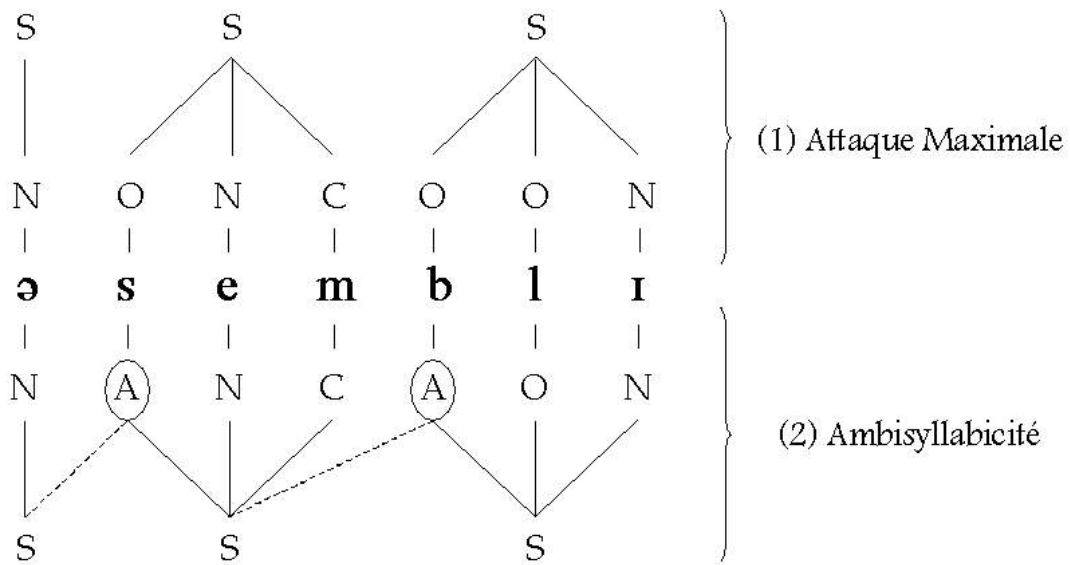


Figure 40. Double codage des consonnes intervocaliques dans la base de données Aix-MARSEC selon le principe d'attaque maximale et celui d'ambisyllabité

Le premier codage des consonnes intervocaliques se fait en fonction du principe d'attaque maximale (partie haute de la Figure 40) : un maximum de consonnes est placé en position d'attaque tout en respectant les contraintes phonotactiques de cooccurrence. Ainsi, dans le mot *assembly* donné en exemple, la consonne /s/ est en position d'attaque. Dans le groupe consonantique /mb/, seule la séquence /b/ est une attaque licite, elle est donc placée en position d'attaque et le /m/ en position de coda. Le deuxième codage se fait en fonction du principe d'ambisyllabité (partie basse de la figure) : quand une consonne peut être attaque et coda, elle est ambisyllabique. En cas de groupe consonantique, la première consonne de l'attaque licite est ambisyllabique.

Dans l'exemple *assembly*, la consonne /s/ peut être attaque et coda, elle est donc codée comme étant potentiellement ambisyllabique. Pour le groupe consonantique /mb/, la séquence /b/ est une attaque licite, le /b/ est donc codé comme étant potentiellement ambisyllabique. Notons que pour un tel codage, il faudrait également prendre en compte les contraintes portant sur la coda ; or, comme nous l'avons précisé dans le Chapitre 5, il est beaucoup plus complexe de prendre en compte à la fois la bonne formation de l'attaque et la bonne formation de la coda pour le codage automatique. Nous avons donc opté pour un codage uniquement en fonction de l'attaque et nous suivons ici Pulgram (1970) : toute irrégularité porte sur la coda plutôt que sur l'attaque (*Principle of irregular coda*).

A partir de ce double codage, nous pouvons tester les différences qui existent entre les deux catégories de consonnes en terme de durée : les consonnes en position d'attaque (selon le principe d'attaque maximale) et les consonnes codées comme étant potentiellement ambisyllabiques.

7.1.2. Outil statistique

Comme pour les tests concernant les modèles supra-syllabiques, nous ferons dans un premier temps toutes les analyses dans le passage C01 et dans un deuxième temps sur la totalité des données Aix-MARSEC. Nous comparerons ensuite les résultats obtenus afin de voir si les tendances sont similaires. Tous les tests sont réalisés grâce au logiciel de statistique R et la totalité des résultats est donnée en Annexes 3.1 (passage C01) et 3.2 (totalité des données). Toutes les analyses sont effectuées sur les durées normalisées de phonèmes car, comme nous l'avons vu dans le Chapitre 6, la normalisation des données permet de neutraliser les effets dus à l'identité des phonèmes. Nous ferons toutefois référence aux résultats obtenus sur les durées brutes lorsque cela sera nécessaire.

Le test statistique utilisé dans cette partie est le test-t ; celui-ci permet de comparer deux catégories deux à deux, à savoir la durée (brute ou normalisée) des consonnes potentiellement ambisyllabiques et la durée des consonnes en position d'attaque. Une différence non significative tendrait à supposer qu'il n'y a pas lieu de distinguer les consonnes ambisyllabiques des consonnes non ambisyllabiques. En revanche, si la différence s'avère être significative, il convient de creuser davantage l'hypothèse d'ambisyllabité. Le même test est utilisé pour la comparaison de la durée des voyelles précédentes.

7.1.3. Expérience en production

En complément des différents tests dans la base de données Aix-MARSEC, nous avons trouvé intéressant d'explorer cette problématique sous la forme d'une expérience en production. La question que l'on se pose est de savoir comment tester le phénomène d'ambisyllabité (en dehors des tests statistiques précédemment décrits). La solution n'est pas aisée puisqu'il est quasi impossible d'observer les stratégies de syllabification dans de la parole naturelle, du moins de manière objective. Il faut obtenir de sujets

natifs qu'ils découpent des énoncés en syllabes de manière distincte pour qu'on puisse observer les stratégies de syllabification de manière objective.

A cette fin, nous avons opté pour une expérience en production dont le but est d'inciter des sujets britanniques natifs à découper clairement des énoncés en syllabes et de comptabiliser et de comparer les différents types de regroupements syllabiques des consonnes intervocaliques à l'intérieur des mots. Cette méthode pose le problème du passage à la parole contrainte de laboratoire qui risque donc d'être peu naturelle, mais il est difficile d'imaginer un contexte de production naturelle où des sujets découpent leurs messages instinctivement en syllabes, et ce de manière suffisamment claire pour que nous puissions observer objectivement les stratégies de syllabification.

La première supposition faite dans le cadre de cette expérience est que, si l'ambisyllabité fait partie intégrante du système phonologique de l'anglais, ce double regroupement de la consonne intervocalique devrait se retrouver dans la syllabification des sujets de manière relativement systématique.

Si la production de consonnes ambisyllabiques est effectivement systématique, nous tenterons de déterminer quels facteurs influencent l'apparition de telles consonnes en fonction notamment des résultats de Treiman et Danis (1988). Les facteurs retenus sont les suivants :

- Catégorie de consonne intervocalique
- Complexité orthographique de la consonne intervocalique
- Patron accentuel
- Qualité de la voyelle précédente

Nous testerons par conséquent l'influence de ces paramètres dans le but de voir si nous retrouvons les résultats de Treiman et Danis.

7.2. Test sur la base de données Aix-MARSEC

7.2.1. Résultats dans le passage C01

Dans le passage C01, on compte 2 627 cas de phonèmes intervocaliques codés ambisyllabiques. Nous allons tester dans un premier temps la différence de durée entre les différentes catégories de consonne deux à deux ; pour cela, nous utilisons un test-t en comparant les attaques et les consonnes ambisyllabiques puis les attaques et les codas (cf. Annexe 3.1 pour les résultats détaillés).

Le test-t permettant de comparer les attaques et les consonnes ambisyllabiques révèle une différence très significative en terme de durée brute ($p=2.823e-11$) mais cette différence se révèle être non significative si on prend en compte la durée normalisée ($p=0.177$). Le résultat obtenu sur les durées normalisées est plus fiable puisque l'effet de l'identité des phonèmes est neutralisé. Il n'y a donc pas de différence significative entre les attaques et les consonnes intervocaliques codées ambisyllabiques. Ce résultat représente un argument contre l'hypothèse d'ambisyllabité, bien qu'il ne soit pas suffisant pour tirer quelque conclusion. Notons que la différence de durée (brute) entre les deux catégories de consonnes est d'environ 6.5ms, les consonnes ambisyllabiques étant légèrement plus longues que les attaques.

Si on compare les attaques et les codas, on s'aperçoit que la différence est très significative pour les durées brutes ($p<2.2e-16$) et normalisées ($p=0.0002$), avec toutefois une valeur p nettement plus élevée pour les durées normalisées. Une rapide comparaison des durées moyennes des deux types de consonnes révèle que les codas sont légèrement plus longues que les attaques ; la durée moyenne des codas est d'environ 83ms (0.057 en terme de durée normalisée) et celle des attaques de 68ms (-0.43 pour la durée normalisée). Notons que cette différence peut être due à un effet d'allongement final puisque les codas, à l'instar des attaques, peuvent se trouver en position finale d'unité intonative. Si on réalise la même analyse en excluant les codas en position finale d'unité intonative, la différence concernant les durées normalisées n'est plus significative ($p=0.314$), la durée moyenne des codas étant désormais de 73ms. On peut donc conclure qu'il n'y a pas de différence significative en terme de durée entre les attaques et les codas (non finales).

Pour affiner ces tests, nous observons également la durée (brute et normalisée) de la voyelle qui précède ces consonnes. Un test-t nous permet de comparer les deux types de voyelles, devant une consonne ambisyllabique ou une consonne non ambisyllabique, en excluant les voyelles en position finale de mot (celles-ci sont suivies d'une attaque et pas d'une coda). Il ressort que la différence entre les deux est très significative pour la durée normalisée ($p<2.2e-16$) et pour la durée brute ($p<2.2e-16$). Si on observe plus en détail les durées brutes moyennes respectives, on s'aperçoit que la différence entre chaque catégorie est de l'ordre de 15ms pour une durée moyenne globale des voyelles en question autour de 83ms. Les noyaux vocaliques précédant les consonnes ambisyllabiques sont plus courts (environ 74ms) que les autres noyaux (89ms), ceci

s'expliquant certainement par le fait qu'environ 70% des consonnes ambisyllabiques sont des obstruantes. Ce résultat semblerait donc montrer que l'effet d'une consonne ambisyllabique sur son environnement est différent de celui d'une consonne en attaque.

7.2.2. Résultats dans la totalité des données Aix-MARSEC

Les mêmes tests sont effectués sur la totalité des données du corpus, représentant environ 10 000 cas de consonnes ambisyllabiques (cf. Annexe 3.2 pour les résultats détaillés). Le test-t visant à tester la différence entre la durée des attaques et des consonnes ambisyllabiques révèle que celle-ci est très significative pour les durées brutes ($p=2.465e-12$) et normalisées ($p<2.2e-16$). Ce résultat diffère de celui obtenu sur le passage C01 puisque la différence n'était pas significative en terme de durée normalisée. En revanche, la différence en terme de durée brute entre les deux catégories est ici de 2.6ms, ce qui est nettement en dessous du seuil de perception. On peut par conséquent se demander si cette significativité de la valeur p n'est pas une conséquence directe de la quantité de données, d'autant plus que cette différence était non significative dans le passage C01 (avec une différence de durée brute de l'ordre de 6.5ms).

Un test-t est également effectué pour tester la différence entre les attaques et les codas. Cette analyse révèle une différence très significative en terme de durée brute ($p<2.2e-16$) et significative en terme de durée normalisée ($p=0.001$), cette dernière reste significative si on exclut les codas en position finale d'unité intonative (supposées porter un allongement final). On note toutefois que la valeur p pour les durées normalisées est relativement élevée et il se peut qu'elle atteigne ici aussi un niveau de significativité en raison du grand nombre de données.

La différence de durée entre les voyelles précédant les consonnes ambisyllabiques et les voyelles précédant les codas est très significative pour les durées brutes ($p<2.2e-16$) et normalisées ($p<2.2e-16$), ce résultat confirme par conséquent celui obtenu dans le passage C01.

7.2.3. Conclusion

Les divers résultats obtenus dans la base de données Aix-MARSEC sont intermédiaires. En effet, la différence de durée normalisée entre les attaques et les consonnes

ambisyllabiques n'est pas significative dans le passage C01 mais significative dans la totalité des données. Nous retiendrons le résultat obtenu dans C01 car il nous semble plus fiable en terme de quantité de données et de justesse des durées normalisées. De plus, dans la totalité des données, la différence de durée brute entre les deux catégories de consonne est de 2.6ms, ce qui se situe largement en dessous du seuil de perception. Cette conclusion semble donc plutôt représenter un argument contre l'hypothèse d'ambisyllabité. En revanche, la différence de durée des voyelles précédant les consonnes ambisyllabiques et celles précédant les codas est très significative dans le passage C01 et dans la totalité des données. Il semble donc que les consonnes ambisyllabiques influencent différemment la durée des voyelles précédentes en comparaison avec les consonnes en position de coda, ce qui représente un argument plutôt en faveur de l'hypothèse d'ambisyllabité. Le troisième résultat concerne la différence de durée entre les attaques et les codas ; elle ressort comme non-significative dans C01 lorsqu'on exclut les codas en position finale d'unité intonative (pour neutraliser l'effet de l'allongement final), mais très significative dans la totalité des données (avec le même filtrage des données).

Ces résultats préliminaires ne suffisent toutefois pas à tirer quelque conclusion tranchée et définitive concernant l'hypothèse d'ambisyllabité. Il conviendrait dans un travail ultérieur de vérifier si la distribution des consonnes intervocaliques et des voyelles observées ne représente pas un biais dans nos résultats. En effet, on peut supposer qu'en cas, par exemple, de proportion d'obstruantes nettement supérieure à celle de sonorantes dans les consonnes ambisyllabiques, on aura un effet différent sur la voyelle qui précède ainsi que la durée de la consonne ambisyllabique. La même supposition s'applique à la voyelle précédente : si les voyelles précédant les consonnes ambisyllabiques sont par exemple majoritairement des diphtongues, on imagine qu'elles auront tendance à être plus longues de par leurs caractéristiques intrinsèques mais pas nécessairement en fonction de la consonne suivante. Il convient par conséquent d'évaluer l'effet de la distribution des consonnes intervocaliques et des voyelles précédentes avant de pouvoir donner une conclusion plus tranchée de ces résultats.

Nous avons décidé dans un deuxième temps de réaliser une expérience en production dans le but d'explorer la problématique d'ambisyllabité plus en détail. La

section suivante est donc consacrée à la description de la mise en place de cette expérience ainsi qu'à l'interprétation des résultats que nous avons obtenus.

7.3. Expérience en production

En complément des différents tests effectués sur la base de données Aix-MARSEC, nous avons décidé d'organiser une expérience en production visant à observer plus en détail les stratégies de syllabification de sujets anglophones natifs. Le but global de cette expérience est que les sujets découpent sans ambiguïté en syllabes des phrases construites au préalable afin d'observer leurs découpages syllabiques (regroupement de la consonne intervocalique en position de coda, d'attaque ou double rattachement).

L'originalité de cette expérience est que nous incitons les sujets à découper les phrases en syllabes sans toutefois mentionner le mot « syllabe », ceci dans le but d'éviter toute influence d'une définition préconçue de la syllabe.

Comme nous l'avons vu dans le Chapitre 2 (section 2.3.3.4), les tâches proposées aux sujets dans les travaux précédents dans ce domaine sont de nature diverse. Elles se présentent le plus souvent sous la forme de tâches d'inversion de syllabes (Treiman et Danis 1988 ; Treiman, Straub et Lavery 1994), de répétitions d'une (Fallows 1981) ou deux syllabes (Giegerich 1992) d'un mot dissyllabique ou d'insertions de pauses (Derwing 1992).

Les deux premières tâches posent le problème majeur de modification de la structure rythmique de l'énoncé ou en l'occurrence du mot. En effet, l'accent sera automatiquement déplacé et la structure rythmique par conséquent modifiée. Dans le mot *guitar* /gɪ'ta:/ par exemple, l'accent porte sur la deuxième syllabe ; on a donc selon le modèle soit 'anacrouse + NRU' (Jassem), soit 'silent ictus + pied' (Abercrombie). Or, après inversion des syllabes, on a une NRU ou un pied dissyllabique, la structure rythmique étant alors fondamentalement différente. De plus, si on reprend l'argument d'Anderson (1986), la présence d'une frontière de pied bloque la resyllabification. Si on prend un exemple du type *sugar* et qu'on inverse les deux syllabes, on obtient une structure de type /gə'ʃu/ dans laquelle la deuxième syllabe représente un pied ; la consonne intervocalique étant en position initiale de pied, elle est nécessairement regroupée avec la deuxième syllabe.

En revanche, dans la tâche d'insertion de pause, un tel problème ne se pose pas puisque la structure rythmique est intacte. Dans ce test, nous avons incité les sujets à ralentir leur débit et à découper les phrases en syllabes. Cette tâche, même si elle présente des inconvénients majeurs, permet d'éviter les problèmes de modification de la structure rythmique, comme c'est le cas avec la tâche d'inversion de syllabes.

7.3.1. *Elaboration des consignes*

La première partie de ce travail a donc été la mise au point de consignes visant par des moyens détournés à obtenir des sujets un découpage syllabique sans pour autant le leur demander explicitement.

Pour cela, nous donnons comme prétexte à chacun des sujets que nous travaillons sur un système de reconnaissance automatique de l'anglais dont le but est de reconnaître, décoder et comprendre les différents messages qui lui sont adressés. Pour augmenter son efficacité, il est nécessaire d'entraîner le système et c'est dans ce but que nous recherchons des sujets pour lire une série de phrases anglaises afin de tester le système. Nous expliquons ensuite que le système ne reconnaît pas la totalité des messages et que les sujets pouvaient être amenés à effectuer trois types de lecture. En effet, pour chacune des phrases, les sujets font une première lecture naturelle. En fonction de cette lecture, le sujet reçoit un message en retour lui disant si la phrase a été décodée (message de remerciement) ou pas. Si la phrase n'a pas été décodée, un autre message demande au sujet de répéter cette même phrase (qui s'affiche à nouveau) mais cette fois plus lentement. Au terme de cette deuxième lecture, un autre message précise si l'énoncé a été compris (message de remerciement) ou pas, auquel cas le système demande au sujet de faire une troisième et dernière répétition. Au cours de cette troisième lecture, il est cette fois demandé à chacun des sujets de répéter encore plus lentement, plus fort et en surarticulant, comme s'ils s'adressaient à une personne éloignée ou à une personne malentendante. Ces différentes étapes sont résumées graphiquement en Annexe 3.3.

Par conséquent, le test se fonde sur trois lectures potentielles de chacune des phrases présentées :

- Phase 1 (considérée comme "naturelle")
- Phase 2 (considérée comme "lente")

- Phase 3 (considérée comme "hyperarticulée")

C'est au cours de cette Phase 3 que nous incitons implicitement les sujets à découper les phrases en syllabes. En effet, nous avons précisé au cours des consignes qu'ils doivent surarticuler et, dans le cadre de cette surarticulation, découper les phrases en petites unités avec si possible une courte pause entre ces unités, en expliquant que le système a plus de facilité à décoder les phrases lorsque celles-ci sont clairement découpées pour ensuite reconstituer le message global. Notons que nous avons pris comme facteurs pour cette troisième lecture le ralentissement du débit, l'augmentation de l'intensité et l'hyperarticulation, ceci en partie en fonction de l'étude de Eefting (1991) sur le débit.

7.3.1.1. Phase d'entraînement (sujets francophones)

Avant de débiter l'expérience avec des sujets anglophones, nous avons fait passer le test à dix sujets francophones (7 locutrices, 3 locuteurs) afin de vérifier si, en fonction de la tâche demandée, nous obtenions le résultat voulu lors de la Phase 3, à savoir un découpage syllabique sans ambiguïté. Cette phase d'entraînement correspond uniquement à une phase de mise au point méthodologique qui ne sera pas prise en compte dans les analyses.

Il convient de préciser que c'est lors de cette phase d'entraînement que les consignes ont le plus évolué. En effet, en fonction des réponses données par les sujets francophones, nous avons au fur et à mesure ajusté les consignes pour obtenir le résultat attendu. Pour effectuer ce test, nous avons utilisé une présentation PowerPoint, celle-ci est présente en annexe sur le CD-ROM 2 (dossier 'Présentations', 'français', fichier 'test_français.ppt').

Dans un premier temps, nous demandions systématiquement aux sujets de répéter chaque phrase trois fois (Phases 1, 2 et 3). Mais au cours de cette phase d'entraînement, nous nous sommes rendue compte que la triple répétition posait le problème majeur d'anticipation : leur lecture avait tendance à devenir peu naturelle car les sujets anticipaient la Phase 3 en faisant une réalisation hyperarticulée dès la Phase 1. Nous avons donc décidé de modifier l'organisation globale du test ; nous avons inséré des phrases plus courtes dont nous ne tiendrons pas compte dans l'analyse et qui seront considérées comme "comprises" par le système dès la première lecture, voire la deuxième. Autrement dit, le nombre de lectures varie désormais entre un et trois afin de

neutraliser tout effet d'anticipation. Le nombre de phrases est donc plus important mais nous évitons un biais dû à l'anticipation des sujets.

De plus, certaines phrases sont également présentées à deux reprises. Cette répétition vise elle aussi à contribuer à une attention accrue de la part du sujet qui, généralement, se rend compte qu'il a déjà lu la phrase.

Dix phrases françaises sont utilisées dans cette première partie d'entraînement, trois phrases avec une seule lecture (Phase 1), deux phrases avec deux lectures (Phases 1 et 2) et cinq phrases avec trois lectures (Phases 1, 2 et 3). Ces dernières sont plus nombreuses puisque notre but est justement de tester l'efficacité des consignes de la Phase 3. Le choix des phrases françaises est tout à fait arbitraire puisque seules les consignes sont ici testées.

Malgré la tâche, les sujets francophones ne découpaient pas systématiquement en syllabes au cours de la Phase 3 (en groupes rythmiques pour la plupart, en mots pour certains). Nous avons donc ajouté comme consigne pour la Phase 3 qu'ils découpent en unités les plus petites possibles, en partant de l'hypothèse que leur réflexe sera de découper en syllabes plutôt qu'en phonèmes puisque nous avons choisi de nous adresser à des sujets non spécialistes. De plus, suite à une remarque faite par un des sujets francophones, nous avons ajouté aux consignes déjà existantes que la troisième lecture correspondait à une lecture relativement mécanique, proche d'une production synthétique.

7.3.1.2. Phase d'entraînement (sujet anglophone)

Une fois la tâche testée avec des sujets francophones, nous avons procédé à une phase d'entraînement supplémentaire avec une locutrice anglophone, locutrice qui ne sera pas réutilisée ultérieurement pour le test.

Au cours de ce test, la tâche de la Phase 3 s'est avérée infructueuse car le découpage en « unités les plus petites possible » ne correspondait pas à la syllabe. Nous avons par conséquent décidé d'ajouter une phase-test au début de l'expérience. Ainsi, nous pouvons vérifier si le sujet répond correctement aux consignes et si le résultat escompté est obtenu. Lorsque les sujets continuent à découper les phrases en unités autres que la syllabe (le plus souvent le mot), nous leur avons donné un exemple de phrase découpée en syllabes, mais dans laquelle il n'y a aucun cas de consonne potentiellement

ambisyllabique (en fonction d'une production standard) dans le but de pas influencer les sujets sur les stratégies de syllabification. La phrase utilisée est la suivante :

'Mary was absent from the course'

Dans une variété d'anglais non rhotique, le phonème /r/ ne peut se trouver en position finale de syllabe, il est donc nécessairement en position d'attaque de deuxième syllabe. De plus, /bs/ n'est licite ni en coda, ni en attaque, la frontière syllabique se trouve par conséquent entre /b/ et /s/.

Ainsi, les sujets commencent le test uniquement lorsque leur production correspond à la tâche que nous leur avons demandée. Aucune contrainte temporelle n'est imposée aux sujets, ils peuvent aller au rythme qui leur convient. En cas de répétition, seule la dernière lecture est prise en compte.

Etant donné que nous n'avions pas révélé la tâche, nous avons pu faire un deuxième essai avec le même sujet anglophone et nous avons obtenu cette fois un découpage syllabique des énoncés lors de la Phase 3. C'est uniquement après ces diverses tentatives que nous avons commencé notre expérience avec des sujets anglophones dont nous retiendrons les résultats.

7.3.1.3. Ajout d'une deuxième partie

Au cours de cette phase d'élaboration et d'évaluation des consignes, nous avons décidé d'ajouter une deuxième partie au test au cours de laquelle nous demandons également aux sujets de lire une série de phrases mais en leur demandant explicitement de les découper en syllabes et de marquer une courte pause entre chacune d'elles. Ces phrases sont uniquement celles qui ont été lues trois fois au cours de la première partie du test.

Cette deuxième phase, que nous nommerons Phase 4, nous permet d'avoir un élément de comparaison avec la production des phrases dans le cadre de la Phase 3. En effet, cette phase nous permet d'avoir une idée de la conscience phonologique des sujets en terme de syllabification. Ainsi, nous pouvons comparer le découpage syllabique au cours des Phases 3 et 4 et voir si le découpage réalisé consciemment (Phase 4) est identique au découpage produit de manière "inconsciente", en production contrainte (Phase 3).

Par souci de clarté, nous ferons référence aux Phases 1, 2 et 3 pour les trois types de lecture effectués dans le cadre de la première partie et à la Phase 4 pour la lecture de la deuxième partie.

7.3.2. *Elaboration de l'expérience*

Une fois les consignes établies et testées pendant la phase d'entraînement, nous avons procédé à l'élaboration de l'expérience, à savoir les phrases que les sujets doivent lire et le format sous lesquelles elles se présentent.

7.3.2.1. Stimuli

Les stimuli utilisés sont des phrases variées que nous avons créées pour les besoins de l'expérience. Nous avons dans un premier temps choisi un ensemble de mots plurisyllabiques dans lesquels nous avons fait varier les divers paramètres présentés dans les différentes études psycholinguistiques comme ayant une influence sur l'apparition de consonnes ambisyllabiques en anglais (Chapitre 2). Nous avons ensuite agencé ces mots dans des phrases variant en terme de longueur, de complexité syntaxique et de modalité.

En fonction des résultats observés dans les études psycholinguistiques, les divers paramètres utilisés dans le choix des mots sont les suivants :

- Patron accentuel du mot : variation de l'emplacement de l'accent lexical et variation de la longueur des mots (une, deux ou trois syllabes, et quelques cas de mots avec un nombre supérieur de syllabes)
- Type de consonnes intervocaliques : maximum de consonnes différentes en position intervocalique
- Nombre de consonnes intervocaliques : taille du groupe consonantique intervocalique
- Consonne orthographique simple ou double : consonne simple ou doublement de la consonne
- Type de voyelle précédant la consonne intervocalique : variété de voyelles précédant les consonnes intervocaliques.

Nous ne prétendons pas à l'exhaustivité des occurrences de chacun de ces paramètres ni de l'interaction des divers paramètres manipulés, nous avons plutôt essayé d'en

combiner un maximum dans un nombre minimum de phrases pour que l'expérience ne soit pas trop longue.

Ces différents mots sont agencés en 40 phrases qui sont ensuite présentées aux sujets (cf. Annexe 3.4.1). Parmi ces 40 phrases, 17 sont répétées trois fois (Phases 1, 2 et 3) ; nous avons varié le nombre de répétitions de chacune des phrases et certaines phrases apparaissent deux fois afin d'éviter tout effet d'anticipation des sujets et de conserver leur attention. En dehors de ces 17 phrases, 10 sont répétées deux fois (Phases 1 et 2) et 13 ne sont répétées qu'une seule fois (Phase 1).

Parmi ces 40 phrases, 21 sont sélectionnées pour la deuxième partie de l'expérience (cf. Annexe 3.4.2) au cours de laquelle on demande explicitement aux sujets de découper les phrases en syllabes.

Il convient de préciser que, contrairement aux expériences effectuées en psycholinguistique, il ne s'agit pas ici uniquement de non-mots ou de mots isolés, mais de phrases variées regroupant des mots existants. Même si ces phrases n'ont rien de naturel, il ne s'agit pas d'une phrase porteuse avec un mot différent à chaque production.

7.3.2.2. Aspects techniques

Nous avons utilisé pour l'expérience une interface sous la forme de deux présentations *PowerPoint* (une pour chacune des deux parties) que les sujets manipulent à leur vitesse et à leur convenance. Les deux présentations sont présentes sur le CD-ROM 2 (dossier 'Présentations'). Pour la première présentation, les consignes sont répétées à chaque fois, c'est-à-dire que chaque phrase est précédée d'une diapositive rappelant les consignes en fonction de la phase de lecture : lecture naturelle pour la Phase 1, lecture lente pour la Phase 2, lecture lente, surarticulée et plus forte pour la Phase 3. On espère ainsi maintenir le respect des consignes.

Les enregistrements sont tous effectués avec des micro-casques en chambre sourde au Laboratoire Parole et Langage, avec une numérisation des données audio (fréquence d'échantillonnage de 44.1kHz, codage linéaire sur 16 bits). Pendant l'expérience, l'unité centrale de l'ordinateur utilisé est disposée suffisamment loin du micro-casque afin d'éviter tout bruit de fond lié au ventilateur.

L'expérience dure environ 30 minutes avec une pause entre les deux présentations. Les phases de test pour la première et la deuxième partie sont réalisées avec l'expérimentateur et ne sont pas enregistrées. Une fois les consignes assimilées, les sujets se retrouvent seuls dans la chambre sourde. Deux fichiers-son sont alors créés, un pour chaque présentation.

7.3.2.3. Sujets

Les sujets participant à l'expérience sont des sujets anglophones natifs (lecteurs ou étudiants à l'Université de Provence) se situant tous dans la tranche d'âge 18-25 ans. Nous avons fait passer le test à 14 anglophones britanniques, 4 locuteurs et 10 locutrices, la majorité venant du Sud de l'Angleterre. Deux locutrices sont irlandaises et deux autres viennent du nord de l'Angleterre, mais ces quatre sujets ne ressortent pas comme ayant des caractéristiques divergentes. De plus, la disponibilité des sujets natifs ne permet pas de pouvoir comparer les caractéristiques propres à chaque accent. Il est toutefois envisageable de continuer cette expérience avec pour but de voir s'il existe des différences entre divers dialectes. Les sujets ont tous passé les deux parties du test et sont chacun identifiés par une lettre de l'alphabet.

De manière globale, les sujets n'ont pas éprouvé de grosses difficultés à accomplir la tâche qui leur était demandée. Grâce à la phase-test précédant chacune des deux parties, nous pouvions vérifier que les sujets avaient bien assimilé les consignes pour éviter de devoir les interrompre pendant l'enregistrement. Certains sujets ont cependant moins respecté les consignes que d'autres ; le locuteur B, par exemple, n'a pas systématiquement découpé les phrases en syllabes au cours de la Phase 3.

7.3.3. Mise en forme des données

Les deux fichiers-son correspondant aux deux parties sont segmentés en phrases (avec une, deux et/ ou trois lectures) puis sauvegardés comme tels. Les fichiers-son sont nommés selon la convention suivante : lettre correspondant au sujet + numéro de la phrase + phase de lecture. Ainsi, la lecture du sujet A de la première phrase (une seule lecture) est nommée Aphr1ph1.wav. Pour la deuxième phrase avec Phases 1, 2, 3 et 4, on obtient les fichiers-son Aphr2ph3.wav (Phases 1, 2 et 3) et Aphr2ph4.wav (Phase 4).

Afin de pouvoir exploiter ces données, il est nécessaire d'obtenir une représentation du découpage syllabique. Pour cela, nous avons tout d'abord étiqueté certains fichiers avec le logiciel Praat afin d'avoir une représentation concrète des différents découpages réalisés par les sujets sous la forme d'un étiquetage et d'un alignement temporel. Pour les Phases 2, 3 et 4, nous avons créé un fichier TextGrid (fichier d'étiquettes dans Praat) correspondant à chaque fichier-son, avec une transcription orthographique pour les Phases 2, 3 et 4 et un niveau supplémentaire d'étiquetage avec segmentation au niveau syllabique et transcription phonémique (large) en SAMPA pour les Phases 3 et 4. Nous avons jugé que les fichiers-étiquettes n'étaient pas nécessaires pour la Phase 1 puisque celle-ci n'est pas exploitée (aucun découpage syllabique).

La Figure 41 ci-après représente un extrait de l'étiquetage du fichier Aphr13ph3.wav dans le logiciel Praat.

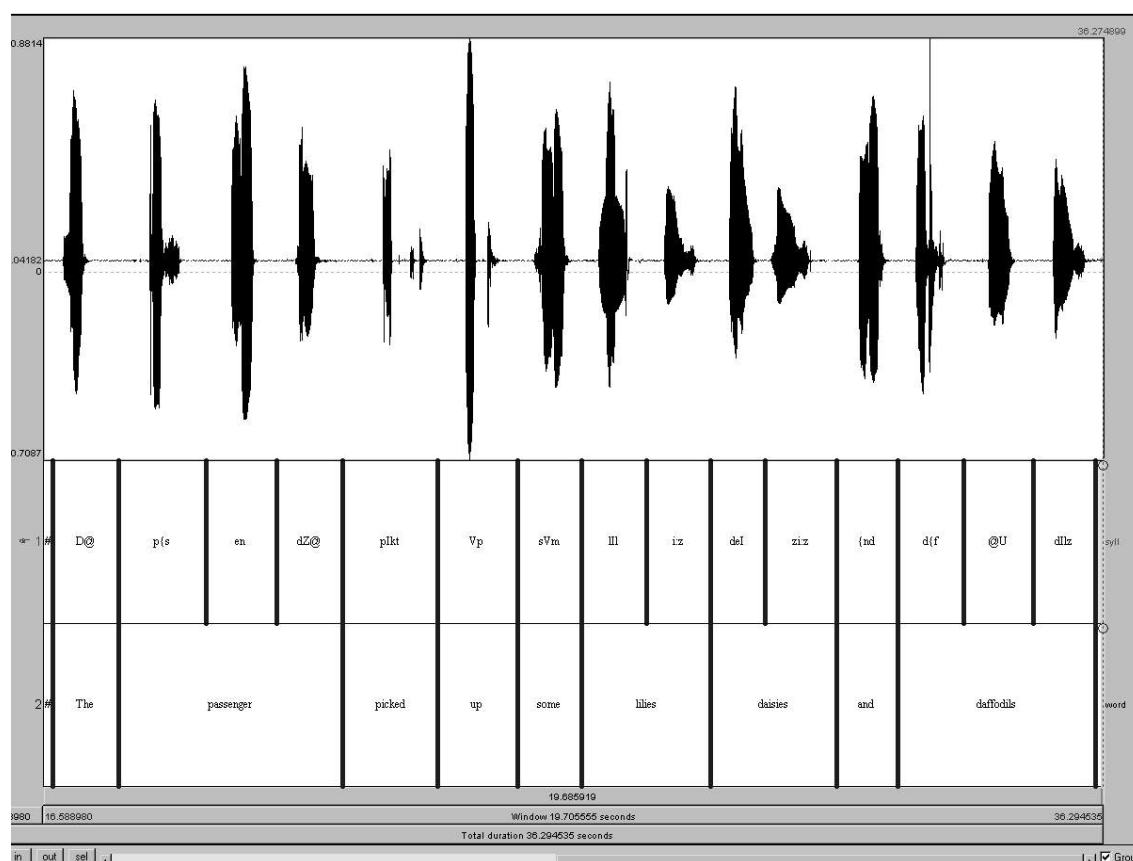


Figure 41. Extrait de l'étiquetage de la Phase 3 du fichier Aphr13ph3.wav (the passenger picked up some lilies, daisies and daffodils) avec au premier niveau la transcription phonémique et l'alignement temporel au niveau de la syllabe et au deuxième niveau la transcription orthographique et l'alignement temporel au niveau du mot

Les TextGrids sont présents en annexes sur le CD-ROM 2 (dossier ‘TextGrid’) ; les fichiers-son pour les locuteurs A-J se trouvent sur le CD-ROM 2 (dossier ‘Enregistrements’), les fichiers restant sont sur le CD-ROM 3.

Pour observer objectivement les découpages syllabiques, nous avons comptabilisé les pauses apparaissant entre les mots et entre les syllabes. Pour cela, un premier pré-traitement automatique est appliqué à chacun des fichiers-son dont le but est de détecter automatiquement les pauses. Cette opération est effectuée grâce à un script Praat rédigé par C. Auran²⁸ permettant de détecter automatiquement les pauses selon un seuil temporel minimal choisi et en fonction de la courbe d’intensité du segment analysé (cf. CD-ROM 2, fichier ‘get_pauses_v2.6.praat’). Le seuil donné ici est de 200ms, seuil généralement adopté dans la détection de pauses. La sortie de cette implémentation est un fichier d’étiquettes Praat avec un premier niveau consacré au marquage des pauses et un deuxième avec le regroupement des segments inter-pauses.

Cette phase de traitement automatique est nécessaire pour avoir un codage objectif et systématique des pauses, en évitant ainsi une interprétation des phrases produites par les sujets. On considère par conséquent qu’il y a découpage syllabique uniquement lorsque le sujet a marqué une pause supérieure ou égale à 200ms. En revanche, les fichiers avec détection automatique des pauses ne peuvent être directement exploités car de nombreuses erreurs de surdétection subsistent, notamment dans le cas des consonnes non voisées. Ces fichiers nous ont donc uniquement servi de référence pour notre segmentation manuelle dans Praat dans le but d’obtenir un codage objectif. Les TextGrids obtenus automatiquement et les TextGrids manuels sont présents sur les CD-ROM 2 et 3 avec les fichiers-son de chaque locuteur.

Nous avons ensuite catégorisé manuellement ces pauses dans un tableau selon qu’elles apparaissent entre deux mots ou qu’elles correspondent à une frontière syllabique, auquel cas nous précisons s’il s’agit d’un regroupement en attaque, en coda ou d’un regroupement ambisyllabique. Ainsi, nous obtenons un tableau de données avec pour chaque phrase cinq types d’informations : le nombre total de pauses, le nombre de

²⁸ Téléchargeable à partir de : http://aune.lpl.univ-aix.fr:16080/~auran/ressources/get_pauses_v2.6.praat

pauses inter-mots, le nombre de pauses inter-syllabes selon le principe d'attaque maximale, de coda maximale ou d'ambisyllabité.

Le tableau final (cf. CD-ROM 2, dossier 'données', fichier 'Tableau1(global).xls') regroupe les 14 locuteurs produisant au total 1470 phrases, soit 105 phrases par locuteur (toutes phases confondues) dont 13 phrases pour la Phase 1 uniquement, 20 pour la Phase 2, 51 pour la Phase 3 et 21 pour la Phase 4. Au total, le corpus regroupe 3036 découpages syllabiques, tous locuteurs et toutes phases de lecture confondus. Un court extrait de ce tableau global est donné ci-après (Tableau 17) : la première colonne précise le locuteur, la deuxième la phrase, la troisième la phase de lecture, la quatrième le nombre total de pauses ('npauses'), la cinquième le nombre de pauses entre les mots ('npm'), la sixième le nombre de regroupements en attaque ('npo'), la septième le nombre de regroupements en coda ('npc') et la huitième le nombre de regroupements ambisyllabiques ('npa').

Loc	Phrase	Phase	npauses	npm	npo	npc	npa
A	1	1	0	0	0	0	0
A	2	1	0	0	0	0	0
A	2	2	1	1	0	0	0
A	2	3	13	8	3	1	1
A	2	4	14	8	4	2	0
A	3	1	1	1	0	0	0
A	3	2	2	2	0	0	0
A	3	4	13	6	4	2	1

Tableau 17. Extrait du tableau global de données (tableau(global).xls, CD-ROM 2)

Précisons que les informations relatives à la première (Phases 1, 2 et 3) et à la deuxième partie du test (Phase 4) sont regroupées dans un seul tableau.

Nous avons ensuite isolé dans deux tableaux séparés tous les cas de regroupements ambisyllabiques et codaïques (CD-ROM 2, dossier 'données', fichiers 'tableau2(ambi).xls' et 'Tableau3(coda).xls'). Pour chaque rattachement ambisyllabique et codaïque, nous avons déterminé manuellement un ensemble d'informations correspondant aux facteurs observés par Treiman et Danis (1988). Ces caractéristiques sont les suivantes :

- Mot en question ('Mot_ambi' ou 'Mot_coda' selon le fichier) et localisation graphique du site de découpage décrit grâce au tiret
- Locuteur ('Loc')
- Numéro de la phrase dont est extrait le mot ('Phr')
- Phase de lecture ('Ph')
- Consonne ('Cons')
- Type de consonne ('Type') : occlusive, fricative, liquide, nasale ou affriquée
- Complexité ('Compl') : '1' pour une consonne simple, '2' pour les consonnes doubles et '&' pour les consonnes mixtes
- Cluster ('Cluster') : codage des groupes consonantiques ('0' lorsqu'il n'y a qu'une seule consonne et '1' lorsqu'il y a au moins deux consonnes différentes)
- Voyelle précédente ('V préc')
- Voyelle suivante ('V suiv')
- Patron accentuel ('Accent') : '1' pour un accent primaire et '2' pour un accent secondaire, '0' pour une syllabe inaccentuée

La complexité orthographique recouvre trois catégories : consonnes simples, doubles ou mixtes. Nous avons ajouté la catégorie mixte pour distinguer les mots avec plusieurs consonnes orthographiques différentes. Par exemple, dans le mot *ashamed*, la consonne intervocalique /ʃ/ est codée '&' (mixte) car il ne s'agit ni d'une consonne orthographique simple (comme dans *city*), ni d'une consonne orthographique double (comme dans *butter*). La catégorie cluster a été ajoutée pour faire la distinction entre des cas comme *butter* et *accent*, les deux ont une consonne orthographique double mais dans le deuxième cas, plus d'une consonne est réalisée.

Un extrait de chaque tableau de données est donné ci-après : Tableau 18 pour les regroupements ambisyllabiques et Tableau 19 pour les regroupements codaïques.

Mot_ambi	Loc	Phrase	Phase	Cons	Type	Compl	Cluster	V préc	V suiv	Accent
rem-minds	A	2	3	m	nas	1	0	ɪ	aɪ	0-1
hel-lo	A	3	4	l	liqu	2	0	æ	əʊ	0-1
abs-solutely	A	4	4	s	fric	1	1	æ	əʊ	2-010
absol-lutely	A	4	3	l	liqu	1	0	ə	u:	20-10
pil-low	A	6	3	l	liqu	2	0	ɪ	əʊ	1-0
suffocat-ted	A	6	4	t	occl	1	0	eɪ	ɪ	100-0
an-nybody	A	7	4	n	nas	1	0	e	ɪ	1-000

Tableau 18. Extrait du tableau2(ambi).xls (CD-ROM 2)

Mots_coda	Loc	Phrase	Phase	Cons	Type	Compl	Cluster	V préc	V suiv	Accent
moth-er	A	2	3	ð	fric	&	0	ʌ	ə	1-0
cam-era	A	2	4	m	nas	1	0	æ	ə	1-00
moth-er	A	2	4	ð	fric	&	0	ʌ	ə	1-0
accomm-odation	A	3	4	m	nas	2	0	ɒ	əʊ	02-010
acc-ommodation	A	3	4	k	occl	2	0	æ	ɒ	0-2010
scen-ery	A	4	4	n	nas	1	0	i:	ə	1-00
suff-ocated	A	6	3	f	fric	2	0	ʌ	əʊ	1-000

Tableau 19. Extrait du Tableau3(coda).xls (CD-ROM 2)

7.3.4. Interprétation des résultats

7.3.4.1. Répartition des différents types de découpages syllabiques

Une fois toutes les pauses répertoriées, nous pouvons observer quelles tendances ressortent des productions des 14 sujets.

Lorsque les locuteurs découpent les phrases en syllabes au cours des Phases 3 et 4, ils le font majoritairement en fonction du principe d'attaque maximale. En effet, comme nous pouvons l'observer dans le Tableau 20, environ 84% des découpages syllabiques se font en fonction du principe d'attaque maximale ; en revanche, dans 13% des cas, la consonne intervocalique est regroupée avec la syllabe précédente (position de coda) et

dans environ 3% des cas la consonne est regroupée avec les deux syllabes (ambisyllabique).

	Attaque	Coda	Ambisyllabique	Total
Nombre	2563	387	86	3036
Pourcentage	84%	13%	3%	

Tableau 20. Répartition en pourcentages des différents types de découpages syllabiques pour les 14 locuteurs.

Etant donné la différence de consigne qui existe entre les quatre phases de lecture, notamment entre la Phase 3 et la Phase 4, il convient de tenir compte pour cette répartition du type de lecture dont il est question. La Figure 42 et le Tableau 21 ci-après représentent la répartition des découpages en fonction du type de lecture.

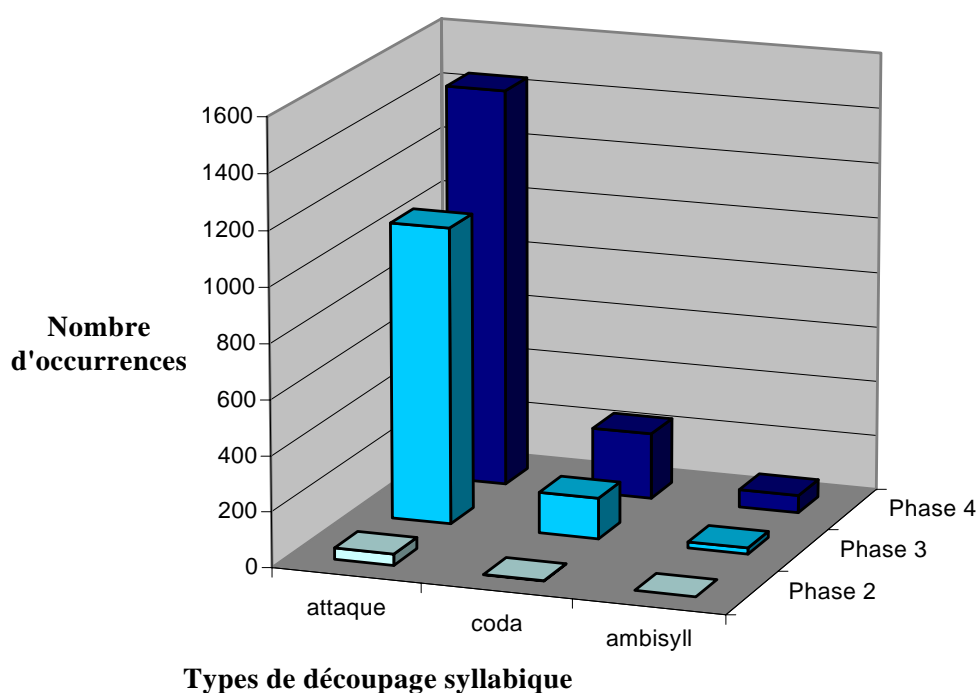


Figure 42. Illustration de la répartition des types de découpage syllabique en fonction de la phase de lecture (Phases 2, 3 et 4)

La Phase 1 est exclue des données prises en compte pour le tableau et la figure car aucun sujet n'a effectué de découpage syllabique au cours de cette phase de lecture. Quelques découpages syllabiques apparaissent au cours de la Phase 2 mais ceux-ci sont

rares et témoignent d'une anticipation de la tâche demandée pour la Phase 3 ; ils ne seront pas pris en compte pour les analyses statistiques.

	Pourcentages		
	Attaque	Coda	Ambisyllabique
Phase 2	97,6%	2,4%	0%
Phase 3	86,3%	11,9%	1,8%
Phase 4	82,7%	13,6%	3,7%

Tableau 21. Répartition exprimée en pourcentages des types de découpage (Attaque, Coda et Ambisyllabique) en fonction de la phase de lecture (2, 3 ou 4).

Il ressort de la Figure 42 et du Tableau 21 que le découpage en fonction du principe d'attaque maximale est largement majoritaire. En effet, la proportion (Tableau 21) de découpage selon le principe d'attaque maximale représente respectivement 97.6%, 86.3% et 82.7% des types de découpage dans les Phases 2, 3 et 4 de lecture, une écrasante majorité qui, dans le contexte de ce travail, contredit l'hypothèse d'un découpage systématique en fonction du principe d'ambisyllabité.

La première remarque que l'on peut faire à partir de la Figure 42 est que les sujets ont relativement "bien" répondu aux consignes qui leur ont été données au cours de la Phase 3 puisqu'ils effectuent de nombreux découpages. La Figure 42 et le Tableau 21 révèlent également que pour chacun des trois types de découpage, les occurrences sont plus nombreuses dans la Phase 4 que dans la Phase 3 (dans un rapport d'environ une fois et demi). Ce résultat n'est pas étonnant dans le sens où, dans la Phase 4, nous demandons explicitement aux sujets de découper en syllabes alors que nous ne faisons que les inciter à le faire au cours de la Phase 3.

On remarque également que les proportions des trois types de découpage changent au passage des Phases 3 et 4. En effet, la proportion de découpage en fonction du principe d'attaque maximal décroît au passage à la Phase 3 puis à la Phase 4. Par voie de conséquence, les proportions des deux autres découpages augmentent.

Il est également intéressant de remarquer que le type de découpage syllabique réalisé pour ce test dépend fortement des locuteurs (Figure 43). En effet, certains locuteurs ont tendance à découper uniquement selon le principe d'attaque maximale (plus de 90% des découpages), c'est le cas des locuteurs D, E, F, J, K et M. D'autres

produisent de nombreux cas de regroupements codaïques comme les locuteurs A, C, G, I, L et N. Seul le locuteur B réalise un certain nombre de regroupements ambisyllabiques (tendance restant toutefois minoritaire). Ces remarques sont résumées dans la Figure 43 ci-après.

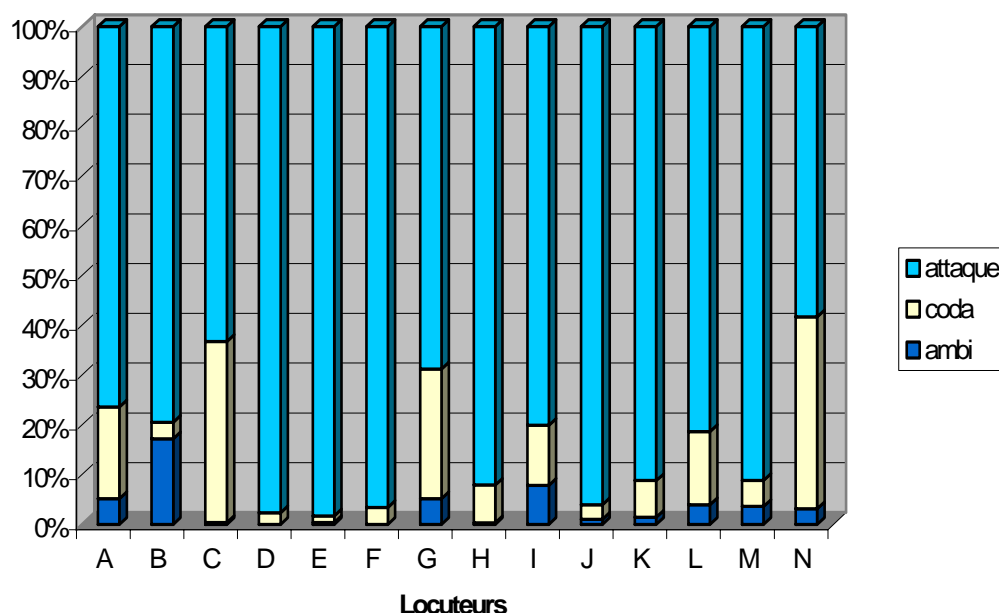


Figure 43. Différents types de découpage syllabique en fonction des locuteurs

Si on compare maintenant ces cas d'ambisyllabité aux regroupements codaïques (Figure 43), on s'aperçoit qu'il n'y a pas ou peu de correspondance. En effet, les sujets produisant le plus de cas d'ambisyllabité ne sont pas également ceux qui ont tendance à regrouper les consonnes intervocaliques avec la syllabe précédente (coda). On pourrait plutôt parler de complémentarité puisque, en dehors des regroupements (majoritaires) en attaque, les sujets ont tendance à faire des regroupements codaïques OU ambisyllabiques. Cette tendance peut notamment s'observer pour les sujets A, B, C, G, L et N. Seul le locuteur I semble avoir des parts égales de regroupements codaïques et ambisyllabiques.

Après cette vue d'ensemble des résultats, nous allons observer plus en détail les occurrences des trois types de découpage syllabique, tout d'abord les occurrences de regroupements ambisyllabiques, puis les regroupements codaïques et enfin les découpages selon le principe d'attaque maximale.

7.3.4.2. Occurrences de consonnes ambisyllabiques

On dénombre dans le corpus 86 cas de regroupements ambisyllabiques. La Figure 44 ci-après représente une des réalisations du locuteur B (Bphr4ph4.wav) dans laquelle on trouve deux cas de regroupement ambisyllabique (entourés sur la figure) : *suf-focated* et *pil-low*.

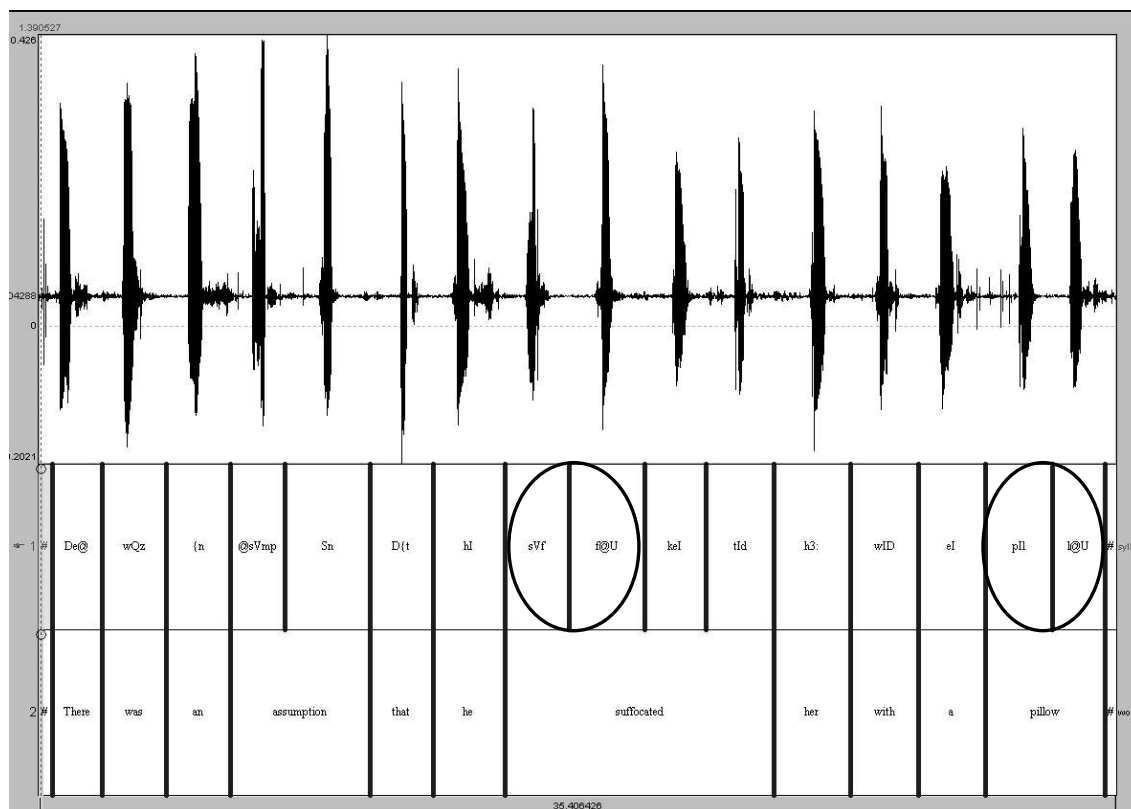


Figure 44. Illustration dans Praat de la Phase 4 de la phrase 4 lue par le locuteur B

Comme nous l'avons souligné plus haut (cf. Tableau 20), les regroupements ambisyllabiques ne représentent qu'environ 3% du nombre total des découpages syllabiques dans le corpus (86 cas) : les découpages ambisyllabiques sont donc très minoritaires. De plus, aucun des 14 locuteurs ne découpe majoritairement en fonction de ce principe. Certains sujets ont tendance à produire des regroupements ambisyllabiques (notamment A, B et I), mais aucun n'applique systématiquement ce type de regroupement. Par conséquent, on parle plus de tendance (plus ou moins marquée) plutôt que de fait systématique. Nous pouvons voir dans la Figure 45 ci-après le nombre de regroupements ambisyllabiques par locuteur. Une certaine variabilité

inter-locuteur ressort de cette figure mais également un nombre limité d'occurrences des regroupements ambisyllabiques (avec un maximum de regroupements de 21 pour le locuteur B).

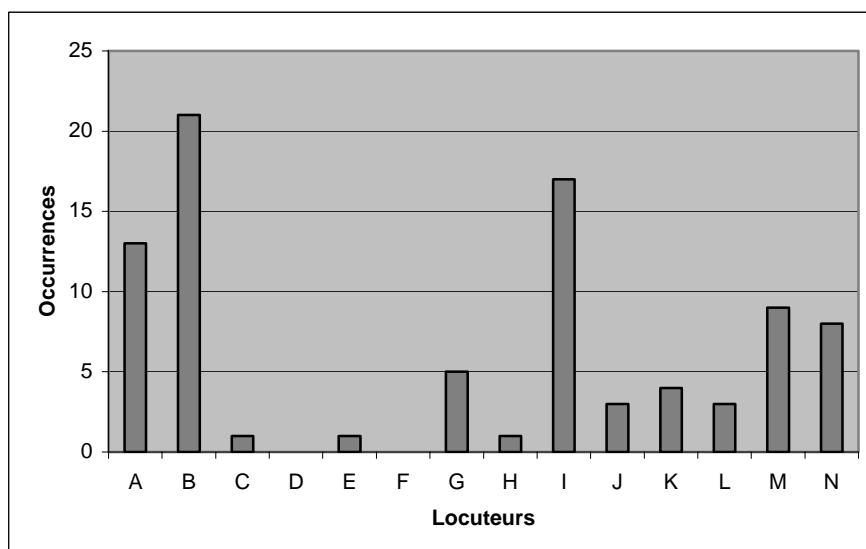


Figure 45. Nombre d'occurrences de cas d'ambisyllabité en fonction des locuteurs

Nous allons tester l'effet des différents facteurs observés notamment par Treiman et Danis (1988) sur l'apparition des consonnes ambisyllabiques. Ces facteurs sont le patron accentuel, le type de consonne intervocalique, la complexité orthographique de la consonne et le type de voyelle précédente. Nous allons donc voir si ces facteurs s'avèrent être pertinents dans nos données.

Pour tester chaque facteur, nous calculons son coefficient *chi*² sur les effectifs. Ce coefficient permet de déterminer si l'effet d'un facteur donné est dû au hasard ou s'il est réellement pertinent. Ainsi, pour chaque facteur énuméré ci-dessus, nous avons opposé les deux caractéristiques qui le définissent :

- Patron accentuel : position pré- vs post-accentuelle
- Catégorie de consonne : obstruante vs sonorante
- Complexité orthographique : consonne simple vs consonnes doubles²⁹
- Voyelle précédente : voyelle courte vs voyelle longue

²⁹ Nous n'avons pas tenu compte dans ce test des consonnes codées mixtes car elles sont au nombre de quatre pour les consonnes ambisyllabiques.

Nous avons ensuite créé des matrices d'effectifs pour chaque facteur avec les effectifs pour les deux caractéristiques en contexte ambisyllabique et non ambisyllabique.

Etant donné la taille du corpus, nous n'avons pas pu manuellement comptabiliser le nombre exact d'occurrences non ambisyllabiques pour chaque caractéristique (comme par exemple toutes les réalisations de sonorantes non ambisyllabiques). Nous avons donc décidé de comptabiliser la totalité de ces occurrences à partir de la liste des phrases présentées aux sujets : nous avons donc compté à partir des phrases tous les cas de consonnes intervocaliques pré- vs post-accentuelles, sonorantes vs obstruantes, orthographiquement simples vs complexes, ainsi que tous les cas de voyelle précédente courte vs longue.

On obtient par conséquent une matrice avec quatre types différents d'effectifs. Un exemple d'une telle matrice est donnée dans le Tableau 22 ci-après pour la catégorie de consonne intervocalique. On a donc 64 cas de sonorantes ambisyllabiques, 1434 de sonorantes non ambisyllabiques, 22 cas d'obstruantes ambisyllabiques et 2232 d'obstruantes non ambisyllabiques. Les occurrences non ambisyllabiques sont par conséquent des estimations plutôt que des cas réalisés par les sujets. La totalité des matrices pour les cas d'ambisyllabité est présentée en Annexe 3.5.1 et les résultats des χ^2 en Annexe 3.6.1.

	Ambisyllabiques	Non ambisyllabiques	TOTAL
Sonorantes	64	1434	1498
Obstruantes	22	2232	2254
TOTAL	86	3666	3752

Tableau 22. Matrice d'effectifs pour la catégorie de consonne intervocalique selon qu'il s'agit d'une sonorante ou une obstruante et selon qu'elle est ambisyllabique ou pas.

A partir de ces matrices, nous avons calculé le coefficient χ^2 pour déterminer si l'effet du facteur testé est significatif (en fonction des effectifs estimés). Ce calcul est effectué dans le logiciel de statistiques R grâce à la formule `chisq.test(x)`, où `x` est la matrice des effectifs. Les résultats pour les quatre facteurs testés sont donnés dans le Tableau 23 ci-après. La première colonne précise les facteurs testés, la deuxième les caractéristiques prises en compte pour chaque facteur, la troisième le nombre

d'occurrences, la quatrième les pourcentages, la cinquième les valeurs du coefficient χ^2 et la sixième la valeur p .

FACTEURS	DIMENSIONS	NOMBRE	POURCENTAGES	CHI ²	P
Type de consonne	Sonorantes	64	74,4	42,2018	8,23e-11
	vs				
	Obstruantes	22	25,6		
Patron accentuel (accents primaires uniquement)	Pré-accentuel	12	14	5,6445	0,01751
	vs				
	Post-accentuel	56	65,1		
Patron accentuel (incluant les accents secondaires)	Pré-accentuel	12	14	101,294	<2,2e-16
	vs				
	Post-accentuel	67	77,9		
Complexité orthographique	Simple	31	36	43,285	4,73e-11
	vs				
	Double	51	59,3		
Voyelle précédente	Voyelle courte	82	95,3	6,1959	0,0128
	vs				
	Voyelle longue	4	4,7		

Tableau 23. Résultats des χ^2 pour les quatre facteurs testés sur les cas de regroupement ambisyllabique

Il ressort de ces résultats que les effets des cinq facteurs sont significatifs ($p < 0.018$), les proportions de chaque dimension en contexte ambisyllabique et non ambisyllabique sont donc statistiquement différentes. L'effet des facteurs "type de consonne" et "complexité orthographique" est très marqué alors que l'effet des facteurs "patron accentuel" (en ne tenant compte que des accents primaires) et "voyelle précédente" ont des différences de proportion moins marquées. En revanche, si on reteste l'effet du patron accentuel en tenant compte des accents secondaires, l'effet est très significatif.

Après observation des effectifs, on peut dire que dans 74.4% des cas, la consonne intervocalique ambisyllabique est une sonorante. On rejoint ici les observations de Treiman et Danis puisque les auteurs constatent que les consonnes ambisyllabiques sont majoritairement des sonorantes. On peut également observer dans ces données que dans près de 60% des cas, la consonne orthographique est double. Ici aussi, nos résultats

confirment ceux obtenus par Treiman et Danis. Ainsi, lorsque les sujets rencontrent une consonne double, comme dans *pillow*, ils semblent être influencés par la redondance de la consonne orthographique et ils ont alors tendance à placer la frontière syllabique entre ces deux consonnes, ils produisent alors deux fois la consonne correspondante, une fois en tant que coda de la première syllabe et une fois en tant qu'attaque de la deuxième syllabe.

En ce qui concerne l'effet du patron accentuel, nous avons considéré dans un premier temps les positions pré et post-accentuelles en ne tenant compte que des accents primaires. La valeur p dans ce cas là est d'environ 0,018 ce qui signifie que l'effet est significatif, même si son effet est moins marqué que d'autres (comme le type de consonne par exemple). Dans un deuxième temps, nous avons regroupé les positions pré et post-accentuelles en tenant compte cette fois des accents primaires et secondaires ; l'effet du patron accentuel est cette fois très marqué ($p < 2.2e-16$). Près de 78% des consonnes ambisyllabiques sont post-accentuelles (accent primaire ou secondaire), ce qui signifie que les consonnes ambisyllabiques apparaissent majoritairement en position post-accentuelle mais rarement en contexte pré-accentuel (comme l'avaient également observé Treiman et Danis).

L'effet de la voyelle précédente est également significatif ($p=0.013$), bien que moins marqué que certains autres facteurs. Dans environ 95% des cas d'ambisyllabité, la voyelle précédente est une voyelle courte. Ce résultat, qui reste à confirmer avec davantage de données, rejoint un des arguments en faveur de l'ambisyllabité, à savoir la nécessité de fermer les syllabes accentuées ouvertes avec voyelles courtes tout en conservant une attaque non vide. Notons que, malgré un pourcentage aussi important, la valeur p est relativement élevée car le nombre d'occurrences est très faible (seulement quatre cas de voyelle longue pour 82 cas de voyelle courte)

7.3.4.3. Regroupement en position codaïque

On compte 387 cas de regroupements codaïques dans le corpus. On trouve deux exemples dans la lecture de la phrase 1 (Phase 4) par le locuteur A (Figure 46 ci-après).

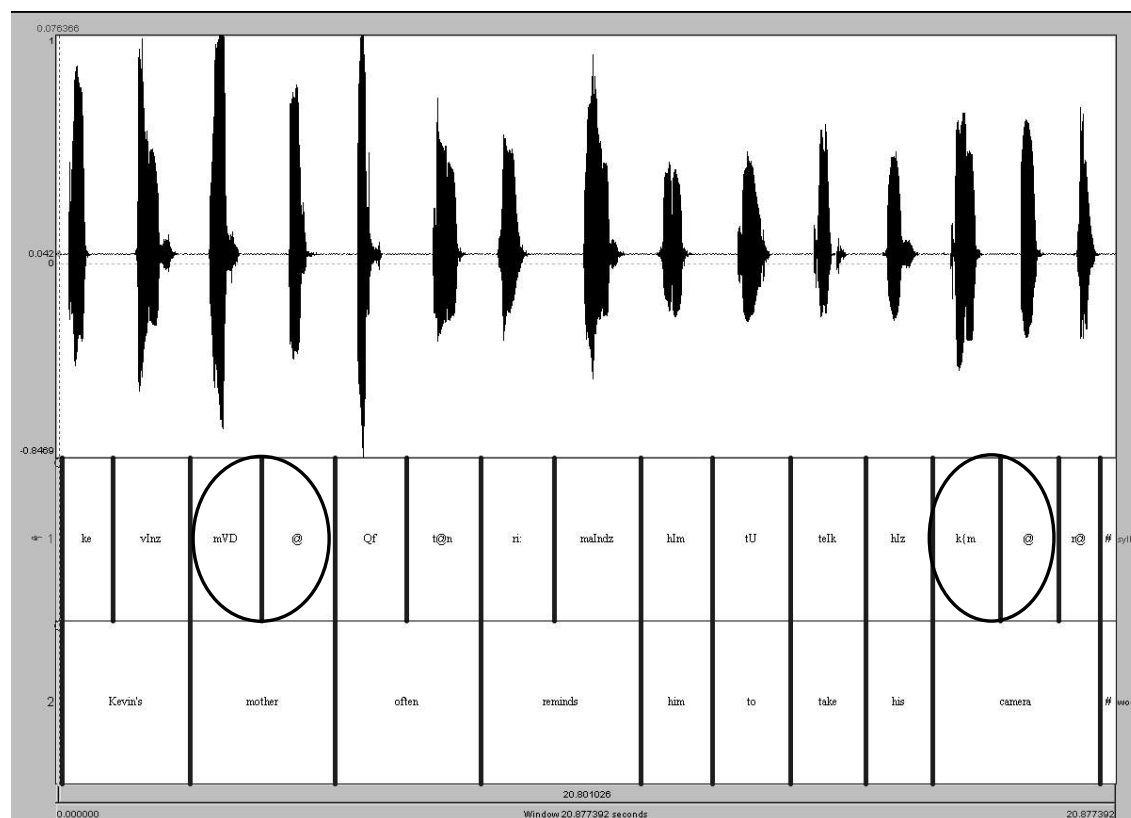


Figure 46. Phase 4 de la phrase 1 lue par le locuteur A

13% des regroupements de consonnes intervocaliques se font en position codaïque de syllabe précédente (cf. Tableau 20). Ce type de regroupement représente également une part minoritaire de syllabification. Aucun sujet ne syllabifie systématiquement en fonction de ce principe, même si certains sujets ont tendance à faire de nombreux regroupements codaïques, comme par exemple les sujets A, C, G, I, L et N (Figure 43). Il est intéressant de voir que, globalement, ce type de syllabification est environ quatre fois plus important que les regroupements ambisyllabiques.

Nous avons vu que l'argument principal concernant ce type de syllabification est d'ordre phonotactique : on ne peut pas avoir en anglais de syllabe accentuée avec voyelle courte, il faut que celle-ci soit fermée par la consonne intervocalique. Les arguments en faveur des regroupements codaïques et ambisyllabiques sont très similaires, la différence est en fait de savoir si une attaque vide est acceptable ou pas. Dans les deux cas, la consonne intervocalique doit être rattachée à la syllabe précédente si celle-ci est accentuée et ouverte avec une voyelle courte. En revanche, le double rattachement de la consonne intervocalique (ambisyllabité) relève plus de la tendance universelle à éviter les attaques vides. Au vu de cette remarque, nous avons décidé de

tester sur les regroupements codaïques les mêmes facteurs que pour les cas d'ambisyllabité.

Nous avons par conséquent testé l'influence des quatre facteurs décrits par Treiman et Danis sur l'apparition de regroupements codaïques. Nous avons créé dans un premier temps des matrices d'effectifs pour chaque facteur pour ensuite pouvoir calculer le coefficient χ^2 de chaque facteur. Les matrices pour ce type de regroupement sont présentées en Annexe 3.5.2 et les résultats des χ^2 en Annexe 3.6.2. Les résultats obtenus sont résumés dans le Tableau 24 et on peut voir que l'effet est très significatif pour tous les facteurs testés.

FACTEURS	DIMENSIONS	NOMBRE	POURCENTAGES	CHI ²	P
Type de consonne	Obstruantes	204	52,7	9,4101	0,00216
	vs				
	Sonorantes	183	47,3		
Patron accentuel (accents primaires uniquement)	Pré-accentuel	75	19,4	12,6183	0,00038
	vs				
	Post-accentuel	253	65,4		
Patron accentuel (incluant les accents secondaires)	Pré-accentuel	80	20,7	10,7654	0,00103
	vs				
	Post-accentuel	274	70,8		
Complexité orthographique	Simple	171	46,3	121,143	< 2,2e-16
	vs				
	Double	170	42,3		
	vs				
	Mixte	46	11,4		
Voyelle précédente	Voyelle courte	363	93,8	23,4482	1,28e-06
	vs				
	Voyelle longue	24	6,2		

Tableau 24. Résultats des χ^2 pour les regroupements codaïques

Contrairement aux résultats précédents, dans 52.7% des cas de regroupements codaïques, la consonne intervocalique est une obstruante. Cette différence est principalement due au grand nombre de fricatives regroupées en position codaïque : les

fricatives représentent environ 30% des consonnes (les liquides représentant environ 24%).

Dans environ 65% des cas, la consonne intervocalique codaïque est en position finale de syllabe avec accent primaire, cette proportion passe à environ 70% si on tient également compte des accents secondaires. On retrouve ici un résultat similaire aux regroupements ambisyllabiques, ce qui est tout à fait compréhensible puisque les deux regroupements répondent aux mêmes exigences de fermeture des syllabes accentuées ouvertes avec voyelle courte. Globalement, les consonnes codaïques sont donc principalement post-accentuelles.

En ce qui concerne la complexité de la consonne intervocalique, nous avons ici inclus la caractéristique ‘mixte’ étant donné que les effectifs le permettaient. Dans 46.3% des cas, la consonne intervocalique orthographique est simple, contre 42.3% pour les consonnes doubles et 11.4% pour les consonnes mixtes. Il y a donc une tendance à ce que la consonne intervocalique soit codaïque lorsque la consonne orthographique est simple ou double, contrairement aux cas où la consonne orthographique est mixte (comme par exemple dans le mot *ashamed*). En revanche, il ne ressort aucune préférence marquée pour les consonnes simples ou les consonnes doubles.

Dans près de 94% des cas, la voyelle précédente est une voyelle courte. Ce résultat combiné au résultat sur les patrons accentuels correspond à l’argument de fermeture de syllabe accentuée avec voyelle courte. De plus, si on s’intéresse uniquement aux consonnes suivant un accent primaire, 91.6% d’entre elles suivent une voyelle courte. Ce résultat va tout à fait dans le sens de la fermeture des syllabes accentuées ouvertes avec voyelle courte, mais reste toutefois marginal.

7.3.4.4. Regroupement selon le principe d’attaque maximale

Il y a au total 2563 cas de regroupements en attaque dans le corpus. Par exemple, dans la phrase 16 lue par le locuteur A pendant la Phase 3 de lecture (Figure 47 ci-après), on compte 12 cas de regroupement d’une consonne intervocalique en attaque.

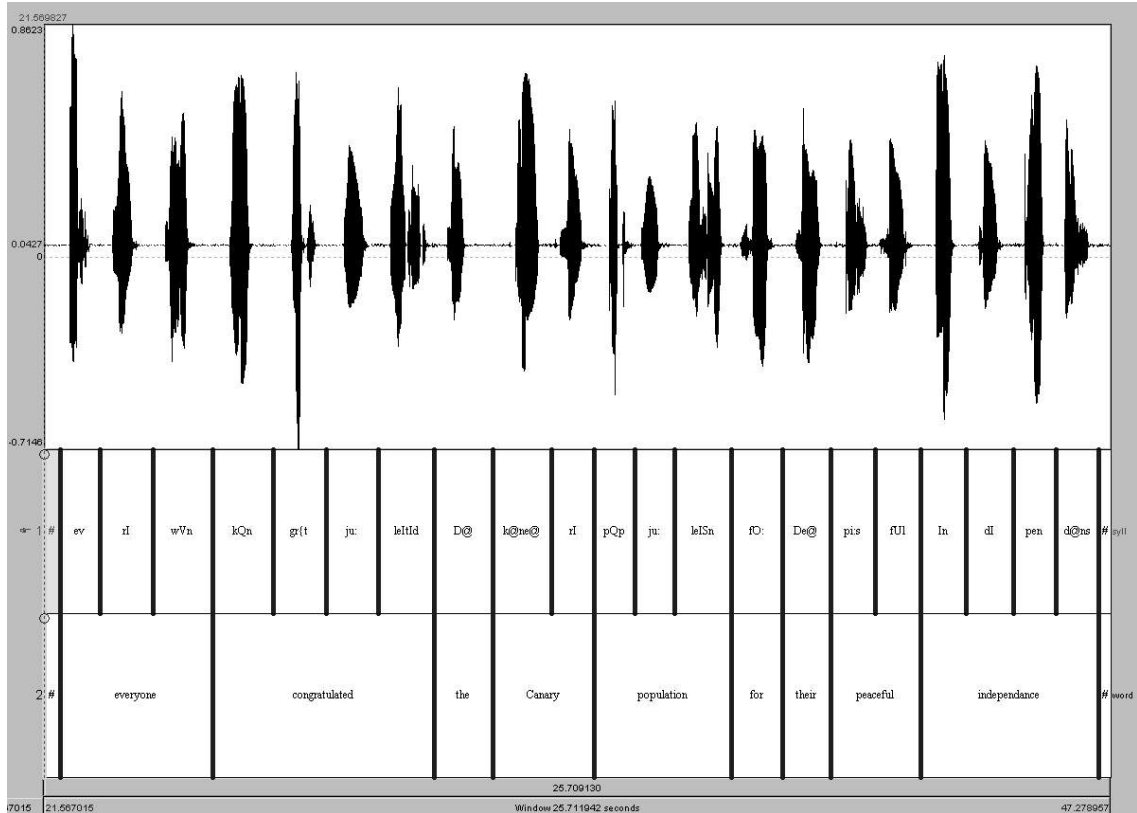


Figure 47. Phase 3 de la phrase 16 lue par le locuteur A

Ce découpage est celui majoritairement utilisé par les sujets puisque dans 84% des cas, le découpage se fait en fonction du principe d'attaque maximale : un maximum de consonnes intervocaliques est placé en position d'attaque de deuxième syllabe tout en respectant les règles phonotactiques de cooccurrence. Ce résultat va à l'encontre des hypothèses de rattachement systématique de la consonne intervocalique à la syllabe précédente lorsque celle-ci est accentuée (phénomène amplifié si la syllabe est ouverte et la voyelle courte).

Ainsi, dans 84% des cas, les sujets suivent le principe universel de syllabification ouverte. Il existe une variabilité inter-locuteur importante (cf. Figure 43) mais tous, sans exception, syllabifient majoritairement en fonction de ce principe.

Contrairement aux regroupements codaïques et ambisyllabiques, nous n'avons pas effectué le même codage manuel de toutes les occurrences de regroupement en attaque par manque de temps puisqu'il y a au total 3279 cas de regroupement en attaque à coder. On ne peut par conséquent pas effectuer les analyses de *chi*² pour tester l'effet des différents facteurs jusqu'ici observés.

7.3.4.5. Comparaison des découpages des Phases 3 et 4

Rappelons que la différence entre ces deux phases de lecture réside dans les consignes données aux sujets : ils sont incités à découper les phrases en syllabes dans la Phase 3 alors qu'on leur demande explicitement de le faire au cours de la Phase 4. D'un point de vue strictement auditif, les deux phases ne semblent pas diverger en terme de naturel. Cette section est consacrée à la comparaison des différents types de découpage syllabique dans les deux phases.

Pour chacun des locuteurs, nous avons comparé les cas de correspondance et de non correspondance de découpage syllabique entre les Phases 3 et 4. Pour cela, nous avons relevé manuellement tous les sites de découpage syllabique (quel qu'il soit) en commun dans les Phases 3 et 4 pour chaque locuteur (1161 pour les 14 locuteurs). Nous avons ensuite comparé ces découpages un à un dans le but de dégager des tendances, qu'elles soient globales ou propres à un locuteur. Nous observons donc ici uniquement les cas où, à une frontière syllabique donnée, les locuteurs ont effectué un découpage dans la Phase 3 ET dans la Phase 4, que le regroupement soit identique ou pas.

Le nombre de cas pour chaque locuteur est très variable, un locuteur ayant relativement bien répondu aux consignes aura de nombreux cas de comparaison. En revanche, les locuteurs ayant effectué peu de découpages (notamment les sujets B, G et L) ont (très) peu de cas de comparaison.

Globalement, il y a plus de cas de correspondance que de non correspondance pour tous les locuteurs. Autrement dit, parmi les sites avec un découpage syllabique dans la Phase 3 et la Phase 4, la syllabification est identique dans la majorité des cas, les pourcentages d'adéquation parmi les sujets allant de 80 à 98.5%. On remarque toutefois qu'aucun des sujets n'obtient 100% d'adéquation : même ceux qui syllabifient majoritairement en attaque ne le font jamais dans 100% des cas. Pour chaque locuteur, il existe une zone de variabilité plus ou moins étendue ; on a donc une forte variabilité intra et inter-locuteur.

La syllabification de certains sites est très stable pour quasiment tous les sujets. Ces sites (caractères en gras) sont entre autres dans les mots suivants : *Kevin's*, *independence*, *peaceful*, *fancy*, *party*, *behaviour*, *remove*, *city*, *occupied*, *happy*, *anchovy*, *Rachel*, *Fountains* et *nothing*. Ces cas de syllabification sont relativement stables notamment de par la nature de la consonne intervocalique, les contraintes phonotactiques et/ ou la présence d'une frontière morphémique. Par exemple, il est

assez compréhensible que le mot *peaceful* soit toujours syllabifié /pi:s.fəl/ puisque /sf/ n'est ni une coda ni une attaque licite et il existe une frontière morphémique forte entre les deux syllabes. Les mots *city* (/sɪ.tɪ/) et *happy* (/hæ.pɪ/) ne présentent pas de tels arguments, on suppose que la syllabification stable vient alors de la nature de la consonne intervocalique (plosive dans les deux cas).

Par opposition, certains mots posent des problèmes aux sujets ; on trouve par exemple les mots *pillow*, *selling*, *lullaby*, *lilies*, *daffodils*, *officers*, *fishermen* et *illegally* pour lesquels il y a souvent divergence entre Phase 3 et Phase 4. On remarque qu'il s'agit ici de liquides et de fricatives, la plupart étant orthographiquement doubles. Même pour les sujets qui syllabifient massivement les consonnes intervocaliques en attaque, ces mots posent problème et leur syllabification est peu stable ; c'est notamment le cas des sujets K et M qui syllabifient *pillow* /pɪl.əʊ/, /pɪl.ləʊ/ ou /pɪ.ləʊ/. Ce mot présente effectivement des caractéristiques favorisant les regroupements codaïques ou ambisyllabiques : la consonne est une sonorante, la consonne orthographique est double et elle suit une syllabe accentuée ouverte avec voyelle courte. Les mêmes remarques s'appliquent au mot *selling*, avec en plus la présence d'une frontière morphémique.

Une autre frontière syllabique posant problème est la séquence /ks/ dans le mot *unacceptable*. Certains syllabifient /ək.sep/, d'autres /əks.ep/ ou encore /ə.ksep/ (alors que la séquence /ks/ n'est pas une attaque licite).

Les mots *camera* et *scenery* posent également des problèmes liés à l'élision potentielle du schwa. En effet, leur syllabification est différente selon que le schwa est réalisé ou pas. Dans *camera*, s'il y a élision (/kæmrə/), le /m/ aura tendance à être regroupé en coda mais, lorsqu'il n'y a pas élision, on trouve des traces de cette prononciation alternative. Pour le locuteur A par exemple (mais également chez les sujets C, I et J), on a les syllabifications /kæm.rə/ (Phase 3) et /kæm.ə.rə/ (Phase 4), on peut supposer que le sujet hésite quant à la réalisation du schwa et ceci se répercute sur le rattachement de la consonne intervocalique. Le même type de syllabification se trouve pour le mot *scenery*.

Le regroupement de consonnes intervocaliques simples en position de coda peut dans certains cas influencer la réalisation de la voyelle qui suit (avec une attaque vide). C'est le cas par exemple dans le mot *lullaby* (sujet A) syllabifié /lʌ.lə.baɪ/ (Phase 3) et

/lʌl.æ.baɪ/ (Phase 4). Dans la deuxième réalisation, la consonne /l/ est regroupée avec la première syllabe, la syllabe inaccentuée suivante n'a donc plus d'attaque et ceci se répercute sur la qualité de la voyelle qui devient pleine (/æ/). Toutefois, on peut également expliquer cette réalisation par l'influence, souvent marquée, de l'orthographe.

Ces diverses remarques ne sont qu'une première étape dans l'exploitation des données ; nous tentons pour l'instant de dégager des tendances plus ou moins systématiques, mais il conviendrait d'explorer plus en détail ces productions.

7.3.4.6. Bilan de l'expérience

Bien que cette expérience nous permette d'obtenir des résultats très intéressants sur les stratégies de syllabification en anglais britannique, il convient de préciser qu'il présente également deux inconvénients majeurs. Le premier est le manque de naturel ; cet aspect est difficilement évitable dans la mesure où le but est que les sujets découpent des énoncés en syllabes, en faisant en sorte que la syllabification soit sans ambiguïté pour que les regroupements soient catégorisés objectivement par l'expérimentateur. Nous avons évité toute interprétation au cours de cette catégorisation manuelle puisque nous avons abandonné de nombreux cas où la pause entre les deux syllabes était inférieure à 200ms.

Deuxièmement, la présentation des stimuli aux sujets est faite de manière visuelle sous la forme de phrases écrites. Il est par conséquent assez normal que l'on retrouve une influence importante de l'orthographe sur la production des sujets. En effet, on trouve dans les données de nombreux exemples illustrant cette influence. Par exemple, certains sujets, notamment le locuteur G, réalisent les pronoms indéfinis *a* et *an* /æ/ et /æn/. Lors d'une présentation de ce travail (Bouzon et Hirst 2004b), il nous a été suggéré de faire apprendre les phrases aux sujets puis de leur faire produire avec les mêmes consignes, ceci pour qu'ils ne soient pas influencés par la version orthographique. Cette suggestion demanderait un test beaucoup plus long pour les sujets puisqu'il faudrait d'abord apprendre les phrases. De plus, l'expérimentateur serait obligé de rester avec les sujets ; or, nous pensons que la présence de l'expérimentateur dans la chambre sourde pourrait être un facteur supplémentaire entraînant une production non naturelle.

Cette influence de l'orthographe est à mettre en parallèle avec une autre remarque importante, à savoir une certaine tendance, mais pas systématique, à réorganiser la structure rythmique des énoncés/ mots produits. En effet, lorsqu'ils découpent les phrases en syllabes, certains sujets ont tendance à réaliser chaque syllabe comme étant une unité rythmique/ pied accentuel monosyllabique. La conséquence directe porte sur la qualité de la voyelle qui n'est alors plus réduite. Les exemples sont nombreux, on peut citer *potatoes* (/pɒ.teɪ.təʊz/), *lullaby* (/lʌl.æ.bai/), *unacceptable* (/ʌn.æk.sept.æ.bl/), *amount* (/æ.maʊnt/), *police* (/pəʊ.li:s/), *tobacco* (/təʊ.bæ.kəʊ/) et *ashamed* (/æ.ʃeɪmd/).

La réorganisation rythmique n'est pas systématique, que ce soit chez un locuteur ou sur la totalité des locuteurs, ce qui donne des réalisations alternatives intéressantes. Par exemple, le mot *daffodils* peut être réalisé (1) /dæf.əʊ.dɪlz/ mais également (2) /dæ.fə.dɪlz/. Dans (1), le sujet réalise trois unités rythmiques (la deuxième voyelle est pleine) auquel cas la consonne /f/ est rattachée à la première unité évitant ainsi une syllabe accentuée ouverte avec voyelle courte. Dans (2), il n'y a pas de réorganisation rythmique et le mot représente une unité rythmique (deuxième voyelle réduite), la syllabification se fait alors en fonction du principe d'attaque maximale pour éviter une syllabe inaccentuée avec une attaque vide. Cette alternance s'observe également fréquemment avec le mot *suffocated* qui est réalisé /sʌf.əʊ.keɪ.tɪd/ ou /sʌ.fə.keɪ.tɪd/.

Par conséquent, nous pouvons reprocher à notre test d'induire une réorganisation de la structure rythmique des énoncés. En revanche, contrairement à la tâche d'inversion de syllabe (Treiman et Danis), cette réorganisation est une possibilité, attestée par certains locuteurs seulement, mais elle n'est pas directement induite par la tâche. C'est l'interprétation des locuteurs au moment de leur production qui fait qu'ils vont réorganiser la structure rythmique des énoncés.

Nous n'avons pas ici exploité la Phase 2 de lecture et, pourtant, une remarque intéressante peut être faite. Lorsque les sujets ralentissent leur débit, certains sujets ont tendance à découper les énoncés soit en mots soit en unités rythmiques. Quand ils découpent en unités rythmiques, il est intéressant de voir qu'ils le font selon le modèle de Jassem, à savoir en anacrouses + NRU c'est-à-dire en TRU (*Total Rhythm Unit*). Si on prend par exemple la lecture de la phrase 25 par le locuteur G (GPhr25Ph3.wav), le découpage est le suivant :

They were 'judging | 'Rachel | for il 'legally | 'selling | to 'bacco
 ANA NRU NRU ANA NRU NRU ANA NRU

Par conséquent, lorsque le sujet ralentit sa vitesse d'élocution, il découpe l'énoncé en unités rythmiques totales. Ce découpage est loin d'être systématique, mais il est intéressant de remarquer qu'il revient relativement fréquemment, que ce soit au niveau de l'énoncé entier comme c'est le cas dans l'exemple ci-dessus, ou pour seulement une portion de l'énoncé. Ces découpages tendent à montrer que le découpages en unités rythmiques selon le modèle de Jassem est plus intuitif que le découpage en pieds accentuels selon le modèle d'Abercrombie/ Halliday. Il convient toutefois de tester cette hypothèse plus en détail avant de faire une telle conclusion.

Les perspectives d'amélioration de cette expérience sont nombreuses ; il conviendrait tout d'abord de retravailler sur les consignes afin d'obtenir un résultat plus naturel. Une autre tâche linguistique permettrait peut-être que les sujets effectuent une production plus naturelle. Plus généralement, il serait souhaitable de faire passer le test à un nombre plus grand de sujets anglophones afin de voir si les tendances se confirment. Il serait également intéressant de voir s'il existe des variations lorsqu'on compare différentes variétés dialectales de l'anglais. Nous avons d'ailleurs déjà enregistré deux locutrices anglo-américaines et nous comptons agrandir ce groupe afin de voir s'il existe des différences de stratégie de syllabification entre les deux dialectes de l'anglais. Pour l'instant, ces deux sujets produisent 86.5% de cas de syllabification en attaque, 10.3% de regroupement codaïque et 3.2% d'ambisyllabité. Il est évident que ces données sont loin d'être suffisantes pour en tirer des conclusions, mais il semblerait que la même tendance se retrouve en anglais américain.

7.4. Résumé et conclusion

Ces résultats (avec les méthodologies qu'ils induisent) ont par conséquent plutôt tendance à contredire l'hypothèse du principe d'ambisyllabité en anglais britannique. Dans la base de données Aix-MARSEC, les résultats ne permettent pas de donner une conclusion tranchée concernant l'ambisyllabité, mais ils offrent des pistes de recherche à explorer. La différence en terme de durée entre les attaques et les consonnes ambisyllabiques ne ressort pas comme étant pas significative ; en revanche, elle l'est pour les voyelles précédentes. Il semblerait donc que les consonnes en attaque et les

consonnes ambisyllabiques ne se distinguent pas en fonction de leur durée mais leur influence sur la voyelle précédente semble différer. Ces tests doivent toutefois être étoffés en testant notamment l'effet de la distribution des consonnes et des voyelles en question.

Suite à ces résultats, nous avons réalisé une expérience en production au cours de laquelle des sujets anglophones britanniques sont incités à découper des énoncés en syllabes. Malgré des variations inter et intra-locuteurs, cette expérience montre que le découpage syllabique se fait majoritairement (84%) en fonction du principe d'attaque maximale. Seulement 3% des découpages sont effectués en fonction du principe d'ambisyllabité ; les 13% restant correspondent à des regroupements codaïques des consonnes intervocaliques.

Il convient de préciser qu'il s'agit ici d'une proposition d'approche pour tester ce phénomène ; d'autres méthodes peuvent être envisagées comme par exemple l'élaboration d'un corpus de type *Map Task* avec un locuteur anglais natif et un locuteur de nationalité différente parlant peu l'anglais. On peut alors imaginer que le locuteur natif devra surarticuler certains noms de lieu incompréhensibles pour le locuteur non natif ou même imaginer des contextes où le locuteur non natif demande au locuteur natif de répéter jusqu'à ce qu'il découpe certains mots en syllabes. Le même type de test pourrait être effectué avec des non-mots à la place des noms de lieu, même si nous pensons qu'il est préférable de ne pas se détacher de la parole authentique. Toutefois, il est probable qu'il faille un corpus énorme avant d'obtenir un nombre suffisant d'occurrences de regroupement syllabique.

D'autres méthodes de nature différente peuvent être également envisagées comme par exemple l'utilisation des potentiels évoqués. Cette méthode permet l'enregistrement des fluctuations électriques entre deux points du scalp (électro-encéphalogramme ou EEG), conséquences de l'activité électrique intracérébrale. Il faudrait ensuite adapter les stimuli en fonction du résultat espéré puisque les potentiels évoqués sont des réponses des voies nerveuses et du cerveau qui se produisent à la suite d'une stimulation. Les réponses sont par conséquent spécifiques à la nature du stimulus employé pour les obtenir.

La méthodologie et les résultats ici obtenus représentent par conséquent un point de départ de travail qu'il conviendrait d'étoffer en utilisant des méthodes complémentaires et/ ou différentes.

CONCLUSION

L'objectif de ce travail est d'explorer les deux problématiques de définition de l'unité rythmique et d'ambisyllabité, dans une perspective d'amélioration de la prédiction de l'organisation temporelle de l'anglais. Trois types d'approche prédictive ont été proposés en fonction de l'unité de prédiction, les modèles segmentaux, syllabiques et supra-syllabiques, les deux derniers résultant du problème fondamental de la variabilité importante de la durée des phonèmes et donc de la difficulté à prédire efficacement leur durée. Les modèles segmentaux sont qualifiés d'ascendants puisque le but est de prédire la durée des phonèmes pour ensuite obtenir la durée des unités supérieures. Par opposition, les modèles syllabiques et supra-syllabiques sont qualifiés de descendants dans le sens où on cherche à prédire la durée des syllabes ou autre unité supra-syllabique pour ensuite déduire la durée des phonèmes qui les composent. Ces deux approches descendantes posent toutefois deux questions fondamentales.

Premièrement, le but d'une approche supra-syllabique est de prédire la durée des unités rythmiques pour ensuite déterminer la durée des syllabes et des phonèmes qui les composent. Nous opposons dans ce travail deux modèles supra-syllabiques : le modèle classique d'Abercrombie (1964) et de Halliday (1967), dont l'unité rythmique est le pied accentuel, et le modèle de Jassem (1952), dont les unités rythmiques sont l'anacrouse et la NRU. Les deux unités principales (pied et NRU) sont caractérisées par le phénomène d'isochronie faible, mais le traitement des syllabes inaccentuées diffèrent fondamentalement. Les deux modèles sont comparés grâce à une série d'analyses de régression effectuées dans la base de données Aix-MARSEC représentant plus de cinq heures d'anglais britannique authentique et alignée à différents niveaux de représentation. Les résultats montrent qu'il n'existe pas d'isochronie stricte en anglais britannique au niveau de la syllabe, de l'unité rythmique, quelle qu'elle soit, du mot et de l'unité intonative. La durée de ces unités n'est par conséquent pas constante quelle que soit leur complexité. En revanche, malgré des résultats statistiques peu marqués,

nous observons un certain degré de compression (isochronie faible) des phonèmes et des syllabes aux niveaux de la NRU, du pied, des deux types d'anacrouse, du mot et de l'unité intonative, mais pas au niveau de la syllabe. En effet, il ressort que la durée des syllabes augmente quasi linéairement en fonction de leur complexité. En ce qui concerne les unités rythmiques, il existe un degré de compression plus important dans la NRU par rapport à l'anacrouse, la compression dans le pied étant proche de celle de l'anacrouse. Ce résultat confirme la nécessité de distinguer l'anacrouse de la NRU (contrairement au modèle classique d'Abercrombie/ Halliday) ; le modèle de Jassem semble donc mieux rendre compte de la structuration rythmique de l'anglais britannique.

La suite des tests porte sur l'effet de l'accent sur la durée des phonèmes et des syllabes car les deux modèles supposent que les syllabes comprises dans l'unité rythmique (pied ou NRU) se partagent équitablement la durée de l'unité rythmique, quel que soit leur statut accentuel. L'effet de l'accent est très significatif dans le pied ; en revanche, les résultats obtenus dans la NRU sont très surprenants puisque, si on exclut les deux derniers phonèmes des NRU en position finale d'unité intonative (supposés porter l'allongement final), on s'aperçoit que l'effet de l'accent n'est pas significatif. Ce résultat tend à montrer que l'effet de l'accent dans le pied est entièrement causé par la distinction entre l'anacrouse et la NRU, ce qui donne un argument supplémentaire en faveur du modèle de Jassem. Cette absence d'effet de l'accent est un résultat intéressant étant donné que l'effet de l'accent sur la durée des phonèmes est largement décrit dans la littérature ; or, il semblerait que ce soit l'appartenance à la NRU plutôt que l'effet de l'accent qui influe sur la durée des phonèmes.

Ces résultats sont intéressants car il semblerait que le modèle classique d'Abercrombie/ Halliday, largement adopté pour rendre compte de l'organisation rythmique de l'anglais, soit à remettre en cause. Le modèle proposé par Jassem semble en revanche donner une meilleure représentation de cette structuration rythmique. Il convient de tester plus en détail le modèle de Jassem dans le but de voir si sa supériorité vis-à-vis du modèle classique se confirme. La faiblesse des résultats statistiques concernant les analyses de régression peut être due à la mauvaise qualité de l'alignement automatique en phonèmes. Depuis peu, nous possédons environ 50 minutes de la base de données Aix-MARSEC alignés manuellement ; il serait par conséquent intéressant de refaire les tests statistiques sur cette portion du corpus pour

voir si les résultats se confirment et si les effets ressortent de manière plus tranchée une fois le biais de l'alignement nettement atténué. Pour approfondir ce test, il serait également intéressant de comparer les prédictions temporelles de chaque modèle sur de la parole authentique afin de voir lequel des deux modèles permet une meilleure prédiction des durées phonémiques et/ ou syllabiques.

La deuxième problématique de ce travail concerne la syllabification. En effet, les modèles syllabiques cherchent à prédire la durée des syllabes et soulèvent par conséquent la question, pour le moins épineuse, de la syllabification en anglais. La définition des frontières syllabiques en anglais a fait l'objet de nombreux débats sans toutefois aboutir sur une décision définitive. La question de la syllabification se pose notamment en cas de consonne intervocalique simple ou de groupe consonantique complexe à l'intérieur d'un mot ou à la frontière entre deux mots. C'est donc cette problématique de regroupement des éléments intervocaliques qui nous intéresse dans ce travail et que nous testons selon deux approches différentes en nous limitant toutefois aux consonnes intervocaliques à l'intérieur des mots.

Dans un premier temps, nous testons la différence de durée entre les consonnes potentiellement ambisyllabiques et les consonnes regroupées en attaque selon le principe d'attaque maximale dans la base de données Aix-MARSEC. Les résultats de ces tests sont mitigés puisqu'ils tendent, dans le passage C01, à montrer qu'il n'existe pas de différence de durée (normalisée) significative entre les deux types de consonnes intervocaliques ; ces résultats ne se retrouvent pas dans la totalité des données, mais ils peuvent être biaisés par le calcul des durées normalisées et par la quantité de données. De plus, la différence de durée moyenne brute des deux types de consonne se situe largement en dessous du seuil de perception. En revanche, la différence de durée concernant la voyelle précédente est significative à la fois dans C01 et dans la totalité des données, ce qui suggère un effet différent de la consonne intervocalique sur la voyelle précédente selon qu'elle se trouve en position d'attaque ou en position ambisyllabique. Toutefois, ces résultats ne nous permettent pas de donner quelque conclusion tranchée sur le principe d'ambisyllabité ; cette piste doit être explorée davantage, notamment en testant l'effet de la distribution des consonnes intervocaliques et des voyelles précédentes.

En complément de ces tests statistiques sur de la parole authentique, nous avons réalisé une expérience en production avec 14 sujets anglophones ; le but est de faire lire

des phrases à ces sujets en les incitant à les découper en syllabes. Il ressort de cette expérience une importante variation inter-locuteur ; toutefois on observe que dans 84% des cas, la syllabification est effectuée selon le principe d'attaque maximale qui consiste à placer un maximum de consonnes en position d'attaque de syllabe tout en respectant les contraintes phonotactiques de cooccurrence. Dans 3% des cas, la consonne intervocalique est ambisyllabique et dans 13% des cas elle est regroupée en position de coda de la syllabe précédente. Il semble donc que, dans le contexte de cette expérience, l'ambisyllabité ne ressorte pas comme faisant partie intégrante du système phonologique des sujets britanniques.

A partir de ces données, nous avons ensuite testé l'effet des facteurs décrits par Treiman et Danis (1988) comme affectant l'apparition des consonnes ambisyllabiques. Ces facteurs sont la position de l'accent, la qualité de la consonne, la qualité de la voyelle précédente et la complexité orthographique de la consonne intervocalique. Ces tests sont effectués à la fois sur les occurrences de regroupements ambisyllabiques mais également codaïques, grâce au test statistique du *chi*² sur les effectifs. Ces facteurs ont un effet significatif sur l'apparition des consonnes ambisyllabiques et codaïques. Les consonnes ambisyllabiques sont majoritairement des sonorantes, orthographiquement doubles, en position post-accentuelle (notamment si on tient compte des accents secondaires) et précédées d'une voyelle courte. Les consonnes codaïques sont plutôt des obstruantes, orthographiquement simples (mais avec des proportions respectives à peine plus importantes que celles des sonorantes et des consonnes orthographiques simples), en position post-accentuelle (notamment avec les accents secondaires) et précédées d'une voyelle courte. Même s'ils ne portent que sur 16% des regroupements syllabiques présents dans le corpus, ces résultats confirment que les sujets ont tendance à fermer les syllabes accentuées avec voyelle courte (argument principal en faveur de ces deux types de regroupement syllabique). Globalement, les résultats semblent montrer que les sujets, quand ils se trouvent dans un contexte de découpage syllabique, syllabifient principalement en fonction du principe d'attaque maximale. Ils regroupent les consonnes intervocaliques avec la syllabe précédente ou les deux syllabes environnantes que dans certains cas minoritaires.

Ce travail ne représente toutefois qu'un point de départ et il convient de tester plus en détail cette problématique en utilisant d'autres outils et d'autres approches méthodologiques. Nous avons évoqué à la fin du Chapitre 7 quelques méthodes qui

pourraient être utilisées pour explorer davantage la question de syllabification en anglais, comme par exemple d'autres tâches à présenter aux sujets (comme celle de *Map Task*) dans le but de les faire découper des énoncés en syllabes tout en conservant une production relativement naturelle. L'utilisation de potentiels évoqués pourrait également s'avérer très fructueuse. Outre les questions méthodologiques, il serait également intéressant de recueillir davantage de données en multipliant le nombre de locuteurs ainsi que d'étendre ce type de test à différentes variétés dialectales de l'anglais.

Comme nous l'avons précisé en introduction générale, le but de ce travail n'est pas l'élaboration d'un modèle de prédiction de l'organisation temporelle de l'anglais. En revanche, nous espérons que les résultats que nous avons obtenus concernant la syllabification et la structuration rythmique se révèlent être une piste fructueuse pour l'amélioration de la prédiction de l'organisation temporelle de l'anglais britannique et, à terme, de la production de parole synthétique.

REFERENCES BIBLIOGRAPHIQUES

- Abercrombie, D. (1964). Syllable quantity and enclitics in English. In Abercrombie, D., Fry, P., MacCarthy, N. et Trim, J. (eds): *In Honor of Daniel Jones*. London: Longman, pp.216-222.
- Abercrombie, D. (1967). *Elements of General Phonetics*. Great Britain: Edinburgh University Press.
- Allen, J.S. (1975). Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics*, **3**, pp.75-86.
- Allen, J.S. et Miller, J.L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of the Acoustical Society of America*, **106** (4), pp.2031-2039.
- Anderson, J.M. et Jones, C. (1974). Three theses concerning phonological representation. *Journal of Linguistics*, **10**, pp.1-26.
- Anderson, J.M. (1986). Suprasegmental Dependencies. In Durand, J. (éd.): *Dependency and Non-linear Phonology*, pp.55-135.
- Angoujard, J. (1997). *Théorie de la syllabe*. Paris : CNRS Editions.
- Auran, C. et Bouzon, C. (2003). Phonotactique prédictive et alignement automatique : application au corpus MARSEC et perspective. *Travaux Interdisciplinaires du laboratoire PArole et langage d'Aix-en-Provence*, **22**, pp.13-44.
- Auran C., Bouzon, C. et Hirst, D. (2002). *Aix-MARSEC : une proposition de traitement automatique de corpus d'anglais britannique oral*. Journée "Les Corpus Oraux en Anglais et en Français", 15 novembre, Toulouse, disponible à l'adresse suivante : http://www.lpl.univ-aix.fr/~EPGA/marsec_com/toulouse.pps.
- Auran C., Bouzon, C. et Hirst, D.J. (2004a). The Aix-MARSEC project: an evolutive database of spoken British English, *Speech Prosody 2004*, March 23-26 2004, Nara, pp.561-564.
- Auran, C., Bouzon, C. et Hirst, D.J. (2004b). Automatic Grapheme-Phoneme Conversion for Spoken British English Corpora: the Aix-MARSEC Methodology.

- PAC Workshop*, April 23-24 2004, Toulouse.
- Auran, C., Bouzon, C., Hirst, D.J., Levy, C. et Nocéra, P. (2004). Algorithme de prédiction d'élisions de phonèmes et influence sur l'alignement automatique dans le cadre du projet Aix-MARSEC. *XXV^{èmes} Journées d'Etudes sur la Parole*, 18-23 avril, Fès, Maroc, pp.57-60.
- Auran C et Hirst D. (2004). Anaphora, Connectors and Resetting: Prosodic and Pragmatic Parameters Interactions in the Marking of Discourse Structure. *Speech Prosody 2004*, March 23-26, 2004, Nara, pp.259-262.
- Barbosa, P. et Bailly, G. (1994). Characterization of Rhythmic Patterns for Text-to-Speech Synthesis. *Speech and Communication*, **15**, pp.127-138.
- Barbosa, P. et Bailly, G. (1997). Generation of Pauses Within the z-score Model. In Van Santen, J.P.H., Sproat, R.W., Olive, J.P. et Hirschberg, J. (eds): *Progress In Speech Synthesis*, pp.365-381.
- Basbøll, H. (1988). Phonological Theory. In Newmeyer, F.J. (éd.): *Linguistics: the Cambridge survey of linguistic theory foundation, vol. 1*. Cambridge: Cambridge University Press. pp.192-215.
- Beckman, M. (1992). Evidence for Speech Rhythms across Languages. In Tohkura, Y., Vatikiotis-Bateson, E. et Sagisaka, Y. (eds): *Speech Perception, Production and Linguistic Structure*, pp.457-463.
- Bell-Berti, F. (1991). Comments on "some observations on the organization and rhythm of speech". *ICPHS 91, 12^e Congrès International des Sciences Phonétiques*, Aix-en-Provence, France, pp.238-241.
- Benguerel, A.P. & Darcy, J. (1986). Time-warping and the perception of rhythm in speech. *Journal of Phonetics*, **14**, pp.231-246.
- Bertinetto, P.M. (1989). Reflections on the dichotomy "stress" vs "syllable-timing". *Revue de Phonétique Appliquée*, vol. 91-93, pp.99-130.
- Black, J.W. et Singh, S. (1968). The Physiological basis of phonetics. In Malmberg, B. (ed.): *Manual of Phonetics*. Pays-Bas: North-Holland Publisher, pp.105-128.
- Blevins, J. (1995). The Syllable In Phonological Theory. In Goldsmith, J.A. (éd.) : *The Handbook Of Phonological Theory*, pp.206-244.
- Boersma, P. et Weenink, D. (1996). Praat, a System for doing Phonetics by Computer, version 3.4. *Institute of Phonetic Sciences of the University of Amsterdam, Report 132*. Logiciel téléchargeable à partir de www.praat.org
- Bolinger, D.L. (1958). A theory of pitch accent in English. *Word*, **14**, pp.109-149.
- Bolinger, D.L. (1963). Temporal Predictability in the Perception of English Speech. In Cutler, A. and Ladd, D.R. (eds): *Prosody: Models and Measurements*, pp.111-121.

- Bolinger, D.L. (1965). Pitch accent and sentence rhythm. In Abe, I. and Kanekiyo, T. (eds): *Forms of English: Accent, Morpheme, Order*. Cambridge: Harvard University Press, pp.139-180.
- Bolinger, D.L. (1972). Accent is predictable (if you're a mind-reader). *Language*, **48**, pp.633-644.
- Bolinger, D.L. (1981). *Two Kinds of Vowels, Two Kinds of Rhythm*. Bloomington, USA: IULC Publications.
- Bouzon, C. (2002). Influences des caractéristiques des contours mélodiques sur la durée des mots en anglais britannique contemporain. *XXIV^{ème} Journées d'Études sur la Parole*. Nancy, France : LORIA et ATILF, pp. 69-72.
- Bouzon, C. (2004). Annotation et traitement automatique d'un corpus d'anglais britannique oral. *Anglophonia*, *14*, pp.201-219.
- Bouzon, C., Auran, C. et Hirst, D.J. (2003). Phonétisation, alignement et association dans les grands corpus oraux en anglais, *INTO 01*, 22-24 Mai 2003, Rouen. (http://www.lpl.univ-aix.fr/~EPG/marsec_com/rouen.pps)
- Bouzon, C. et Hirst, D.J. (2002). The Influence of Prosodic Factors on the Duration of Words in British English. In Bel, B. et Marlien, I. (eds): *Proceedings from Speech Prosody 2002*, Aix-en-Provence, pp.191-194.
- Bouzon, C. et Hirst, D.J. (2004a). Isochrony and prosodic structure in British English. *Speech Prosody 2004*, March 23-26 2004, Nara, pp.223-226.
- Bouzon, C. et Hirst, D.J. (2004b). Ambisyllabité en anglais britannique contemporain. *12^{ème} Colloque de l'ALOE*, Villetaneuse, 2-3 avril 2004.
- Bouzon, C. et Hirst, D.J. (submitted). Isochrony and prosodic structure. Testing phonological models of rhythm on a large corpus of spoken English. *Journal of Phonetics*.
- Brown, G., Anderson, A., Shillcock, R.C. et Yule, G. (1984). *Teaching Talk*. Cambridge: Cambridge University Press.
- Buxton, H. (1983). Temporal Predictability in the Perception of English Speech. In Cutler, A. and Ladd, D.R. (eds): *Prosody: Models And Measurements*, pp.111-121.
- Campbell, N. (1989). Syllable-level Duration Determination. *Proceedings from Eurospeech*, Paris, pp.698-701.
- Campbell, N. (1992a). *Multi-level Timing in Speech*. PhD Thesis, University of Sussex.
- Campbell, N. (1992b). Syllable Based Segmental Duration. In Bailly, G., Benoit, C. and Sawallis, T.R. (eds): *Talking Machines. Theories, Models, and Designs*, pp.211-224.
- Campbell, N. (1993). Automatic detection of prosodic boundaries in speech. *Speech Communication*, **13**, pp.343-354.
- Campbell, N. (1996). Speech timing in the SEC. In Knowles, G., Wichmann, A. and Alderson, P. (eds): *Working with Speech. Perspectives on Research into the*

- LANCASTER/IBM Spoken English Corpus. Learning About Language.* Great Britain : Longman Group Ltd., pp.214-236.
- Campbell, N. (2000). Timing in Speech: A Multi-Level Process. In Horne, M. (éd.): *Prosody: Theory and Experiment*. Kluwer Academic Publishers, Pays-Bas : Dordrecht, p. 281-334.
- Campbell, N. et Isard, D. (1991). Segment Durations in a Syllable Frame. In Carlson, R. (éd.): *Speech Synthesis and Phonetics*, pp.37-47.
- Carlson, R. (1991). Duration models in use. *ICPHS 91, 12^e Congrès International des Sciences Phonétiques*, Aix-en-Provence, France, pp.243-245.
- Carlson, R. et Granström, B. (1989). Modelling Duration for Different Text Materials. *Proceedings from Eurospeech*, Paris, pp.328-331.
- Catford, J.C. (1977). *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.
- Chen, M. (1970). Vowel length variation as a function of voicing of the consonant environment. *Phonetica*, **22**, pp.129-159.
- Chomsky, N. et Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.
- Classé, A. (1939). *The Rhythm of English Prose*. Oxford: Basil Blackwell.
- Clements, G.N. et Keyser, S.J. (1983). From CV Phonology : A Generative Theory of the Syllable. In : Goldsmith, J.A. (éd.): *Phonological Theory. The Essential Readings*. Malden, MA, USA : Blackwell Publ., pp.185-200 .
- Coker, C.H., Umeda, N. et Brownman, C.P. (1973). *Automatic Synthesis from Text*. IEEE Trans. Audio Electroacoust. AU-21, pp.293-297.
- Coleman, C. (1974). *A study of acoustical and perceptual attributes of isochrony*. PhD Thesis, Univ. Washington.
- Coleman, J. (1992). Synthesis by Rule without Segments or Rewrite-Rules. In Bailly, G., Benoit, C. and Sawallis, T.R. (eds): *Talking Machines. Theories, Models and Designs*, pp.43-60.
- Coleman, J. (1994). Polysyllabic Words in the Yorktalk Synthesis System. In Keating, P.A. (éd.): *Papers in Laboratory Phonology III. Phonological Structure and Phonetic Form*, pp.293-324.
- Collier, R. (1992). A Comment on the Prediction of Prosody. In Bailly, G., Benoit, C. and Sawallis, T.R. (eds): *Talking Machines. Theories, Models, and Designs*, pp.205-208.
- Cooper, W.E. et Danly, M. (1981). Segmental and Temporal Aspects of Utterance-Final Lengthening. *Phonetica*, **38**, pp.106-115.
- Cooper, W.E. et Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, USA: Harvard University Press.

- Cornulier, B. de (1978). Syllabe et suite de phonèmes en phonologie du français. In Cornulier, B. de et Dell, F. (eds): *Etudes de phonologie française*. Ed. CNRS : Marseille, pp.31-69.
- Couper-Kuhlen, E. (1986). *An Introduction to English Prosody*. Tübingen, RFA : Niemeyer.
- Cruttenden, A. (1997). *Gimson's Pronunciation of English. Fifth edition*. England: Arnold.
- Crystal, T.H. et House, A.S. (1982). Segmental duration in connected speech signals: Preliminary results. *Journal of the Acoustical Society of America*, **72**(3), pp.705-716.
- Crystal, T.H. et House, A.S. (1988a). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, **83**(4), pp.1553-1573.
- Crystal, T.H. et House, A.S. (1988b). Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America*, **83**(4), pp.1574-1585.
- Crystal, T.H. et House, A.S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, **88**(1), pp.101-112.
- Cummins, F. (1998). Some Lengthening Factors in English Speech Combine Additively at Most Rates. *Journal of the Acoustical Society of America*, **105**, pp.476-480.
- Cummins, F. et Port, R. (1998). Rhythmic Constraints on Stress Timing in English. In Walley, A.C. and Flege, J.E. (eds): *Journal of Phonetics*, **26**(2), pp.145-171.
- Cutler, A. (1984). Stress and Accent in Language Production and Understanding. In Gibbon, D. and Richter, H. (eds): *Intonation, Accent and Rhythm. Studies in Discourse Phonology*. Walter de Gruyter, Berlin, pp.77-90.
- Cutler, A. (1991). Linguistic Rhythm and Speech Segmentation. In Sundberg, J., Nord, L. and Carlson, R. (eds): *Music, Language, Speech and Brain*, pp.157-166.
- Cutler, A. (1992). Production and Perception of Word Boundaries. In Tohkura, Y., Vatikiotis-Bateson, E. and Sagisaka, Y. (eds): *Speech Perception, Production and Linguistic Structure*, pp.419-425.
- Cutler, A. et Isard, S.D. (1980). The Production of Prosody. In Butterworth, B. (éd.): *Language Production, Vol.1, Speech and Talk*, pp.245-270.
- Cutler, A., Mehler, J., Norris, D. et Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25**, pp.385-400.
- Dasher, R. et Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, **12**, pp.58-69.
- Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, **11**, pp.51-62.

- Dauer, R.M. (1987). Phonetic and phonological components of language rhythm. *Proceedings from the 11th International Congress of Phonetic Sciences*, 5, Tallinn: URSS.
- Delais-Roussarie, E. (2002). Constituer des corpus oraux : méthodes et outils. *Carnets de grammaire. Rapports internes de l'Équipe de Recherche en Syntaxe et Sémantique*, **10**, Toulouse-Le Mirail.
- Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics*, **4**, pp.183-198.
- Derwing, B.L. (1992). A 'Pause Break' Task for Eliciting Syllable Boundary Judgments from Literate and Illiterate Speakers: Preliminary Results for Five Diverse Languages. *Language and Speech*, **35**(2), pp.219-236.
- Deschamps, A. Duchet, J.-L. Fournier, J.-M. et O'Neil, M. (2004). *English Phonology and Graphophonemics*. France: Editions Ophrys.
- Deterding, D. (2001). The Measurement of Rhythm: A Comparison of Singapore and British English. *Journal of Phonetics*, **29**(2), pp.217-230.
- Di Cristo, A. (1999). Le cadre accentuel du français : essai de modélisation, première partie. *Langues*, **2**(3), pp.184-205.
- Di Cristo, A. (2000). La problématique de la prosodie dans l'étude de la parole dite spontanée. *Parole*, **15-16**, pp.189-250.
- Di Cristo, A. (2002). De la métrique et du rythme de la parole ordinaire : l'exemple du français. Bordas, E. (éd.) : *Le rythme de la prose*. Revue de sémio-linguistique des textes et discours, **16**, pp.25-43.
- Di Cristo, A. et Hirst, D.J. (1997). L'accentuation non emphatique en français : stratégies et paramètres. Perrot, J. (éd.) : *Polyphonie pour Ivan Fónagy*, L'Harmattan, Paris, pp.221-223.
- Di Cristo, P. et Hirst, D.J. (1997). Un procédé d'alignement automatique de transcriptions phonétiques sans apprentissage préalable. *4^o Congrès Français d'Acoustique*, 1, Marseille, 14-18 avril, France : SFA, Teknea.
- Dirksen, A. et Coleman, J.S. (1997). All Prosodic Speech Synthesis. In Van Santen, J.P.H., Sproat, R.W., Olive, J.P. and Hirschberg, J. (eds): *Progress in Speech Synthesis*, pp.91-108.
- Dogil, G. et Braun, G. (1988). *The PIVOT model of speech parsing*. Der österreichischen akademie der wissenschaften: Wien.
- Donovan, A. et Darwin, C.J. (1979). The Perceived Rhythm of Speech. *Proceedings from the Ninth International Congress of Phonetic Sciences*, Vol. 2, Copenhagen: Institute of Phonetics, pp.268-274.
- Duchet, J.L. (1994). *Code de l'anglais oral*. France : Editions Ophrys, 2ème édition.

- Durand, J. et Anderson, J.M. (1986). Dependency Phonology. In Durand, J. (éd.): *Dependency and Non-linear Phonology*, pp.1-54.
- Durand, J., Laks, B. et Lyche, C. (2002). Phonologie du français contemporain, usages, variétés et structure. *Bulletin PFC*, **1**, Toulouse-Le Mirail.
- Durand, J. et Pukli, M. (à paraître). The PAC project and its methodology. *Tribune Internationale des Langues Vivantes*.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F. et van der Vrecken, O. (1996). The MBROLA Project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *ICSLP'96*. Philadelphia.
- Eefting, W. (1991). *Timing in Talking. Tempo Variation in Production and its Role in Perception*. Utrecht, Pays-Bas.
- Emerard, F., Mortamet, L. et Cozannet, A. (1992). Prosodic Processing in a Text-to-Speech Synthesis System Using a Database and Learning Procedures. In Bailly, G., Benoit, C. and Sawallis, T.R. (eds): *Talking Machines. Theories, Models, and Designs*, pp.225-254.
- Eriksson, A. (1991). *Aspects of Swedish Speech Rhythm*. Doctoral dissertation, University of Göteborg: Suède.
- Eskénazi, M. (1993). Trends in Speaking Styles Research. *Europseech '93*, pp.501-508.
- Ewen, C. (1982). The internal structure of complex segments. In Van Der Hulst, H. et Smith, N. (eds): *The structure of phonological representations*, vol. 2., Dordrecht: Foris Publications, pp.27-67.
- Ewen, C.J. et van der Hulst, H. (2001). The Phonological Structure of Words. *Cambridge Textbooks in Linguistics*. Great Britain: Cambridge University Press.
- Fallows, D. (1981). Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics*, **17**, pp.179-392.
- Fant, G. (1991). Units of temporal organization, stress groups versus syllables and words. *Proceedings of the XIIth ICPHS*, vol.1, Université de Provence, Aix-en-Provence, pp.247-250.
- Fant, G., Kruckenberg, A. et Nord, L. (1991a). Durational correlates of stress in Swedish, French and English. *Journal of Phonetics*, **19**, pp.351-365.
- Fant, G., Kruckenberg, A. et Nord, L. (1991b). Language Specific Patterns of Prosodic and Segmental Structures in Swedish, French and English. *Proceedings of the XIIth ICPHS*, vol.4, Université de Provence, Aix-en-Provence, pp.118-121.
- Fant, G. et Kruckenberg, A. (1993). Towards an integrated view of stress correlates. *Working Papers*, **41**, ESCA Workshop on Prosody, Lund, Sweden, pp.42-45.
- Farinas, J. (2002). *Une modélisation automatique du rythme pour l'identification des langues*. Thèse de Doctorat, Université Toulouse III, Institut de Recherche en Informatique de Toulouse.

- Faure, G. (1975). *Manuel pratique d'anglais parlé*. France : Hachette.
- Faure, G., Hirst, D.J. et Chafcouloff, M. (1980). Rhythm in English: Isochronism, Pitch and Perceived Stress. In Waugh, L.R. and Schooneveld, C.H. (eds): *The Melody of Language. Intonation and Prosody*, pp.71-80.
- Fourakis, M. et Monahan, C.B. (1988). Effects of Metrical Foot Structure on Syllable Timing. *Language and speech*, **31**(3), pp.283-306.
- Fowler, C. A. (1977). *Timing Control in Speech Production*. Bloomington, USA: IULC Publications.
- Fox, A. (2000). *Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals*. Great Britain : Oxford University Press.
- Fraisse, P. (1974). *Psychologie Du Rythme*. Paris : PUF, Presses Universitaires de France.
- Fry, D.B. (1955). Duration and Intensity as Physical Correlates of linguistic Stress. *Journal of the Acoustical Society of America*, **27**, pp.765-768.
- Fry, D.B. (1958). Experiments in the perception of stress. *Language and Speech*, **1**, pp.126-152.
- Fudge, E.C. (1969). Syllables. In Goldsmith, J.A. (éd.) : *Phonological Theory. The Essential Readings*. Malden, MA, USA : Blackwell Publ., pp.370-391.
- Fudge, E. (1984). *English Word-Stress*. George Allen & Unwin, London.
- Fudge, E.C. (1987a). Branching Structure within the Syllable. *Journal of Linguistics*, **23**(2), pp.359-378.
- Fudge, E.C. (1987b). Review of Dependency and Non-linear Phonology, Durand, J.. *Journal of Linguistics*, **23**(2), pp.459-465.
- Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. Local, J., Ogden, R. et Temple, R. (eds): *Phonetic Interpretation. Papers in Laboratory Phonology VI*, pp.222-236.
- Giegerich, H.J. (1992). *English Phonology. An Introduction*. Cambridge, Great Britain: Cambridge University Press.
- Gimson, A.C. (1989). *An Introduction to the Pronunciation of English* (4th edition, edited by S. Ramasaran), London: Edward Arnold.
- Ginésy, M. (1995). *Mémento de phonétique anglaise*. France : Nathan.
- Goldsmith, J. (1990). *Autosegmental and metrical Phonology*. Oxford: Blackwell Publisher.
- Grabe, E. et Post, B. (2002). Intonational Variation in English. In Bel, B. and Marlien, I. (eds): *Proceedings of the Speech Prosody 2002 Conference*, April, 11-13th 2002, Aix-en-Provence: Laboratoire Parole et Langage, pp.343-346.

- Grabe, E. et Low, E.L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In Gussenhoven, C. and Warner, N. (eds.): *Laboratory Phonology 7. Phonology & Phonetics*. Allemagne : Mouton De Gruyter, pp.515-46.
- Greenberg, S., Chang, S. et Hitchcock, L. (2001). The Relation Between Stress Accent and Vocalic Identity in Spontaneous American English Discourse. *Proceedings from ISCA Workshop on Prosody in Speech Recognition and Understanding*, pp.51-56.
- Greenberg, S., Carvey, H. et Hitchcock, L. (2002). The Relation Between Stress Accent and Pronunciation Variation in Spontaneous American English Discourse. In Bel, B. et Marlien, I. (eds): *Proceedings of the Speech Prosody 2002 Conference*, 11-13 Avril. Aix-en-Provence: Laboratoire Parole et Langage, pp.351-354.
- Grosjean, F. et Collins, M. (1979). Breathing, pausing and reading. *Phonetica*, **36**, pp.98-114.
- Halliday, M.A.K. (1967). *Intonation and Grammar in British English*. Janua Linguarum, Series Practica, 48. La Haye : Mouton et Co.
- Halliday, M.A.K. (1970). *A Course in Spoken English: Intonation*. Great Britain: Oxford University Press.
- Hawkins, S., House, J., Huckvale, M., Local, J. et Ogden, R., (1998). ProSynth: An integrated prosodic approach to device-independent, natural-sounding speech synthesis. *Proceedings of International Conference on Spoken Language Proceeding*, pp.1707-1710.
- Hayes, B. (1986). Inalterability in CV Phonology. In Goldsmith, J.A. (éd.) : *Phonological Theory. The Essential Readings*. Malden, MA, USA : Blackwell Publ., pp.224-237.
- Hayes, B. (1995). *Metrical Stress Theory. Principles and Case Studies*. Chicago, IL, USA: The University of Chicago Press.
- Hill, D.R., Jassem, W. et Witten, I.H. (1978). A statistical approach to the problem of isochrony in spoken British English. *Man-Machine Systems Laboratory Report 78/26/5*, University of Calgary.
- Hill, D.R., Witten, I.H. et Jassem, W. (1978). Some results from a preliminary study of British English speech rhythm. *Research Report 78/26/5*. University of Calgary, Department of computer science.
- Hirst, D.J. (1983). Peak, boundary and cohesion characteristics of prosodic grouping. In House, D. and Touati, P. (eds): *Prosody, Working Papers 41*, ESCA Workshop on Prosody, Dept of Linguistics and Phonetics, Lund, Sweden, pp.32-37.
- Hirst, D.J. (1984). Prosodie et structures de données en phonologie. In Dell, F., Hirst, D.J. et Vergnaud, J.R. (eds): *Forme Sonore du Langage. Structure des Représentations en Phonologie*, pp.43-63.
- Hirst, D.J. (1998). Intonation in British English. In Hirst, D.J. and Di Cristo, A. (eds): *Intonation Systems. A Survey Of Twenty Languages*, pp.56-77.

- Hirst, D.J. et Di Cristo, A. (1984). French Intonation: A Parametric Approach. *Off-print, Die Neueren Sprachen*, **83**(5), pp.554-569.
- Hirst, D.J. et Espesser, R. (1993). Automatic Modelling of Fundamental Frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, **15**, pp.75-85.
- Hirst, D.J., Di Cristo, A. et Espesser, R. (2000). Levels of Representation and Levels of Analysis for the Description of Intonation Systems. In Horne, M. (éd.): *Prosody : Theory and Experiment. Text, Speech and Language Technology*, **14**. Pays-Bas : Kluwer Akademik Publishers, pp.51-87.
- Hitchcock, L. et Greenberg, S. (2001). Vowel Height is Intimately Associated with Stress Accent in Spontaneous American English Discourse. *Eurospeech 2001*, Aalborg.
- Hockett, C. (1965). *A Manual of Phonology*. The University of Chicago Press.
- Hoequist (1983a). Durational correlates of Linguistic Rhythm Categories. *Phonetica*, **40**(1), pp.19-31.
- Hoequist (1983b). Syllable Duration in Stress-, syllable- and Mora-timed Languages. *Phonetica*, **40**(3), pp.203-207.
- Hooper, J.B. (1972). The syllable in phonological theory. *Language*, **48**(3), pp.525-540.
- House, A.S. (1975). On Vowel Duration in English. In Fry, D.B. (éd.): *Acoustic Phonetics. A Course of Basic Readings*, pp.369-377.
- Huggins, A.W.F. (1975). On isochrony and syntax. In Fant, G. and Tatham, M.A. (eds): *Auditory analysis and perception of speech*. New York: Academic Press, pp.455-463.
- Jassem, W. (1949). In defense of broad transcription. *Le Maître Phonétique*, III/90, pp.23-24.
- Jassem, W. (1951). *Intonation in Conversational English (Educated Southern British)*. Polish Academy of Science, Warszawa.
- Jassem, W. (1952). Stress in Modern English. *Bulletin de la société polonaise de linguistique, fascicule XI*. Poland : Cracovie.
- Jassem, W. (1983). *The Phonology of Modern English*. Warszawa: Panstwowe Wydawnictwo Naukowe.
- Jassem, W. (1995). *Exercises in English Pronunciation*. Warszawa: Panstwowe Wydawnictwo Naukowe.
- Jassem, W. (1999). English Stress, Accent and Intonation Revisited. In *Speech and Language Technology Vol. 3, Papers, Reports and Technical Notes*. Poznan: Polish Phonetics Association, pp.33-50.
- Jassem, W. et Gibbon, D. (1980). Re-Defining English Accent and Stress. *Journal of the International Phonetic Association*, **10**(1-2), pp.2-16.

- Jassem, W., Hill, D.R. et Witten, I.H. (1984). Isochrony in English Speech: its Statistical Validity and Linguistic Relevance. In Gibbon, D. and Richter, H. (eds): *Intonation, Accent and Rhythm. Studies in Discourse Phonology*, pp.203-225.
- Jones, C. (1976). Some constraints on medial consonant clusters. *Language*, **52**, pp.121-130.
- Jones, D. (1918; 1972). *An Outline of English Phonetics*. Cambridge University Press, Cambridge.
- Jones, D. (1991). *English Pronouncing Dictionary*. England: Longman.
- Kahn, D. (1976). *Syllable-Based Generalizations in English Phonology*. Bloomington, IN, USA : IULC Publications.
- Kaiki, N., Takeda, K. et Sagisaka, Y. (1992). Linguistic Properties in the Control of Segmental Duration for Speech Synthesis. In Bailly, G., Benoit, C. and Sawallis, T.R. (eds) : *Talking Machines. Theories, Models, and Designs*, pp.255-263.
- Kaye, J.D. et Lowenstamm, J. (1984). De la syllabacité. In Dell, F., Hirst, D.J. et Vergnaud, J.-R. (eds) : *Forme Sonore Du Langage. Structure Des Représentations En Phonologie*, pp.123-159.
- Kingdon, R. (1958). *The Groundwork of English Intonation*. London : Longman.
- Kipp, A., Wesenick, M.-B. et Schiel, F. (1996). Automatic detection and segmentation of pronunciation variants in German speech corpora. *Proceedings of ICSLP 96, 4^e Int. Conf. on Spoken Lang.* USA : University of Delaware & Alfred I. du Pont Institute, pp.106-109.
- Klatt, D.H. (1973a). Interaction between Two Factors that Influence Vowel Duration. *Journal of the Acoustical Society of America*, **54**, pp.1102-1104.
- Klatt, D.H. (1973b). Duration characteristics of pre-stressed word-initial consonant clusters in English. *Technical report QPR 108*, MIT, Cambridge MA.
- Klatt, D.H. (1974). The duration of [s] in English words. *Journal of Speech Hearing Research*, **17**, pp.51-63.
- Klatt, D.H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, pp.129-140.
- Klatt, D.H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, pp.1208-1221.
- Klatt, D.H. (1979). Synthesis by Rule of Segmental Durations in English Sentences. In Lindblom, B. and Ohman, S.E.G. (eds): *Frontiers of Speech Communication Research*, pp.287-299.
- Klatt, D.H. (1982). A Strategy for the Perceptual Interpretation of Durational Cues in English Sentences. In *Working Papers vol.1, MIT Speech Communication*, USA:

- MIT, Massachusetts Institute of Technology Research Laboratory of Electronics, pp.83-91.
- Klatt, D.H. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, **82**, pp.737-793.
- Knowles, G. (1987). Patterns of Spoken English. An Introduction to English Phonetics. London: Longman.
- Kohler, K.J. (1991). Isochrony, units of rhythmic organization and speech rate. *ICPHS 91, 12^e Congrès International des Sciences Phonétiques*, Aix-en-Provence, France, pp.527-261.
- Kurylowicz, J. (1948). Contribution à la théorie de la syllabe. *Bulletin de la Société Polonaise de Linguistique*, **8**, pp.80-104.
- Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.
- Laver, J. (1994). *Principles of phonetics*. Cambridge Textbooks in Linguistics. Cambridge, Great Britain: Cambridge University Press.
- Lea, W.A. (1974). Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding. *Univac Report No. PX10791*. St Paul, Minnesota: Sperry Univac, DSD.
- Lehiste, I. (1970). *Suprasegmentals*. USA : The MIT Press.
- Lehiste, I. (1971). Temporal Organization of Spoken Language. In Hammerich, L.L., Jakobson, R. and Zwirner, E. (eds): *Form and Substance. Phonetic and Linguistic Papers Presented to Eli Fischer Jorgensen 11th February 1971*, pp.159-169.
- Lehiste, I. (1974). The Timing of Utterances and Linguistic Boundaries. In Lass, N. (ed.): *Speech and Hearing Science: Selected Readings*, pp.20-35.
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, **5**, pp.253-263.
- Lehiste, I. (1979). Temporal relations within speech units. In *Proceedings of the Ninth International Congress of Phonetic Sciences*, vol III, Copenhagen, pp.247-254.
- Lehiste, I. (1980). Interaction between Test Word Duration and Length of Utterance. In Waugh, L.R. and Schooneveld, C.H. (eds): *The Melody of Language. Intonation and Prosody*, pp.169-176.
- Levelt, W.J.M. (1989). *Speaking. From Intention to Articulation*. Cambridge, MA, USA: The MIT Press.
- Liberman, M. (1975). *The intonational system of English*. Doctoral dissertation, Massachusetts Institute of Technology.
- Liberman, M. et Prince, A.S. (1977). On Stress and Linguistic Rhythm. In Goldsmith, J.A. (éd.): *Phonological Theory. The Essential Readings*. USA : Blackwell Publ., pp.392-404.

- Lilly, R. et Viel, M. (1998). *La prononciation de l'anglais. Règles phonologiques et exercices de transcription*. Coll. Hachette Université.
- Lloyd, James A. (1940). *Speech Signals in Telephony*. London.
- Local, J. et Ogden, R. (1997). A Model of Timing for Nonsegmental Phonological Structure. In Van Santen, J.P.H., Sproat, R.W., Olive, J.P. and Hirschberg, J. (eds): *Progress in Speech Synthesis*, pp.109-121.
- Luce, P.A. et Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration and the C/V ratio in speech production. *Journal of the Acoustical Society of America*, 78, pp.1949-1957.
- Marcus, S.M. (1979). Perceptual Centers (P-centers). *9th International Congress of Phonetic Sciences*, Copenhagen, Danemark, p.238.
- Mehler, J., Segui, J. et Frauenfelder, U. (1981). The Role of the Syllable in Language Acquisition and Perception. In Myers, T., Laver, J. and Anderson, J. (eds) : *The Cognitive Representation of Speech*, pp.295-305.
- Meynadier, Y. (2001). La syllabe phonétique et phonologique : une introduction. *Travaux Interdisciplinaires du Laboratoire PArôle et Langage d'Aix-en-Provence*, 20, pp.91-148.
- Möbius, B. et Santen, van J.P.H. (2000). Phonetically Motivated Modeling of Prosody. *Prosody 2000*, Faculty of Modern Languages and Literature, Adam Mickiewicz University, Poland, pp.161-166.
- Morton, J., Marcus, S. et Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, 83, pp.405-408.
- Mullooly, R. (2003). An electromagnetic articulography study of resyllabification of rhotic consonants in English. *Proceedings from the XVth ICPHS*, Barcelona, pp.2265-2268.
- Nakatani, L.H., O'Connor, K.D. et Aston, C.H. (1981). Prosodic Aspects of American English Speech Rhythm. *Phonetica*, 38, pp.9-20.
- Nespor, M. (1990). On the rhythm parameter in phonology. In Roca, I.M. (éd.): *Logical Issues in Language Acquisition*. Dordrecht: Foris, pp.157-175.
- Nespor, M. et Vogel, I. (1986). Prosodic Phonology. *Studies in Generative Grammar*, 28. Dordrecht, Pays-Bas : Foris Publ.
- Nocéra, P., Linares, G., Massonié, D. et Lefort, L. (2000). Phoneme Lattice Based A* Search Algorithm for Speech Recognition. *Proceedings of TSD*, pp.300-308.
- Nooteboom, S.G. (1991). Some observations on the temporal organisation and rhythm of speech. *ICPHS 91, 12° Congrès International des Sciences Phonétiques*. Aix-en-Provence, France, pp.228-237.
- O'Connor, J.D. (1965). The perception of time intervals. *Progress Report (Phonetics Laboratory, University College London)*, 2, pp.11-15.

- O'Connor, J.D. (1967; 1980). *Better English Pronunciation*. Cambridge: Cambridge University Press.
- O'Connor, J. D. et Arnold, G. F. (1973). *Intonation of Colloquial English. Second edition*. London: Longman.
- O'Connor, J.D. et Trim, J.L.M. (1953). Vowel, consonant, and syllable – a phonological definition. *Word*, **9**, pp.103-122.
- Ogden, R., Local, J. et Carter, P., (1999). Temporal Interpretation in ProSynth, a Prosodic Speech Synthesis System. *Proc. XIVth International Congress of Phonetic Sciences*, pp.1059-1062.
- Oller, D.K. (1973). The effect of position in utterance on speech segment duration. *Journal of the Acoustical Society of America*, **54**, pp.1235-1247.
- Peterson, G.E. et Lehiste, I. (1960). Duration of Syllable Nuclei in English. In Lehiste, I. (éd.): *Readings in Acoustic Phonetics*, pp.191-201.
- Pfitzinger, H.R. (1996). Two approaches to speech rate estimation. In *Proceedings of SST' 96*, Adelaide, pp.421-426.
- Pfitzinger, H.R. (1999). Local speech rate perception in German speech. *Proceedings from ICPHS*, San Francisco, pp.893-896.
- Pinheiro J.C. et D.M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New-York: Springer.
- Pike, K.L. (1945). *The Intonation of American English*. USA: The University of Michigan Press.
- Port, R., Al-ani, S. et Maeda, S. (1980). Temporal Compensation and Universal Phonetics. *Phonetica*, **37**, pp.235-252.
- Pulgram, E. (1970). *Syllable, Word, Nexus, Cursus*. The Hague: Mouton.
- Rakerd, B., Sennett, W. et Fowler, C.A. (1992). Domain Final Lengthening and Foot Level Shortening in Spoken English. *Status Report on SPEECH RESEARCH*, Haskins Laboratories, New Haven: USA, pp.55-63.
- Ramus, F. (1999). *Rythme des langues et acquisition du langage*. Thèse de doctorat, EHESS, Paris.
- Ramus, F., Nespor, M., et Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**(3), pp.265-292.
- Rice, C. (1992). *Binarity and Ternarity in Metrical Theory: Parametric Extensions*. Doctoral dissertation, University of Texas, Austin.
- Riley, M.D. (1992). Tree Based Modelling of Segmental Durations. In Bailly, G. and Sawallis, T.R. (eds): *Talking Machines. Theories, Models, And Designs*, pp.265-273.

- Roach, P. (1982). On the Distinction between "Stress Timed" and "Syllable Timed" Languages. In Crystal, D. (éd.): *Linguistic Controversies. Essays in Linguistic Theory and Practice, In Honour Of F.R. Palmer*, pp.73-79.
- Roach, P. (1991). *English Phonetics and Phonology: A Practical Course*. Cambridge University Press, second edition.
- Roach, P. (1994). Conversion between prosodic transcription systems: "Standard British" and ToBI. *Speech Communication*, 15, pp. 91-99.
- Roach, P. (1998). Some languages are spoken more quickly than others. In Bauer, L. and Trudgill, P. (eds): *Language Myths*. Penguin, pp.150-158.
- Roach, P. (1999). Studying rhythm and timing in English speech: scientific curiosity, or a classroom necessity? In *The Fifth International Conference on Computers in Medicine*, Lodz, Poland. The College of Computer Science.
- Santen, van J.P.H. (1992a). Deriving Text-to-Speech Durations from Natural Speech. In Bailly, G., Benoit, C. and Sawallis, T.R. (eds): *Talking Machines. Theories, Models, and Designs*, pp.275-285.
- Santen, van J.P.H. (1992b). Contextual effects on vowel duration. *Speech Communication*, 11, pp.513-546.
- Santen, van J.P.H. (1997). Segmental Duration and Speech Timing. In Sagisaka, Y., Campbell, N. and Higuchi, N. (eds): *Computing Prosody. Computational Models For Processing Spontaneous Speech*, pp.225-249.
- Santen, van J.P.H. (1998). Timing. In Sproat, R. (éd.): *Multilingual Text-to-Speech Synthesis. The Bell Labs Approach*. Dordrecht: Kluwer Academic Publishers, pp.115-139.
- Schiller, N.O., Meyer, A.S. et Levelt, W.J.M. (1997). The Syllabic Structure of Spoken Words: Evidence from the Syllabification of Intervocalic Consonants. *Language and Speech*, 40(2), pp.103-140.
- Scott, D.R., Isard, S.D. et de Boysson-Bardies, B. (1985). Perceptual isochrony in English and French. *Journal of Phonetics*, 13, pp.155-162.
- Séguinot, C. (1979). A Phonostylistic Study of Rhythm in English. In Léon, P. et Rossi, M. (eds): *Problèmes de prosodie, Vol. 2: expérimentations, modèles et fonctions*, pp.179-189.
- Selkirk, E.O. (1982). The Syllable. In Van Der Hulst, H. et Smith, N. (eds): *The Structure Of Phonological Representations, Part 2*, pp.337-383.
- Selkirk, E. (1984a). *Phonology and Syntax: the Relation between Sound and Structure*. Cambridge, Massachusetts: MIT Press.
- Selkirk, E.O. (1984b). On the Major Class Features and Syllable Theory. In Lieber, R. (éd.): *Consonant Gradation In Fula: An Autosegmental Approach*, pp.107-136.

- Selkirk, E. (1986). On The Nature on Prosodic Constituency: Comments on Beckman and Edwards's Paper. In Kingston, J. and Beckman, M.E. (eds): *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech*, pp.179-200.
- Shen, Y. et Peterson, G.G. (1962). Isochronism in English. *University of Buffalo Studies in Linguistics, Occasional Papers*, **9**, pp.1-36.
- Sluijter, A.M.C. et van Heuven, V.J. (1993). Perceptual cues of linguistic stress: intensity revisited. *Working Papers*, **41**. ESCA Workshop on Prosody, Lund, Sweden, pp.246-249.
- Tatham, M. et Morton, K. (2002). Computational modelling of speech production, English rhythm. In Braun, A. and Masthoff, H.R. (eds): *Phonetics and its Applications, Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday*. Franz Steiner Verlag, pp.383-405.
- Trager, G.L. et Smith, H.L.J. (1957). An outline of English structure. *Studies in Linguistics: Occasional Papers* 3. American Council of Learned Societies, Washington D.C. (5th printing).
- Treiman, R. (1989). The Internal Structure Of The Syllable. In Carlson, G. N. and Tanenhaus, M.K. (eds): *Linguistic Structure In Language Processing*, pp.27-52.
- Treiman, R. et Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **14**, pp.145-152.
- Treiman, E., Straub, K. et Lavery, P. (1994). Syllabification of Bisyllabic Nonwords: Evidence from Short Term Memory Errors. *Language and Speech*, **37**(1), pp.45-60.
- Treiman, R. et Kessler, B. (1995). In Defense Of An Onset Rime Syllable Structure For English. *Language and Speech*, **38**(2), pp.127-142.
- Troubetzkoy, N.S. (1939). *Grundzüge der Phonologie*. Prague. Travaux du Cercle Linguistique de Prague. Traduction française de Cantineau, J. (1949). *Principes de phonologie*. Paris : Klincksieck.
- Turk, A. (1994). Articulatory Phonetic Clues to Syllable Affiliation: Gestural Characteristics of Bilabial Stops (syllables). In Keating, P.A. (éd.): *Papers In Laboratory Phonology III. Phonological Structure And Phonetic Form*. pp.107-135.
- Turk, A. et Sawush, J.R. (1997). The Domain of Accentual Lengthening in American English. *Journal of Phonetics*, **25**(1), pp.25-41.
- Turk, A.E. et White, L. (1999). Structural Influences on Accentual Lengthening in English. *Journal of Phonetics*, **27**(2), p. 171-206.
- Turk, A.E. et Shattuck-Hufnagel, S. (2000). Word-Boundary-Related Duration Patterns in English. *Journal of Phonetics*, **28**(4), pp.397-440.
- Uldall, E.T. (1971). Isochronous Stresses in RP. In Hammerich, L.L., Jakobson, R. and Zwirner, E. (eds): *Form and Substance. Phonetic and Linguistic Papers Presented to Eli Fischer Jorgensen 11th February 1971*, pp.205-210.

- Umeda, N. (1975). Vowel Duration in American English. *Journal of the Acoustical Society of America*, **58**, pp.434-444.
- Viel, M. (1995). *La phonétique de l'anglais*. Coll. Que sais-je? Paris : Presses universitaires de France.
- Vogel, I., Bunnell, H.T. et Hoskins, S. (1995). The Phonology and Phonetics of the Rhythm Rule. In Connell, B. and Arvaniti, A. (eds): *Papers in Laboratory Phonology IV. Phonology and Phonetic Evidence*, pp.111-127.
- Wells, J.C. (1990a). *Longman Pronunciation Dictionary*. First edition. Harlow: Longman.
- Wells, J.C. (1990b). Syllabification and allophony. In Ramsaran, S. (éd.): *Studies in the Pronunciation of English*. London: Routledge, pp.76-86.
- Wells, J.C. (2000). *Longman Pronunciation Dictionary*. Second edition. Harlow: Pearson Education Limited.
- Wenk, B.J. et Wioland, F. (1982). Is French Really Syllable Timed? *Journal of Phonetics*, **10**, pp.193-215.
- Wenk, B.J. (1983). Effets de rythme dans le français parlé. *Recherches sur le français parlé*, **5**, Publications de l'Université de Provence, Aix-en-Provence, pp.147-162.
- Wood, S. (1973). What happens to vowels and consonants when we speak faster? *Working Papers 9*, Phonetics Laboratory Lund University, Lund, pp.8-39.
- Zellner-Keller, B. (1996). Structures temporelles et structures prosodiques en français lu. *Revue française de linguistique appliquée*, **1**, pp.7-23.
- Zellner-Keller, B. (1998). *Caractérisation et prédiction du débit de parole en français, une étude de cas*. Thèse de doctorat, université de Lausanne.
- Zhang, G. (1996). *Foot-timing and word-timing in English*. PhD dissertation, University of Delaware.

INDEX

INDEX DES MOTS-CLES

A

Accent lexical
51, 98, 135, 136, 137, 139, 141, 194,
204, 205, 206, 274

Aix-MARSEC
5, 21, 22, 33, 34, 172, 178, 187, 188,
190, 195, 202, 204, 206, 207, 208,
214, 220, 221, 222, 223, 225, 260,
261, 262, 263, 264, 265, 267, 269,
298, 303, 304, 305, 310, 311, 334,
429

Alignement
21, 22, 24, 52, 54, 178, 179, 187,
190, 194, 196, 198, 200, 201, 202,
206, 207, 208, 209, 213, 214, 218,
222, 232, 277, 304, 310, 311, 312,
315

Ambisyllabicit 
20, 21, 22, 23, 25, 31, 33, 34, 39, 65,
67, 68, 69, 80, 81, 82, 85, 86, 87, 88,
89, 90, 91, 94, 95, 96, 97, 98, 99,
100, 172, 210, 212, 213, 214, 260,
261, 262, 263, 264, 265, 266, 267,
268, 273, 274, 278, 279, 280, 282,
283, 284, 285, 286, 287, 288, 289,
290, 292, 293, 295, 298, 299, 303,
305, 306, 333, 414, 416, 417, 418,
419, 423, 424, 425

Anacrouse
60, 110, 111, 112, 113, 115, 117,
120, 121, 125, 128, 129, 133, 142,
172, 202, 206, 207, 209, 210, 211,
214, 216, 217, 218, 219, 220, 222,
225, 226, 234, 235, 236, 239, 241,
251, 257, 258, 269, 297, 298, 303,
304, 338, 339, 344, 345, 351, 352,

353, 356, 357, 358, 363, 364, 369,
370, 374, 375, 380, 381

Attaque
22, 34, 42, 44, 49, 50, 54, 56, 57, 72,
73, 76, 77, 78, 79, 80, 81, 82, 84, 85,
86, 87, 88, 89, 91, 94, 95, 96, 98, 99,
100, 126, 138, 148, 150, 153, 198,
202, 204, 207, 210, 213, 214, 260,
261, 263, 264, 265, 266, 267, 269,
273, 278, 279, 281, 283, 284, 289,
290, 292, 293, 294, 295, 297, 298,
299, 305, 306, 414, 415, 417, 418

C

Cat gorisation des langues 144, 157

Coda
25, 54, 56, 73, 76, 77, 78, 79, 80, 81,
82, 85, 86, 88, 89, 90, 91, 94, 95, 96,
97, 98, 100, 202, 203, 204, 207, 210,
260, 263, 265, 266, 267, 268, 269,
273, 278, 279, 280, 281, 284, 289,
290, 291, 292, 293, 295, 298, 299,
306, 333, 414, 415, 417, 418, 424,
425, 427, 430

Contraintes de cooccurrence
31, 78, 203

Corpus
21, 33, 46, 54, 83, 148, 149, 152,
153, 168, 170, 178, 179, 180, 181,
182, 183, 184, 185, 187, 188, 189,
190, 191, 192, 194, 195, 196, 199,
200, 201, 204, 205, 207, 208, 209,
214, 221, 232, 261, 267, 279, 285,
287, 289, 292, 299, 304, 306, 310,
312, 315, 429

I

Isochronie.....
 24, 32, 33, 34, 43, 48, 59, 60, 109,
 110, 111, 113, 114, 115, 117, 121,
 122, 133, 142, 143, 144, 145, 146,
 147, 148, 149, 150, 151, 152, 153,
 154, 155, 156, 157, 158, 160, 161,
 166, 171, 172, 210, 218, 219, 220,
 222, 224, 225, 226, 227, 228, 229,
 230, 231, 233, 238, 247, 257, 303,
 313, 318, 319, 324

L

Langues à rythme accentuel.....
 32, 43, 136, 146, 149, 157, 158, 160,
 162, 163, 164, 168, 170

Langues à rythme syllabique
 43, 149, 157, 158, 160, 164, 167, 170

M

MARSEC.....
 21, 33, 173, 178, 187, 189, 206

Modèle segmental
 31, 44, 45, 46, 47, 55, 64, 303

Modèle supra-syllabique.....
 20, 32, 33, 44, 59, 60, 103, 118, 217,
 264, 303

Modèle syllabique.....
 31, 44, 57, 64, 69, 261, 303, 305

N

NRU
 60, 110, 115, 142, 202, 206, 207,
 209, 214

O

Organisation temporelle.....
 30, 31, 32, 34, 39, 40, 41, 42, 43, 47,
 48, 49, 50, 52, 53, 57, 59, 60, 61, 64,
 65, 69, 86, 94, 100, 104, 105, 109,
 113, 121, 122, 127, 129, 132, 134,
 142, 172, 303, 307

P

Parole authentique.....
 34, 100, 180, 185, 188, 195, 207,
 214, 261, 299, 305

Phonétisation.....
 178, 191, 192, 193, 195, 196, 199,
 200, 201, 207, 429

Phonotactique.....
 57, 68, 70, 73, 76, 81, 82, 85, 86, 88,
 89, 94, 97, 100, 169, 195, 197, 198,
 203, 263, 290, 293, 294, 306

Pied
 22, 24, 25, 42, 43, 44, 53, 54, 55, 58,
 59, 73, 87, 102, 105, 107, 108, 109,
 111, 115, 116, 117, 118, 119, 120,
 121, 122, 123, 124, 125, 128, 129,
 130, 131, 132, 133, 134, 142, 145,
 146, 148, 149, 150, 160, 163, 165,
 172, 202, 206, 207, 209, 210, 211,
 212, 214, 216, 217, 218, 219, 220,
 222, 225, 226, 227, 229, 230, 231,
 234, 235, 236, 237, 238, 239, 240,
 241, 242, 243, 246, 248, 249, 250,
 251, 253, 254, 255, 256, 257, 258,
 269, 297, 298, 303, 304, 336, 337,
 341, 343, 348, 350, 355, 360, 361,
 367, 372, 373, 377, 378, 379, 386,
 392, 393, 400, 401, 407

Principe d'attaque maximale.....68, 82

Proéminence.....
71, 76, 102, 105, 107, 109, 131, 135,
136, 137, 138, 139, 140, 141, 164

Prosodie.....
21, 29, 30, 33, 39, 40, 43, 48, 49, 50,
53, 57, 58, 82, 99, 100, 107, 108,
115, 127, 172, 187, 188, 189, 196,
204, 205, 206, 208, 216, 218, 220,
229, 258, 315, 326

R

Règles d'élision.....
178, 191, 196, 197, 199, 200, 201

Resyllabification.....
68, 80, 81, 84, 90, 91, 94, 204, 214,
269, 322

Rime.....
44, 50, 51, 54, 56, 57, 68, 70, 76, 78,
79, 80, 81, 82, 88, 95, 99, 261

S

SEC.....21, 188, 189, 194

Syllabification.....
31, 33, 34, 38, 39, 58, 64, 68, 69, 80,
81, 82, 84, 86, 87, 88, 89, 91, 92, 93,
94, 96, 99, 100, 172, 204, 261, 264,
265, 269, 273, 290, 293, 294, 295,
296, 297, 298, 305, 306, 307, 316

T

TRU.....30, 110, 112, 298

U

Unité rythmique.....
30, 31, 32, 33, 34, 38, 39, 44, 59, 60,
65, 100, 102, 103, 106, 107, 108,
109, 110, 111, 112, 113, 115, 117,
118, 119, 120, 121, 122, 127, 129,
130, 133, 134, 141, 142, 145, 149,
150, 151, 157, 161, 172, 178, 202,
204, 205, 206, 207, 209, 210, 211,
212, 213, 214, 217, 218, 219, 251,
257, 297, 298, 303, 304

ANNEXES

TABLE DES ANNEXES

1.	ANNEXE 1. CORRESPONDANCE SYMBOLES SAMPA/ IPA	335
2.	ANNEXES 2. ISOCHRONIE ET CONSTITUANCE	336
2.1.	Isochronie stricte.....	336
2.1.1.	Passage C01 : Tableau 7 (page 222).....	336
2.1.2.	Totalité des données : Tableau 8 (page 224)	342
2.2.	Isochronie faible	349
2.2.1.	Passage C01	349
2.2.2.	Totalité des données.....	366
2.3.	Effet du paramètre accentuel	384
2.3.1.	Passage C01	384
2.3.2.	Totalité des données.....	399
3.	ANNEXES 3. SYLLABIFICATION EN ANGLAIS BRITANNIQUE.....	414
3.1.	Résultats du test t (passage C01)	414
3.2.	Résultats du test t (totalité des données).....	417
3.3.	Fonctionnement du test de production.....	420
3.4.	Phrases de l'expérience en production et différentes phases de lecture	421
3.4.1.	Première partie de l'expérience	421
3.4.2.	Deuxième partie de l'expérience	422
3.5.	Matrices des effectifs	423
3.5.1.	Matrices des effectifs pour les regroupements ambisyllabiques	423
3.5.2.	Matrices des effectifs pour les regroupements codaïques	424
3.6.	Résultats des χ^2	425
3.6.1.	Regroupements ambisyllabiques	425
3.6.2.	Regroupements codaïques	427

CONTENU DES CD-ROM.....	429
CD-ROM 1. LA BASE DE DONNEES AIX-MARSEC.....	429
📁 Audio.....	429
📁 Dictionnaires	429
📁 Sripis.....	429
📁 TextGrid	429
📄 données.zip.....	429
📄 thèse.pdf	429
CD-ROM 2. EXPERIENCE EN PRODUCTION	430
📁 Données	430
📁 Enregistrements.....	430
📁 Présentations.....	430
📄 get_pauses_v2.6.praat (script de détection des pauses)	430
CD-ROM 3. EXPERIENCE EN PRODUCTION (SUITE)	431
📁 Enregistrements.....	431

1. Annexe 1. Correspondance symboles SAMPA/ IPA

Consonnes

p	p	dZ	dʒ	S	ʃ	l	l
b	b	f	f	Z	ʒ	ʍ	ɮ
t	t	v	v	h	h	w	w
d	d	T	θ	m	m	j	j
k	k	D	ð	n	n		
g	g	s	s	N	ŋ		
tS	tʃ	z	z	r	r		

Voyelles

I	ɪ	U	ʊ	OI	ɔɪ	A:	ɑ:
e	e	@	ə	u:	u:	O:	ɔ:
{	æ	i:	i:	@U	əʊ	I@	ɪə
Q	ɒ	eI	eɪ	aU	aʊ	e@	eə
V	ʌ	aI	aɪ	3:	ɜ:	U@	ʊə

2. Annexes 2. Isochronie et constituance

2.1. Isochronie stricte

2.1.1. Passage C01 : Tableau 7 (page 227)

Durée de la syllabe ('ds') en fonction de sa complexité ('nps')

```
> lm(C01syllfilt$ds~ C01syllfilt$nps)
```

Call:

```
lm(formula = C01syllfilt$ds ~ C01syllfilt$nps)
```

Residuals:

Min	1Q	Median	3Q	Max
-245.06	-50.88	-11.33	34.40	604.94

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.153	2.766	2.225	0.0261 *
C01syllfilt\$nps	74.726	1.051	71.104	<2e-16 ***

Residual standard error: 75.7 on 6785 degrees of freedom

Multiple R-Squared: 0.427, Adjusted R-squared: 0.4269

F-statistic: 5056 on 1 and 6785 DF, p-value: < 2.2e-16

Durée du pied ('df') en fonction du nombre de phonèmes dans le pied ('npf')

```
> lm(C01nrufilt$df~ C01nrufilt$npf)
```

Call:

```
lm(formula = C01nrufilt$df ~ C01nrufilt$npf)
```

Residuals:

Min	1Q	Median	3Q	Max
-296.35	-69.99	-10.22	61.74	1054.47

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	125.8099	4.6963	26.79	<2e-16 ***
C01nrufilt\$npf	54.6358	0.7481	73.03	<2e-16 ***

Residual standard error: 102.5 on 2918 degrees of freedom

Multiple R-Squared: 0.6464, Adjusted R-squared: 0.6462

F-statistic: 5333 on 1 and 2918 DF, p-value: < 2.2e-16

Durée du pied ('df') en fonction du nombre de syllabes dans le pied ('nsf')

```
> lm(C01nrufilt$df~ C01nrufilt$nsf)
```

Call:

```
lm(formula = C01nrufilt$df ~ C01nrufilt$nsf)
```

Residuals:

Min	1Q	Median	3Q	Max
-286.69	-86.69	-12.10	77.72	1034.49

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	175.867	5.317	33.07	<2e-16 ***
C01nrufilt\$nsf	115.411	2.110	54.69	<2e-16 ***

Residual standard error: 121.1 on 2918 degrees of freedom

Multiple R-Squared: 0.5061, Adjusted R-squared: 0.506

F-statistic: 2991 on 1 and 2918 DF, p-value: < 2.2e-16

Durée de la NRU ('dr') en fonction du nombre de phonèmes dans la NRU ('npr')

```
> lm(C01nrufilt$dr[C01nrufilt$ru=="R"]~ C01nrufilt$npr[C01nrufilt$ru=="R"])
```

Call:

```
lm(formula=C01nrufilt$dr[C01nrufilt$ru=="R"]~C01nrufilt$npr[C01nrufilt$ru=="R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-275.700	-66.700	-9.755	61.245	574.300

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	129.5910	4.5730	28.34	<2e-16 ***
C01nrufilt\$npr[C01nrufilt\$ru == "R"]	51.5274	0.9252	55.70	<2e-16 ***

Residual standard error: 95.29 on 2919 degrees of freedom

Multiple R-Squared: 0.5152, Adjusted R-squared: 0.515

F-statistic: 3102 on 1 and 2919 DF, p-value: < 2.2e-16

Durée de la NRU ('dr') en fonction du nombre de syllabes dans la NRU ('nsr')

```
> lm(C01nrufilt$dr[C01nrufilt$ru=="R"]~ C01nrufilt$nsr[C01nrufilt$ru=="R"])
```

Call:

```
lm(formula=C01nrufilt$dr[C01nrufilt$ru=="R"]~C01nrufilt$nsr[C01nrufilt$ru=="R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-273.31	-83.31	-13.31	75.69	616.78

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	193.118	5.153	37.47	<2e-16 ***
C01nrufilt\$nsr[C01nrufilt\$ru == "R"]	100.097	2.747	36.43	<2e-16 ***

Residual standard error: 113.5 on 2919 degrees of freedom
 Multiple R-Squared: 0.3126, Adjusted R-squared: 0.3124
 F-statistic: 1327 on 1 and 2919 DF, p-value: < 2.2e-16

Durée de l'anacrouse ('dr') en fonction du nombre de phonèmes dans l'anacrouse ('npr')

```
> lm(C01anafilt$dr[C01anafilt$ru=="A"]~ C01anafilt$npr[C01anafilt$ru=="A"])
Call:
lm(formula=C01anafilt$dr[C01anafilt$ru=="A"]~C01anafilt$npr[C01anafilt$ru=="A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-182.810	-36.879	-5.862	32.121	308.630

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	27.896	3.743	7.453	1.92e-13 ***
C01anafilt\$npr[C01anafilt\$ru == "A"]	54.491	1.201	45.386	< 2e-16 ***

Residual standard error: 54.39 on 1042 degrees of freedom
 Multiple R-Squared: 0.6641, Adjusted R-squared: 0.6638
 F-statistic: 2060 on 1 and 1042 DF, p-value: < 2.2e-16

Durée de l'anacrouse ('dr') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

```
> lm(C01anafilt$dr[C01anafilt$ru=="A"]~ C01anafilt$nsr[C01anafilt$ru=="A"])
Call:
lm(formula=C01anafilt$dr[C01anafilt$ru=="A"]~C01anafilt$nsr[C01anafilt$ru=="A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-153.62	-45.41	-6.21	34.79	444.59

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	16.007	4.826	3.317	0.000943 ***
C01anafilt\$nsr[C01anafilt\$ru == "A"]	119.204	3.229	36.921	< 2e-16 ***

Residual standard error: 61.77 on 1042 degrees of freedom
 Multiple R-Squared: 0.5668, Adjusted R-squared: 0.5664
 F-statistic: 1363 on 1 and 1042 DF, p-value: < 2.2e-16

Durée de l'anacrouse initiale ('dr') en fonction du nombre de phonèmes dans l'anacrouse initiale ('npr')

```
> S<-lm(C01sfilt$dr[C01sfilt$ru=="S"]~ C01sfilt$npr[C01sfilt$ru=="S"])
```

Call:

```
lm(formula = C01sfilt$dr[C01sfilt$ru == "S"] ~ C01sfilt$npr[C01sfilt$ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-117.702	-35.335	-5.335	30.121	230.121

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	25.424	6.363	3.995	7.92e-05 ***
C01sfilt\$npr[C01sfilt\$ru == "S"]	54.456	1.867	29.164	< 2e-16 ***

Residual standard error: 53.91 on 340 degrees of freedom

Multiple R-Squared: 0.7144, Adjusted R-squared: 0.7136

F-statistic: 850.6 on 1 and 340 DF, p-value: < 2.2e-16

Durée de l'anacrouse initiale ('dr') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

```
> S<-lm(C01sfilt$dr[C01sfilt$ru=="S"]~ C01sfilt$nsr[C01sfilt$ru=="S"])
```

Call:

```
lm(formula = C01sfilt$dr[C01sfilt$ru == "S"] ~ C01sfilt$nsr[C01sfilt$ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-137.16	-43.71	-7.16	32.84	261.84

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.265	8.053	1.275	0.203
C01sfilt\$nsr[C01sfilt\$ru == "S"]	123.447	5.041	24.490	< 2e-16 ***

Residual standard error: 60.68 on 340 degrees of freedom

Multiple R-Squared: 0.6382, Adjusted R-squared: 0.6371

F-statistic: 599.8 on 1 and 340 DF, p-value: < 2.2e-16

Durée du mot ('dw') en fonction du nombre de phonèmes dans le mot ('npw')

```
> lm(C01wdfilt$dw~ C01wdfilt$npw)
```

Call:

```
lm(formula = C01wdfilt$dw ~ C01wdfilt$npw)
```

Residuals:

Min	1Q	Median	3Q	Max
-321.23	-65.67	-15.67	55.42	602.03

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	41.3656	2.9564	13.99	<2e-16 ***
C01wdfilt\$npw	66.6512	0.5983	111.40	<2e-16 ***

Residual standard error: 95.99 on 3983 degrees of freedom
 Multiple R-Squared: 0.757, Adjusted R-squared: 0.757
 F-statistic: 1.241e+04 on 1 and 3983 DF, p-value: < 2.2e-16

Durée du mot ('dw') en fonction du nombre de syllabes dans le mot ('nsw')

```
> lm(C01wdfilt$dw ~ C01wdfilt$nsw)
```

Call:

```
lm(formula = C01wdfilt$dw ~ C01wdfilt$nsw)
```

Residuals:

Min	1Q	Median	3Q	Max
-289.34	-92.50	-22.50	69.79	697.50

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	55.795	3.736	14.93	<2e-16 ***
aC01wdfilt\$nsw	156.709	1.892	82.84	<2e-16 ***

Residual standard error: 118 on 3983 degrees of freedom
 Multiple R-Squared: 0.6328, Adjusted R-squared: 0.6327
 F-statistic: 6863 on 1 and 3983 DF, p-value: < 2.2e-16

Durée de l'unité intonative ('di') en fonction du nombre de phonèmes dans l'unité intonative ('npi')

```
> lm(C01iufilt$di ~ C01iufilt$npi)
```

Call:

```
lm(formula = C01iufilt$di ~ C01iufilt$npi)
```

Residuals:

Min	1Q	Median	3Q	Max
-722.1	-303.3	-161.2	116.6	2326.8

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	424.462	34.019	12.48	<2e-16 ***
C01iufilt\$npi	65.099	1.995	32.64	<2e-16 ***

Residual standard error: 479.9 on 965 degrees of freedom

Multiple R-Squared: 0.5247, Adjusted R-squared: 0.5242
F-statistic: 1065 on 1 and 965 DF, p-value: < 2.2e-16

Durée de l'unité intonative ('di') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

```
> lm(C01iufilt$di~ C01iufilt$nsi)
Call:
lm(formula = C01iufilt$di ~ C01iufilt$nsi)
```

Residuals:

Min	1Q	Median	3Q	Max
-743.5	-310.3	-149.9	136.4	2379.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	448.255	34.189	13.11	<2e-16 ***
C01iufilt\$nsi	156.460	4.925	31.77	<2e-16 ***

Residual standard error: 486.7 on 965 degrees of freedom
Multiple R-Squared: 0.5112, Adjusted R-squared: 0.5107
F-statistic: 1009 on 1 and 965 DF, p-value: < 2.2e-16

Durée de l'unité intonative ('di') en fonction du nombre de pieds dans l'unité intonative ('nfi')

```
> lm(C01iufilt$di~ C01iufilt$nfi)
Call:
lm(formula = C01iufilt$di ~ C01iufilt$nfi)
```

Residuals:

Min	1Q	Median	3Q	Max
-1056.03	-316.91	-82.78	201.64	2385.17

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	375.74	37.29	10.08	<2e-16 ***
C01iufilt\$nfi	367.05	11.93	30.78	<2e-16 ***

Residual standard error: 494.5 on 965 degrees of freedom
Multiple R-Squared: 0.4954, Adjusted R-squared: 0.4949
F-statistic: 947.3 on 1 and 965 DF, p-value: < 2.2e-16

Durée de l'unité intonative ('*di*') en fonction du nombre de mots dans l'unité intonative ('*nwi*')

```
> lm(C01iufilt$di~ C01iufilt$nwi)
Call:
lm(formula = C01iufilt$di ~ C01iufilt$nwi)
```

Residuals:

Min	1Q	Median	3Q	Max
-1041.5	-352.7	-112.2	231.6	2253.4

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	503.869	37.371	13.48	<2e-16 ***
C01iufilt\$nwi	248.229	9.103	27.27	<2e-16 ***

Residual standard error: 523.2 on 965 degrees of freedom
 Multiple R-Squared: 0.4352, Adjusted R-squared: 0.4346
 F-statistic: 743.6 on 1 and 965 DF, p-value: < 2.2e-16

2.1.2. Totalité des données : Tableau 8 (page 229)

Durée de la syllabe ('*ds*') en fonction de sa complexité ('*nps*')

```
> summary(lm(ds~nps))
Call:
lm(formula = ds ~ nps)
```

Residuals:

Min	1Q	Median	3Q	Max
-322.08	-57.27	-15.07	43.73	4934.93

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	22.6595	1.1121	20.38	<2e-16 ***
npsnum	71.2033	0.4246	167.68	<2e-16 ***

Residual standard error: 92.87 on 66538 degrees of freedom
 Multiple R-Squared: 0.297, Adjusted R-squared: 0.297
 F-statistic: 2.812e+04 on 1 and 66538 DF, p-value: < 2.2e-16

Durée du pied ('df') en fonction du nombre de phonèmes dans le pied ('npf')

```
> summary(lm(df~npf))
```

```
Call:
```

```
lm(formula = df ~ npfnum)
```

```
Residuals:
```

```
    Min     1Q  Median     3Q     Max
-368.76 -82.05 -16.83  63.17 3518.95
```

```
Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 145.2658    1.8359   79.13  <2e-16 ***
npfnum       51.9274    0.3242  160.19  <2e-16 ***
```

```
Residual standard error: 130.5 on 31242 degrees of freedom
```

```
Multiple R-Squared: 0.451,    Adjusted R-squared: 0.4509
```

```
F-statistic: 2.566e+04 on 1 and 31242 DF, p-value: < 2.2e-16
```

Durée du pied ('df') en fonction du nombre de syllabes dans le pied ('nsf')

```
> summary(lm(df~nsf))
```

```
Call:
```

```
lm(formula = df ~ nsfnum)
```

```
Residuals:
```

```
    Min     1Q  Median     3Q     Max
-479.86 -89.10 -18.10  71.65 3521.90
```

```
Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 187.8457    1.8577  101.1  <2e-16 ***
nsfnum      110.2525    0.8173  134.9  <2e-16 ***
```

```
Residual standard error: 140 on 31242 degrees of freedom
```

```
Multiple R-Squared: 0.3681,    Adjusted R-squared: 0.3681
```

```
F-statistic: 1.82e+04 on 1 and 31242 DF, p-value: < 2.2e-16
```

Durée de la NRU ('dr') en fonction du nombre de phonèmes dans la NRU ('npr')

```
> summary(lm(dr[ru=="R"]~npr [ru=="R"]))
```

```
Call:
```

```
lm(formula = dr[ru == "R"] ~ nprnum[ru == "R"])
```

```
Residuals:
```

```
    Min     1Q  Median     3Q     Max
-350.78 -74.96 -14.96  61.13 3535.04
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	130.5985	1.7286	75.55	<2e-16 ***
nprnum[ru == "R"]	51.4545	0.3853	133.55	<2e-16 ***

Residual standard error: 114.2 on 31242 degrees of freedom
 Multiple R-Squared: 0.3634, Adjusted R-squared: 0.3634
 F-statistic: 1.784e+04 on 1 and 31242 DF, p-value: < 2.2e-16

Durée de la NRU ('*dr*') en fonction du nombre de syllabes dans la NRU ('*nsr*')

```
> summary(lm(dr[ru=="R"]~nsr [ru=="R"]))
Call:
lm(formula = dr[ru == "R"] ~ nsrnum[ru == "R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-383.12	-83.12	-13.12	69.91	3532.94

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	184.025	1.765	104.26	<2e-16 ***
nsrnum[ru == "R"]	103.032	1.037	99.33	<2e-16 ***

Residual standard error: 124.8 on 31242 degrees of freedom
 Multiple R-Squared: 0.24, Adjusted R-squared: 0.24
 F-statistic: 9867 on 1 and 31242 DF, p-value: < 2.2e-16

Durée de l'anacrouse ('*dr*') en fonction du nombre de phonèmes dans l'anacrouse ('*npr*')

```
> summary(lm(dr [ru=="A"]~npr [ru=="A"]))
Call:
lm(formula = drnum[ru == "A"] ~ nprnum[ru == "A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-233.63	-48.11	-12.74	32.74	2778.26

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	40.854	1.817	22.48	<2e-16 ***
nprnum[ru == "A"]	53.630	0.578	92.79	<2e-16 ***

Residual standard error: 82.58 on 9925 degrees of freedom
 Multiple R-Squared: 0.4645, Adjusted R-squared: 0.4645
 F-statistic: 8610 on 1 and 9925 DF, p-value: < 2.2e-16

Durée de l'anacrouse ('dr') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

```
> summary(lm(dr [ru=="A"]~nsr [ru=="A"]))
```

Call:

```
lm(formula = drnum[ru == "A"] ~ nsrnum[ru == "A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-236.37	-57.27	-16.82	41.73	2832.73

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	27.713	2.258	12.28	<2e-16 ***
nsrnum[ru == "A"]	119.554	1.520	78.64	<2e-16 ***

Residual standard error: 88.58 on 9925 degrees of freedom

Multiple R-Squared: 0.3839, Adjusted R-squared: 0.3838

F-statistic: 6184 on 1 and 9925 DF, p-value: < 2.2e-16

Durée de l'anacrouse initiale ('dr') en fonction du nombre de phonèmes dans l'anacrouse initiale ('npr')

```
> summary(lm(dr [ru=="S"]~npr [ru=="S"]))
```

Call:

```
lm(formula = drnum[ru == "S"] ~ nprnum[ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-231.95	-50.28	-11.28	31.64	4959.72

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	34.441	3.253	10.59	<2e-16 ***
nprnum[ru == "S"]	52.918	0.947	55.88	<2e-16 ***

Residual standard error: 108.4 on 4844 degrees of freedom

Multiple R-Squared: 0.392, Adjusted R-squared: 0.3918

F-statistic: 3123 on 1 and 4844 DF, p-value: < 2.2e-16

Durée de l'anacrouse initiale ('dr') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

```
> summary(lm(dr [ru=="S"]~nsr [ru=="S"]))
```

Call:

```
lm(formula = drnum[ru == "S"] ~ nsrnum[ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-228.19	-57.47	-17.33	32.81	4962.53

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	17.605	3.875	4.543	5.68e-06 ***
nsrnum[ru == "S"]	119.862	2.391	50.124	< 2e-16 ***

Residual standard error: 112.8 on 4844 degrees of freedom
 Multiple R-Squared: 0.3415, Adjusted R-squared: 0.3414
 F-statistic: 2512 on 1 and 4844 DF, p-value: < 2.2e-16

Durée du mot ('dw') en fonction du nombre de phonèmes dans le mot ('npw')

```
> summary(lm(dw~npw))
```

Call:

```
lm(formula = dwnum ~ npwnum)
```

Residuals:

Min	1Q	Median	3Q	Max
-508.62	-70.04	-17.35	54.34	4747.72

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	40.7253	1.1024	36.94	<2e-16 ***
npwnum	68.3101	0.2589	263.87	<2e-16 ***

Residual standard error: 115.3 on 45623 degrees of freedom
 Multiple R-Squared: 0.6041, Adjusted R-squared: 0.6041
 F-statistic: 6.963e+04 on 1 and 45623 DF, p-value: < 2.2e-16

Durée du mot ('dw') en fonction du nombre de syllabes dans le mot ('nsw')

```
> summary(lm(dw~nsw))
```

Call:

```
lm(formula = dwnum ~ nswnum)
```

Residuals:

Min	1Q	Median	3Q	Max
-537.97	-95.65	-23.73	73.35	4756.27

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	57.5672	1.3095	43.96	<2e-16 ***
nswnum	158.0805	0.7704	205.19	<2e-16 ***

Residual standard error: 132.1 on 45623 degrees of freedom
 Multiple R-Squared: 0.4799, Adjusted R-squared: 0.4799

F-statistic: 4.21e+04 on 1 and 45623 DF, p-value: < 2.2e-16

Durée de l'unité intonative ('di') en fonction du nombre de phonèmes dans l'unité intonative ('npi')

```
> summary(lm(di~npi))
```

Call:

```
lm(formula = dinum ~ npinum)
```

Residuals:

Min	1Q	Median	3Q	Max
-1056.7	-337.4	-128.4	214.0	5063.6

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	487.5755	9.8725	49.39	<2e-16 ***
npinum	67.7581	0.5713	118.60	<2e-16 ***

Residual standard error: 495.4 on 10362 degrees of freedom

Multiple R-Squared: 0.5758, Adjusted R-squared: 0.5758

F-statistic: 1.407e+04 on 1 and 10362 DF, p-value: < 2.2e-16

Durée de l'unité intonative ('di') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

```
> summary(lm(di~nsi))
```

Call:

```
lm(formula = dinum ~ nsinum)
```

Residuals:

Min	1Q	Median	3Q	Max
-1116.0	-343.4	-121.4	228.5	5053.4

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	508.636	9.967	51.03	<2e-16 ***
nsinum	163.962	1.422	115.29	<2e-16 ***

Residual standard error: 503.4 on 10362 degrees of freedom

Multiple R-Squared: 0.5619, Adjusted R-squared: 0.5619

F-statistic: 1.329e+04 on 1 and 10362 DF, p-value: < 2.2e-16

Durée de l'unité intonative ('di') en fonction du nombre de pieds dans l'unité intonative ('nfi')

```
> summary(lm(di~nfi))
```

Call:

```
lm(formula = dinum ~ nfinum)
```

Residuals:

Min	1Q	Median	3Q	Max
-1411.0	-333.0	-89.7	232.6	4787.0

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	370.39	11.38	32.55	<2e-16 ***
nfinum	366.33	3.29	111.34	<2e-16 ***

Residual standard error: 513.2 on 10362 degrees of freedom

Multiple R-Squared: 0.5447, Adjusted R-squared: 0.5447

F-statistic: 1.24e+04 on 1 and 10362 DF, p-value: < 2.2e-16

Durée de l'unité intonative ('di') en fonction du nombre de mots dans l'unité intonative ('nwi')

```
> summary(lm(di~nwi))
```

Call:

```
lm(formula = dinum ~ nwinum)
```

Residuals:

Min	1Q	Median	3Q	Max
-1411.26	-355.01	-99.23	255.56	5110.39

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	547.201	10.527	51.98	<2e-16 ***
nwinum	232.406	2.217	104.83	<2e-16 ***

Residual standard error: 529.9 on 10362 degrees of freedom

Multiple R-Squared: 0.5147, Adjusted R-squared: 0.5146

F-statistic: 1.099e+04 on 1 and 10362 DF, p-value: < 2.2e-16

2.2. Isochronie faible

2.2.1. Passage C01

2.2.1.1. Tableau 9 (page 231)

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans la syllabe ('nps')

```
> lm(C01spfilt$dp ~ C01spfilt$nps)
```

Call:

```
lm(formula = C01spfilt$dp ~ C01spfilt$nps)
```

Residuals:

Min	1Q	Median	3Q	Max
-64.99	-32.31	-10.42	26.80	317.69

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	78.6316	1.1300	69.583	<2e-16 ***
C01spfilt\$nps	0.8932	0.3881	2.302	0.0214 *

Residual standard error: 46.68 on 17126 degrees of freedom

Multiple R-Squared: 0.0003092, Adjusted R-squared: 0.0002508

F-statistic: 5.297 on 1 and 17126 DF, p-value: 0.02137

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans la syllabe ('nps')

```
> lm(C01spfilt$z ~ C01spfilt$nps)
```

Call:

```
lm(formula = C01spfilt$z ~ C01spfilt$nps)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6137	-0.7207	-0.1557	0.5354	7.8677

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.035256	0.024178	-1.458	0.145
C01spfilt\$nps	0.012760	0.008303	1.537	0.124

Residual standard error: 0.9987 on 17126 degrees of freedom

Multiple R-Squared: 0.0001379, Adjusted R-squared: 7.949e-05

F-statistic: 2.362 on 1 and 17126 DF, p-value: 0.1244

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans le pied ('npf')

```
> lm(C01spfilt$dp~ C01spfilt$npf)
```

Call:

```
lm(formula = C01spfilt$dp ~ C01spfilt$npf)
```

Residuals:

Min	1Q	Median	3Q	Max
-72.867	-34.100	-7.792	24.170	309.479

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	96.5590	0.9156	105.46	<2e-16 ***
C01spfilt\$npf	-2.3459	0.1282	-18.30	<2e-16 ***

Residual standard error: 46.24 on 17126 degrees of freedom

Multiple R-Squared: 0.01918, Adjusted R-squared: 0.01913

F-statistic: 335 on 1 and 17126 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans le pied ('npf')

```
> lm(C01spfilt$z~ C01spfilt$npf)
```

Call:

```
lm(formula = C01spfilt$z ~ C01spfilt$npf)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.7465	-0.7059	-0.1647	0.5259	8.0562

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.225930	0.019690	11.48	<2e-16 ***
C01spfilt\$npf	-0.034283	0.002756	-12.44	<2e-16 ***

Residual standard error: 0.9943 on 17126 degrees of freedom

Multiple R-Squared: 0.008952, Adjusted R-squared: 0.008894

F-statistic: 154.7 on 1 and 17126 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans la NRU ('npr')

```
> lm(C01spfilt$dp[C01spfilt$ru=="R"] ~ C01spfilt$npr[C01spfilt$ru=="R"])
```

Call:

```
lm(formula = C01spfilt$dp[C01spfilt$ru == "R"] ~ C01spfilt$npr[C01spfilt$ru=="R"])
```

Residuals:

Min 1Q Median 3Q Max
-82.397 -36.304 -7.341 23.696 302.621

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	111.435	1.174	94.90	<2e-16 ***
C01spfilt\$npr[C01spfilt\$ru == "R"]	-5.019	0.206	-24.37	<2e-16 ***

Residual standard error: 47.59 on 12777 degrees of freedom
Multiple R-Squared: 0.04441, Adjusted R-squared: 0.04433
F-statistic: 593.8 on 1 and 12777 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans la NRU ('npr')

> lm(C01spfilt\$z[C01spfilt\$ru=="R"] ~ C01spfilt\$npr[C01spfilt\$ru=="R"])

Call:

lm(formula = C01spfilt\$z[C01spfilt\$ru == "R"] ~ C01spfilt\$npr[C01spfilt\$ru == "R"])

Residuals:

Min 1Q Median 3Q Max
-3.8252 -0.7153 -0.1616 0.5288 7.9357

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.401895	0.025003	16.07	<2e-16 ***
C01spfilt\$npr[C01spfilt\$ru == "R"]	-0.066688	0.004385	-15.21	<2e-16 ***

Residual standard error: 1.013 on 12777 degrees of freedom
Multiple R-Squared: 0.01778, Adjusted R-squared: 0.0177
F-statistic: 231.2 on 1 and 12777 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans l'anacrouse ('npr')

> lm(C01spfilt\$dp[C01spfilt\$ru=="A"] ~ C01spfilt\$npr[C01spfilt\$ru=="A"])

Call:

lm(formula = C01spfilt\$dp[C01spfilt\$ru == "A"] ~ C01spfilt\$npr[C01spfilt\$ru=="A"])

Residuals:

Min 1Q Median 3Q Max
-55.315 -29.890 -5.315 21.919 306.494

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	77.9315	1.5915	48.969	< 2e-16 ***
C01spfilt\$npr[C01spfilt\$ru == "A"]	-1.8084	0.4018	-4.501	7.02e-06 ***

Residual standard error: 39.19 on 3082 degrees of freedom
 Multiple R-Squared: 0.00653, Adjusted R-squared: 0.006207
 F-statistic: 20.26 on 1 and 3082 DF, p-value: 7.023e-06

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans l'anacrouse ('npr')

```
> lm(C01spfilt$z[C01spfilt$ru=="A"] ~ C01spfilt$npr[C01spfilt$ru=="A"])
Call:
lm(formula = C01spfilt$z[C01spfilt$ru == "A"] ~ C01spfilt$npr[C01spfilt$ru == "A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.4623	-0.6729	-0.1338	0.5186	6.9342

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.029954	0.038099	0.786	0.432
C01spfilt\$npr[C01spfilt\$ru == "A"]	-0.038510	0.009619	-4.004	6.39e-05 ***

Residual standard error: 0.9381 on 3082 degrees of freedom
 Multiple R-Squared: 0.005174, Adjusted R-squared: 0.004851
 F-statistic: 16.03 on 1 and 3082 DF, p-value: 6.387e-05

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans l'anacrouse initiale ('npr')

```
> lm(C01spfilt$dp[C01spfilt$ru=="S"] ~ C01spfilt$npr[C01spfilt$ru=="S"])
Call:
lm(formula = C01spfilt$dp[C01spfilt$ru == "S"] ~ C01spfilt$npr[C01spfilt$ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-51.841	-27.285	-6.285	19.159	237.641

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	73.8775	2.3848	30.978	< 2e-16 ***
C01spfilt\$npr[C01spfilt\$ru == "S"]	-1.5184	0.5477	-2.772	0.00565 **

Residual standard error: 35.59 on 1263 degrees of freedom
 Multiple R-Squared: 0.006049, Adjusted R-squared: 0.005262
 F-statistic: 7.687 on 1 and 1263 DF, p-value: 0.005645

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans l'anacrouse initiale ('npr')

```
> lm(C01spfilt$z[C01spfilt$ru=="S"] ~ C01spfilt$npr[C01spfilt$ru=="S"])
```

Call:

```
lm(formula = C01spfilt$z[C01spfilt$ru == "S"] ~ C01spfilt$npr[C01spfilt$ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-1.8379	-0.5692	-0.1452	0.4498	5.3675

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.09859	0.05581	-1.766	0.0776 .
C01spfilt\$npr[C01spfilt\$ru == "S"]	-0.02919	0.01282	-2.277	0.0230 *

Residual standard error: 0.833 on 1263 degrees of freedom

Multiple R-Squared: 0.004088, Adjusted R-squared: 0.0033

F-statistic: 5.185 on 1 and 1263 DF, p-value: 0.02295

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans le mot ('npw')

```
> lm(C01spfilt$dp ~ C01spfilt$npw)
```

Call:

```
lm(formula = C01spfilt$dp ~ C01spfilt$npw)
```

Residuals:

Min	1Q	Median	3Q	Max
-68.06	-34.84	-7.41	23.55	313.55

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	90.2766	0.7913	114.09	<2e-16 ***
C01spfilt\$npw	-1.6084	0.1239	-12.98	<2e-16 ***

Residual standard error: 46.46 on 17126 degrees of freedom

Multiple R-Squared: 0.009737, Adjusted R-squared: 0.009679

F-statistic: 168.4 on 1 and 17126 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans le mot ('npw')

```
> lm(C01spfilt$z ~ C01spfilt$npw)
```

Call:

```
lm(formula = C01spfilt$z ~ C01spfilt$npw)
```

Residuals:

Min 1Q Median 3Q Max
-3.6879 -0.7093 -0.1732 0.5256 7.9492

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.135962	0.016972	8.011	1.21e-15 ***
C01spfilt\$npw	-0.023829	0.002658	-8.964	< 2e-16 ***

Residual standard error: 0.9964 on 17126 degrees of freedom
Multiple R-Squared: 0.00467, Adjusted R-squared: 0.004612
F-statistic: 80.35 on 1 and 17126 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans l'unité intonative ('npi')

```
> lm(C01spfilt$dp ~ C01spfilt$npi)
```

Call:

```
lm(formula = C01spfilt$dp ~ C01spfilt$npi)
```

Residuals:

Min 1Q Median 3Q Max
-70.32 -34.34 -7.86 24.17 313.17

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	90.81356	0.87417	103.89	<2e-16 ***
C01spfilt\$npi	-0.49823	0.04097	-12.16	<2e-16 ***

Residual standard error: 46.49 on 17126 degrees of freedom
Multiple R-Squared: 0.008562, Adjusted R-squared: 0.008504
F-statistic: 147.9 on 1 and 17126 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans l'unité intonative ('npi')

```
> lm(C01spfilt$z ~ C01spfilt$npi)
```

Call:

```
lm(formula = C01spfilt$z ~ C01spfilt$npi)
```

Residuals:

Min 1Q Median 3Q Max
-3.7513 -0.7116 -0.1635 0.5224 7.8755

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.1847457	0.0187186	9.87	<2e-16 ***
C01spfilt\$npi	-0.0094755	0.0008772	-10.80	<2e-16 ***

Residual standard error: 0.9954 on 17126 degrees of freedom
 Multiple R-Squared: 0.006766, Adjusted R-squared: 0.006708
 F-statistic: 116.7 on 1 and 17126 DF, p-value: < 2.2e-16

2.2.1.2. Tableau 10 (page 235)

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans le pied ('nsf')

```
> lm(C01spfilt$dp ~ C01spfilt$nsf)
Call:
lm(formula = C01spfilt$dp ~ C01spfilt$nsf)
```

Residuals:

Min	1Q	Median	3Q	Max
-71.054	-34.505	-8.955	24.495	308.946

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	95.6039	0.8858	107.92	<2e-16 ***
C01spfilt\$nsf	-5.5495	0.3108	-17.86	<2e-16 ***

Residual standard error: 46.26 on 17126 degrees of freedom
 Multiple R-Squared: 0.01828, Adjusted R-squared: 0.01822
 F-statistic: 318.9 on 1 and 17126 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans le pied ('nsf')

```
> lm(C01spfilt$z ~ C01spfilt$nsf)
Call:
lm(formula = C01spfilt$z ~ C01spfilt$nsf)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6723	-0.7124	-0.1668	0.5162	8.0608

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.208180	0.019048	10.93	<2e-16 ***
C01spfilt\$nsf	-0.079652	0.006683	-11.92	<2e-16 ***

Residual standard error: 0.9947 on 17126 degrees of freedom
 Multiple R-Squared: 0.008227, Adjusted R-squared: 0.008169
 F-statistic: 142.1 on 1 and 17126 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans la NRU ('nsr')

```
> lm(C01spfilt$dp[C01spfilt$ru=="R"] ~ C01spfilt$nsr[C01spfilt$ru=="R"])
```

Call:

```
lm(formula = C01spfilt$dp[C01spfilt$ru == "R"] ~ C01spfilt$nsr[C01spfilt$ru == "R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-76.994	-34.319	-5.994	24.681	303.006

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	107.6698	1.0920	98.60	<2e-16 ***
C01spfilt\$nsr[C01spfilt\$ru == "R"]	-11.6756	0.5124	-22.79	<2e-16 ***

Residual standard error: 47.72 on 12777 degrees of freedom

Multiple R-Squared: 0.03905, Adjusted R-squared: 0.03898

F-statistic: 519.2 on 1 and 12777 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans la NRU ('nsr')

```
> lm(C01spfilt$z[C01spfilt$ru=="R"] ~ C01spfilt$nsr[C01spfilt$ru=="R"])
```

Call:

```
lm(formula = C01spfilt$z[C01spfilt$ru == "R"] ~ C01spfilt$nsr[C01spfilt$ru=="R"])))
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6651	-0.7298	-0.1790	0.5283	7.9813

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.34620	0.02322	14.91	<2e-16 ***
C01spfilt\$nsr[C01spfilt\$ru == "R"]	-0.15226	0.01089	-13.98	<2e-16 ***

Residual standard error: 1.015 on 12777 degrees of freedom

Multiple R-Squared: 0.01506, Adjusted R-squared: 0.01498

F-statistic: 195.3 on 1 and 12777 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

```
> lm(C01spfilt$dp[C01spfilt$ru=="A"] ~ C01spfilt$nsr[C01spfilt$ru=="A"])
```

Call:

```
lm(formula = C01spfilt$dp[C01spfilt$ru == "A"] ~ C01spfilt$nsr[C01spfilt$ru=="A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-54.446	-30.448	-4.446	21.551	306.554

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	76.4450	1.7347	44.069	< 2e-16 ***
C01spfilt\$nsr[C01spfilt\$ru == "A"]	-2.9987	0.9628	-3.114	0.00186 **

Residual standard error: 39.25 on 3082 degrees of freedom

Multiple R-Squared: 0.003137, Adjusted R-squared: 0.002814

F-statistic: 9.699 on 1 and 3082 DF, p-value: 0.00186

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

> lm(C01spfilt\$z[C01spfilt\$ru=="A"] ~ C01spfilt\$nsr[C01spfilt\$ru=="A"])

Call:

lm(formula = C01spfilt\$z[C01spfilt\$ru == "A"] ~ C01spfilt\$nsr[C01spfilt\$ru == "A"])

Residuals:

Min	1Q	Median	3Q	Max
-2.4430	-0.6836	-0.1314	0.5190	6.9382

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.003716	0.041515	-0.090	0.9287
C01spfilt\$nsr[C01spfilt\$ru == "A"]	-0.062634	0.023043	-2.718	0.0066 **

Residual standard error: 0.9394 on 3082 degrees of freedom

Multiple R-Squared: 0.002391, Adjusted R-squared: 0.002068

F-statistic: 7.388 on 1 and 3082 DF, p-value: 0.006602

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

> lm(C01spfilt\$dp[C01spfilt\$ru=="S"] ~ C01spfilt\$nsr[C01spfilt\$ru=="S"])

Call:

lm(formula = C01spfilt\$dp[C01spfilt\$ru == "S"] ~ C01spfilt\$nsr[C01spfilt\$ru == "S"])

Residuals:

Min	1Q	Median	3Q	Max
-49.857	-27.596	-7.596	20.143	241.143

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	70.119	2.518	27.845	< 2e-16 ***
C01spfilt\$nsr[C01spfilt\$ru == "S"]	-1.261	1.299	-0.971	0.332

Residual standard error: 35.69 on 1263 degrees of freedom
 Multiple R-Squared: 0.000746, Adjusted R-squared: -4.518e-05
 F-statistic: 0.9429 on 1 and 1263 DF, p-value: 0.3317

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

```
> lm(C01spfilt$z[C01spfilt$ru=="S"] ~ C01spfilt$nsr[C01spfilt$ru=="S"])
Call:
lm(formula = C01spfilt$z[C01spfilt$ru == "S"] ~ C01spfilt$nsr[C01spfilt$ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-1.8649	-0.5484	-0.1549	0.4275	5.4412

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.18553	0.05889	-3.150	0.00167 **
C01spfilt\$nsr[C01spfilt\$ru == "S"]	-0.01598	0.03038	-0.526	0.59896

Residual standard error: 0.8346 on 1263 degrees of freedom
 Multiple R-Squared: 0.000219, Adjusted R-squared: -0.0005725
 F-statistic: 0.2767 on 1 and 1263 DF, p-value: 0.599

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans le mot ('nsw')

```
> lm(C01spfilt$dp ~ C01spfilt$nsw)
Call:
lm(formula = C01spfilt$dp ~ C01spfilt$nsw)
```

Residuals:

Min	1Q	Median	3Q	Max
-66.396	-35.396	-8.402	23.604	313.604

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	88.894	0.763	116.50	<2e-16 ***
C01spfilt\$nsw	-3.497	0.303	-11.54	<2e-16 ***

Residual standard error: 46.51 on 17126 degrees of freedom
 Multiple R-Squared: 0.00772, Adjusted R-squared: 0.007662
 F-statistic: 133.2 on 1 and 17126 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans le mot ('nsw')

```
> lm(C01spfilt$z~ C01spfilt$nsw)
```

Call:

```
lm(formula = C01spfilt$z ~ C01spfilt$nsw)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6351	-0.7106	-0.1734	0.5217	7.9610

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.113536	0.016358	6.941	4.04e-12 ***
C01spfilt\$nsw	-0.050943	0.006495	-7.843	4.65e-15 ***

Residual standard error: 0.997 on 17126 degrees of freedom

Multiple R-Squared: 0.003579, Adjusted R-squared: 0.003521

F-statistic: 61.52 on 1 and 17126 DF, p-value: 4.648e-15

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

```
> lm(C01spfilt$dp~ C01spfilt$nsi)
```

Call:

```
lm(formula = C01spfilt$dp ~ C01spfilt$nsi)
```

Residuals:

Min	1Q	Median	3Q	Max
-69.832	-34.179	-7.701	24.691	312.430

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	89.9624	0.8692	103.50	<2e-16 ***
C01spfilt\$nsi	-1.1306	0.1012	-11.17	<2e-16 ***

Residual standard error: 46.52 on 17126 degrees of freedom

Multiple R-Squared: 0.007236, Adjusted R-squared: 0.007178

F-statistic: 124.8 on 1 and 17126 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

```
> lm(C01spfilt$z~ C01spfilt$nsi)
```

Call:

```
lm(formula = C01spfilt$z ~ C01spfilt$nsi)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.7278	-0.7108	-0.1654	0.5220	7.9174

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.169110	0.018610	9.087	<2e-16 ***
C01spfilt\$nsi	-0.021573	0.002167	-9.957	<2e-16 ***

Residual standard error: 0.9959 on 17126 degrees of freedom
 Multiple R-Squared: 0.005756, Adjusted R-squared: 0.005698
 F-statistic: 99.15 on 1 and 17126 DF, p-value: < 2.2e-16

2.2.1.3. Tableau 11 (page 236)

Durée brute des phonèmes ('dp') en fonction du nombre de pieds dans l'unité intonative ('nfi')

```
> lm(C01spfilt$dp ~ C01spfilt$nfi)
Call:
lm(formula = C01spfilt$dp ~ C01spfilt$nfi)
```

Residuals:

Min	1Q	Median	3Q	Max
-65.11	-32.88	-10.17	26.12	317.36

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	85.3501	0.9048	94.335	< 2e-16 ***
C01spfilt\$nfi	-1.2360	0.2418	-5.111	3.23e-07 ***

Residual standard error: 46.65 on 17126 degrees of freedom
 Multiple R-Squared: 0.001523, Adjusted R-squared: 0.001465
 F-statistic: 26.13 on 1 and 17126 DF, p-value: 3.231e-07

Durée normalisée des phonèmes ('z') en fonction du nombre de pieds dans l'unité intonative ('nfi')

```
> lm(C01spfilt$z ~ C01spfilt$nfi)
Call:
lm(formula = C01spfilt$z ~ C01spfilt$nfi)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6758	-0.7118	-0.1717	0.5301	7.8548

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.125248	0.019343	6.475	9.73e-11 ***
C01syllfilt\$nsf	-0.036419	0.005169	-7.045	1.93e-12 ***

Residual standard error: 0.9973 on 17126 degrees of freedom

Multiple R-Squared: 0.00289, Adjusted R-squared: 0.002831

F-statistic: 49.63 on 1 and 17126 DF, p-value: 1.926e-12

2.2.1.4. Tableau 12 (page 237)

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans le pied ('nsf')

```
> lm(C01syllfilt$ds ~ C01syllfilt$nsf)
```

Call:

```
lm(formula = C01syllfilt$ds ~ C01syllfilt$nsf)
```

Residuals:

Min	1Q	Median	3Q	Max
-192.04	-67.86	-17.86	47.96	677.96

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	256.225	2.999	85.44	<2e-16 ***
C01syllfilt\$nsf	-24.182	1.034	-23.39	<2e-16 ***

Residual standard error: 96.2 on 6785 degrees of freedom

Multiple R-Squared: 0.07459, Adjusted R-squared: 0.07446

F-statistic: 546.9 on 1 and 6785 DF, p-value: < 2.2e-16

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans le pied ('nsf')

```
> lm(log(C01syllfilt$ds) ~ C01syllfilt$nsf)
```

Call:

```
lm(formula = log(C01syllfilt$ds) ~ C01syllfilt$nsf)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.20676	-0.33496	0.03205	0.35089	1.49566

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.433088	0.015727	345.45	<2e-16 ***
C01syllfilt\$nsf	-0.115299	0.005423	-21.26	<2e-16 ***

Residual standard error: 0.5045 on 6785 degrees of freedom

Multiple R-Squared: 0.06246, Adjusted R-squared: 0.06233
F-statistic: 452.1 on 1 and 6785 DF, p-value: < 2.2e-16

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans la NRU ('nsr')

```
> lm(C01syllfilt$ds[C01syllfilt$ru=="R"]~ C01syllfilt$nsr[C01syllfilt$ru=="R"])
Call:
lm(formula=C01syllfilt$ds[C01syllfilt$ru=="R"]~C01syllfilt$nsr[C01syllfilt$ru=="R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-215.25	-64.25	-14.25	48.98	635.75

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	330.857	3.586	92.27	<2e-16 ***
C01syllfilt\$nsr[C01syllfilt\$ru == "R"]	-56.611	1.636	-34.60	<2e-16 ***

Residual standard error: 92.95 on 4825 degrees of freedom
Multiple R-Squared: 0.1988, Adjusted R-squared: 0.1986
F-statistic: 1197 on 1 and 4825 DF, p-value: < 2.2e-16

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans la NRU ('nsr')

```
> lm(log(C01syllfilt$ds[C01syllfilt$ru=="R"])~C01syllfilt$nsr[C01syllfilt$ru == "R"])
Call:
lm(formula=log(C01syllfilt$ds[C01syllfilt$ru=="R"])~C01syllfilt$nsr[C01syllfilt$ru=="R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.27281	-0.26460	0.02476	0.29214	1.28362

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.791096	0.016977	341.11	<2e-16 ***
C01syllfilt\$nsr[C01syllfilt\$ru == "R"]	-0.261275	0.007746	-33.73	<2e-16 ***

Residual standard error: 0.4401 on 4825 degrees of freedom
Multiple R-Squared: 0.1908, Adjusted R-squared: 0.1906
F-statistic: 1138 on 1 and 4825 DF, p-value: < 2.2e-16

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

```
> lm(C01syllfilt$ds[C01syllfilt$ru=="A"]~ C01syllfilt$nsr[C01syllfilt$ru=="A"])
```

Call:

```
lm(formula=C01syllfilt$ds[C01syllfilt$ru=="A"]~C01syllfilt$nsr[C01syllfilt$ru=="A"]
)
```

Residuals:

Min	1Q	Median	3Q	Max
-112.745	-42.745	-5.437	33.640	358.255

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	137.129	3.730	36.763	<2e-16 ***
C01syllfilt\$nsr[C01syllfilt\$ru == "A"]	-2.692	2.086	-1.291	0.197

Residual standard error: 57.39 on 1434 degrees of freedom

Multiple R-Squared: 0.001161, Adjusted R-squared: 0.000464

F-statistic: 1.666 on 1 and 1434 DF, p-value: 0.1970

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

```
> lm(log(C01syllfilt$ds[C01syllfilt$ru=="A"])~ C01syllfilt$nsr[C01syllfilt$ru=="A"])
```

Call:

```
lm(formula=log(C01syllfilt$ds[C01syllfilt$ru=="A"])~C01syllfilt$nsr[C01syllfilt$ru=="A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-1.83728	-0.29308	0.04776	0.31785	1.41269

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.84239	0.02939	164.750	<2e-16 ***
C01syllfilt\$nsr[C01syllfilt\$ru == "A"]	-0.03034	0.01643	-1.846	0.0651 .

Residual standard error: 0.4522 on 1434 degrees of freedom

Multiple R-Squared: 0.002371, Adjusted R-squared: 0.001675

F-statistic: 3.407 on 1 and 1434 DF, p-value: 0.06511

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

```
> lm(C01syllfilt$ds[C01syllfilt$ru=="S"]~ C01syllfilt$nsr[C01syllfilt$ru=="S"])
```

Call:

```
lm(formula=C01syllfilt$ds[C01syllfilt$ru=="S"]~C01syllfilt$nsr[C01syllfilt$ru=="S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-100.42	-32.51	-3.51	27.49	237.49

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	134.598	5.659	23.785	<2e-16 ***
C01syllfilt\$nsr[C01syllfilt\$ru == "S"]	-2.088	2.913	-0.717	0.474

Residual standard error: 51.18 on 522 degrees of freedom

Multiple R-Squared: 0.0009831, Adjusted R-squared: -0.0009308

F-statistic: 0.5137 on 1 and 522 DF, p-value: 0.4739

Durée logarithmique des syllabes ('log(ds') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

```
> lm(log(C01syllfilt$ds[C01syllfilt$ru=="S"])~ C01syllfilt$nsr[C01syllfilt$ru=="S"])
```

Call:

```
lm(formula=log(C01syllfilt$ds[C01syllfilt$ru=="S"])~C01syllfilt$nsr[C01syllfilt$ru=="S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-1.38942	-0.20351	0.05114	0.26650	1.10483

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.82674	0.04586	105.244	<2e-16 ***
C01syllfilt\$nsr[C01syllfilt\$ru == "S"]	-0.01806	0.02361	-0.765	0.445

Residual standard error: 0.4148 on 522 degrees of freedom

Multiple R-Squared: 0.00112, Adjusted R-squared: -0.000794

F-statistic: 0.5851 on 1 and 522 DF, p-value: 0.4447

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans le mot ('nsw')

```
> lm(C01syllfilt$ds~ C01syllfilt$nsw)
```

Call:

```
lm(formula = C01syllfilt$ds ~ C01syllfilt$nsw)
```

Residuals:

Min	1Q	Median	3Q	Max
-177.21	-70.57	-18.21	47.22	702.79

Coefficients:

Estimate	Std. Error	t value	Pr(> t)
----------	------------	---------	----------

(Intercept)	219.418	2.581	85.02	<2e-16 ***
C01syllfilt\$nsw	-12.213	1.004	-12.16	<2e-16 ***

Residual standard error: 98.93 on 6785 degrees of freedom
 Multiple R-Squared: 0.02134, Adjusted R-squared: 0.0212
 F-statistic: 148 on 1 and 6785 DF, p-value: < 2.2e-16

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans le mot ('nsw')

```
> lm(log(C01syllfilt$ds)~ C01syllfilt$nsw)
Call:
lm(formula = log(C01syllfilt$ds) ~ C01syllfilt$nsw)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.194362	-0.328767	0.004377	0.349058	1.624865

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.238359	0.013504	387.922	<2e-16 ***
C01syllfilt\$nsw	-0.049779	0.005253	-9.476	<2e-16 ***

Residual standard error: 0.5176 on 6785 degrees of freedom
 Multiple R-Squared: 0.01306, Adjusted R-squared: 0.01292
 F-statistic: 89.8 on 1 and 6785 DF, p-value: < 2.2e-16

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

```
> lm(C01syllfilt$ds~ C01syllfilt$nsi)
Call:
lm(formula = C01syllfilt$ds ~ C01syllfilt$nsi)
```

Residuals:

Min	1Q	Median	3Q	Max
-181.13	-71.19	-20.32	48.75	690.91

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	223.0649	2.9714	75.07	<2e-16 ***
C01syllfilt\$nsi	-3.9769	0.3438	-11.57	<2e-16 ***

Residual standard error: 99.03 on 6785 degrees of freedom
 Multiple R-Squared: 0.01934, Adjusted R-squared: 0.0192
 F-statistic: 133.8 on 1 and 6785 DF, p-value: < 2.2e-16

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

```
> lm(log(C01syllfilt$ds)~ C01syllfilt$nsi)
```

Call:

```
lm(formula = log(C01syllfilt$ds) ~ C01syllfilt$nsi)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.19693	-0.33552	0.01004	0.35564	1.56415

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.267288	0.015520	339.39	<2e-16 ***
C01syllfilt\$nsi	-0.017989	0.001796	-10.02	<2e-16 ***

Residual standard error: 0.5172 on 6785 degrees of freedom

Multiple R-Squared: 0.01458, Adjusted R-squared: 0.01443

F-statistic: 100.4 on 1 and 6785 DF, p-value: < 2.2e-16

2.2.2. Totalité des données

2.2.2.1. Tableau 13 (page 238)

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans la syllabe ('nps')

```
> lm(dp~nps)
```

Call:

```
lm(formula = dp ~ nps)
```

Residuals:

Min	1Q	Median	3Q	Max
-68.12	-37.50	-15.26	23.74	404.12

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	87.7455	0.4429	198.129	< 2e-16 ***
nps	-0.6223	0.1542	-4.036	5.44e-05 ***

Residual standard error: 55.52 on 168569 degrees of freedom

Multiple R-Squared: 9.663e-05, Adjusted R-squared: 9.07e-05

F-statistic: 16.29 on 1 and 168569 DF, p-value: 5.435e-05

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans la syllabe ('nps')

```
> lm(z~nps)
```

```
Call:
```

```
lm(formula = z ~ nps)
```

```
Residuals:
```

```
   Min      1Q  Median      3Q     Max
-2.1814 -0.6466 -0.1588  0.4534  9.9141
```

```
Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.001039   0.007299   0.142   0.887
nps          0.017887   0.002541   7.038 1.95e-12 ***
```

```
Residual standard error: 0.9151 on 168569 degrees of freedom
```

```
Multiple R-Squared: 0.0002938, Adjusted R-squared: 0.0002879
```

```
F-statistic: 49.54 on 1 and 168569 DF, p-value: 1.951e-12
```

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans le pied ('npf')

```
> lm(dp~npf)
```

```
Call:
```

```
lm(formula = dp ~ npf)
```

```
Residuals:
```

```
   Min      1Q  Median      3Q     Max
-82.49 -39.95 -12.08  26.74 403.02
```

```
Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 104.63121   0.34308  304.98 <2e-16 ***
npf          -3.13724   0.05331  -58.84 <2e-16 ***
```

```
Residual standard error: 54.96 on 168569 degrees of freedom
```

```
Multiple R-Squared: 0.02013, Adjusted R-squared: 0.02012
```

```
F-statistic: 3463 on 1 and 168569 DF, p-value: < 2.2e-16
```

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans le pied ('npf')

```
> lm(z~npf)
```

```
Call:
```

```
lm(formula = z ~ npf)
```

```
Residuals:
```

Min 1Q Median 3Q Max
-2.2463 -0.6314 -0.1602 0.4499 9.9936

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.2634862	0.0056848	46.35	<2e-16 ***
npr	-0.0360387	0.0008834	-40.80	<2e-16 ***

Residual standard error: 0.9107 on 168569 degrees of freedom
Multiple R-Squared: 0.009776, Adjusted R-squared: 0.00977
F-statistic: 1664 on 1 and 168569 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans la NRU ('npr')

```
> lm(dp[ru=="R"]~npr [ru=="R"])
```

Call:

```
lm(formula = dp[ru == "R"] ~ npr [ru == "R"])
```

Residuals:

Min 1Q Median 3Q Max
-93.25 -40.71 -10.71 28.29 412.92

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	118.02596	0.45249	260.8	<2e-16 ***
npr[ru == "R"]	-5.77160	0.08799	-65.6	<2e-16 ***

Residual standard error: 56.35 on 123712 degrees of freedom
Multiple R-Squared: 0.03361, Adjusted R-squared: 0.03361
F-statistic: 4303 on 1 and 123712 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans la NRU ('npr')

```
> lm(z[ru=="R"]~npr [ru=="R"])
```

Call:

```
lm(formula = z[ru == "R"] ~ npr [ru == "R"])
```

Residuals:

Min 1Q Median 3Q Max
-2.3223 -0.6319 -0.1441 0.4594 8.8634

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.378583	0.007353	51.49	<2e-16 ***
npr [ru == "R"]	-0.059399	0.001430	-41.55	<2e-16 ***

Residual standard error: 0.9157 on 123712 degrees of freedom
 Multiple R-Squared: 0.01376, Adjusted R-squared: 0.01375
 F-statistic: 1726 on 1 and 123712 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans l'anacrouse ('npr')

```
> lm(dp[ru=="A"]~npr [ru=="A"])
Call:
lm(formula = dp[ru == "A"] ~ npr[ru == "A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-63.96	-34.70	-10.21	21.54	408.79

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	85.7116	0.6596	129.94	<2e-16 ***
npr [ru == "A"]	-2.7518	0.1682	-16.36	<2e-16 ***

Residual standard error: 49.06 on 28908 degrees of freedom
 Multiple R-Squared: 0.009174, Adjusted R-squared: 0.00914
 F-statistic: 267.7 on 1 and 28908 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans l'anacrouse ('npr')

```
> lm(z[ru=="A"]~npr[ru=="A"])
Call:
lm(formula = z[ru == "A"] ~ npr [ru == "A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.1346	-0.6306	-0.1687	0.4171	9.9690

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.131527	0.012148	10.83	<2e-16 ***
npr[ru == "A"]	-0.043937	0.003098	-14.18	<2e-16 ***

Residual standard error: 0.9036 on 28908 degrees of freedom
 Multiple R-Squared: 0.006912, Adjusted R-squared: 0.006877
 F-statistic: 201.2 on 1 and 28908 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans l'anacrouse initiale ('npr')

```
> lm(dp[ru=="S"]~npr [ru=="S"])
```

Call:

```
lm(formula = dp[ru == "S"] ~ npr [ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-57.24	-32.96	-11.32	19.32	408.68

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	77.8823	0.8310	93.72	<2e-16 ***
npr [ru == "S"]	-1.6411	0.1908	-8.60	<2e-16 ***

Residual standard error: 46.34 on 15945 degrees of freedom

Multiple R-Squared: 0.004617, Adjusted R-squared: 0.004555

F-statistic: 73.96 on 1 and 15945 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans l'anacrouse initiale ('npr')

```
> lm(z[ru=="S"]~npr [ru=="S"])
```

Call:

```
lm(formula = z[ru == "S"] ~ npr [ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-1.7363	-0.5620	-0.1919	0.3514	8.1336

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.078735	0.015029	-5.239	1.64e-07 ***
npr [ru == "S"]	-0.018277	0.003451	-5.296	1.20e-07 ***

Residual standard error: 0.8381 on 15945 degrees of freedom

Multiple R-Squared: 0.001756, Adjusted R-squared: 0.001693

F-statistic: 28.05 on 1 and 15945 DF, p-value: 1.201e-07

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans le mot ('npw')

```
> lm(dp~npw)
```

Call:

```
lm(formula = dp ~ npw)
```

Residuals:

Min 1Q Median 3Q Max
-73.26 -40.64 -11.64 26.46 409.08

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	93.88300	0.29686	316.26	<2e-16 ***
npw	-1.62058	0.05466	-29.65	<2e-16 ***

Residual standard error: 55.38 on 168569 degrees of freedom
Multiple R-Squared: 0.005187, Adjusted R-squared: 0.005181
F-statistic: 878.9 on 1 and 168569 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans le mot ('npw')

> lm(z~npw)

Call:

lm(formula = z ~ npw)

Residuals:

Min 1Q Median 3Q Max
-2.1968 -0.6398 -0.1521 0.4475 9.8826

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.1453325	0.0048989	29.67	<2e-16 ***
npw	-0.0197150	0.0009021	-21.85	<2e-16 ***

Residual standard error: 0.9139 on 168569 degrees of freedom
Multiple R-Squared: 0.002825, Adjusted R-squared: 0.002819
F-statistic: 477.6 on 1 and 168569 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de phonèmes dans l'unité intonative ('npi')

> lm(dp~npi)

Call:

lm(formula = dp ~ npi)

Residuals:

Min 1Q Median 3Q Max
-75.85 -40.31 -11.48 26.76 412.13

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	95.82052	0.29881	320.67	<2e-16 ***
npi	-0.48512	0.01323	-36.66	<2e-16 ***

Residual standard error: 55.3 on 168569 degrees of freedom
Multiple R-Squared: 0.007908, Adjusted R-squared: 0.007902
F-statistic: 1344 on 1 and 168569 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de phonèmes dans l'unité intonative ('npi')

```
> lm(z~npi)
Call:
lm(formula = z ~ npi)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.2341	-0.6355	-0.1579	0.4514	9.9099

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.2045039	0.0049269	41.51	<2e-16 ***
npi	-0.0076681	0.0002182	-35.14	<2e-16 ***

Residual standard error: 0.9119 on 168569 degrees of freedom
Multiple R-Squared: 0.007272, Adjusted R-squared: 0.007266
F-statistic: 1235 on 1 and 168569 DF, p-value: < 2.2e-16

2.2.2.2. Tableau 14 (page 240)

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans le pied ('nsf')

```
> lm(dp~nsf)
Call:
lm(formula = dp ~ nsfnum)
```

Residuals:

Min	1Q	Median	3Q	Max
-76.86	-39.64	-11.42	27.58	401.36

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	103.0765	0.3283	313.95	<2e-16 ***
nsf	-7.2186	0.1270	-56.82	<2e-16 ***

Residual standard error: 55 on 168569 degrees of freedom
Multiple R-Squared: 0.0188, Adjusted R-squared: 0.01879
F-statistic: 3229 on 1 and 168569 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans le pied ('nsf')

```
> lm(z~nsf)
```

```
Call:
```

```
lm(formula = z ~ nsf)
```

```
Residuals:
```

```
   Min      1Q  Median      3Q     Max
-2.2653 -0.6320 -0.1608  0.4402  9.8859
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.265832   0.005433   48.93  <2e-16 ***
nsf          -0.091485   0.002102  -43.52  <2e-16 ***
```

```
Residual standard error: 0.9101 on 168569 degrees of freedom
```

```
Multiple R-Squared: 0.01111, Adjusted R-squared: 0.01111
```

```
F-statistic: 1894 on 1 and 168569 DF, p-value: < 2.2e-16
```

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans la NRU ('nsr')

```
> lm(dp[ru=="R"]~nsr[ru=="R"])
```

```
Call:
```

```
lm(formula = dp[ru == "R"] ~ nsr[ru == "R"])
```

```
Residuals:
```

```
   Min      1Q  Median      3Q     Max
-80.40 -39.40 -10.40  29.60 403.08
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   111.1381    0.4133  268.87  <2e-16 ***
nsr[ru == "R"] -11.7405    0.2141  -54.82  <2e-16 ***
```

```
Residual standard error: 56.64 on 123712 degrees of freedom
```

```
Multiple R-Squared: 0.02372, Adjusted R-squared: 0.02371
```

```
F-statistic: 3006 on 1 and 123712 DF, p-value: < 2.2e-16
```

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans la NRU ('nsr')

```
> lm(z[ru=="R"]~nsr [ru=="R"])
```

```
Call:
```

```
lm(formula = z[ru == "R"] ~ nsr [ru == "R"])
```

```
Residuals:
```

Min 1Q Median 3Q Max
-2.2883 -0.6364 -0.1513 0.4548 8.8664

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.331785	0.006688	49.61	<2e-16 ***
nsr[ru == "R"]	-0.134378	0.003465	-38.78	<2e-16 ***

Residual standard error: 0.9165 on 123712 degrees of freedom

Multiple R-Squared: 0.01201, Adjusted R-squared: 0.012

F-statistic: 1504 on 1 and 123712 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

```
> lm(dp[ru=="A"]~nsr [ru=="A"])
```

Call:

```
lm(formula = dp[ru == "A"] ~ nsr [ru == "A"])
```

Residuals:

Min 1Q Median 3Q Max
-60.137 -35.088 -9.137 20.863 410.863

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	84.1868	0.7124	118.18	<2e-16 ***
nsr[ru == "A"]	-5.0495	0.4019	-12.56	<2e-16 ***

Residual standard error: 49.16 on 28908 degrees of freedom

Multiple R-Squared: 0.00543, Adjusted R-squared: 0.005395

F-statistic: 157.8 on 1 and 28908 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

```
> lm(z[ru=="A"]~nsr [ru=="A"])
```

Call:

```
lm(formula = z[ru == "A"] ~ nsr[ru == "A"])
```

Residuals:

Min 1Q Median 3Q Max
-2.0320 -0.6367 -0.1701 0.4145 9.9343

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.127897	0.013104	9.76	<2e-16 ***
nsr[ru == "A"]	-0.093413	0.007394	-12.63	<2e-16 ***

Residual standard error: 0.9042 on 28908 degrees of freedom
 Multiple R-Squared: 0.005491, Adjusted R-squared: 0.005457
 F-statistic: 159.6 on 1 and 28908 DF, p-value: < 2.2e-16

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

```
> lm(dp[ru=="S"]~nsr[ru=="S"])
Call:
lm(formula = dp[ru == "S"] ~ nsr[ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-54.94	-33.94	-10.80	19.20	409.20

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	77.0778	0.8863	86.968	< 2e-16 ***
nsr [ru == "S"]	-3.1396	0.4516	-6.952	3.74e-12 ***

Residual standard error: 46.37 on 15945 degrees of freedom
 Multiple R-Squared: 0.003022, Adjusted R-squared: 0.002959
 F-statistic: 48.33 on 1 and 15945 DF, p-value: 3.743e-12

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

```
> lm(z[ru=="S"]~nsr [ru=="S"])
Call:
lm(formula = z[ru == "S"] ~ nsr [ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-1.7206	-0.5718	-0.1935	0.3508	8.1127

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.06508	0.01601	-4.064	4.85e-05 ***
nsr [ru == "S"]	-0.04763	0.00816	-5.837	5.42e-09 ***

Residual standard error: 0.8379 on 15945 degrees of freedom
 Multiple R-Squared: 0.002132, Adjusted R-squared: 0.00207
 F-statistic: 34.07 on 1 and 15945 DF, p-value: 5.419e-09

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans le mot ('nsw')

```
> lm(dp~nsw)
```

```
Call:
```

```
lm(formula = dp ~ nsw)
```

```
Residuals:
```

```
   Min     1Q  Median     3Q    Max
-70.03 -40.03 -10.03  26.72 411.09
```

```
Coefficients:
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  92.3975    0.2851   324.1  <2e-16 ***
nsw          -3.3722    0.1333   -25.3  <2e-16 ***
```

```
Residual standard error: 55.42 on 168569 degrees of freedom
```

```
Multiple R-Squared: 0.003783, Adjusted R-squared: 0.003777
```

```
F-statistic: 640.1 on 1 and 168569 DF, p-value: < 2.2e-16
```

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans le mot ('nsw')

```
> lm(z~nsw)
```

```
Call:
```

```
lm(formula = z ~ nsw)
```

```
Residuals:
```

```
   Min     1Q  Median     3Q    Max
-2.1835 -0.6461 -0.1584  0.4491  9.8762
```

```
Coefficients:
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.140747    0.004702   29.93  <2e-16 ***
nsw          -0.048182    0.002198  -21.92  <2e-16 ***
```

```
Residual standard error: 0.9139 on 168569 degrees of freedom
```

```
Multiple R-Squared: 0.002842, Adjusted R-squared: 0.002836
```

```
F-statistic: 480.5 on 1 and 168569 DF, p-value: < 2.2e-16
```

Durée brute des phonèmes ('dp') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

```
> lm(dp~nsi)
```

```
Call:
```

```
lm(formula = dp ~ nsi)
```

```
Residuals:
```


Min 1Q Median 3Q Max
-75.18 -40.46 -11.60 26.97 411.83

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	95.31932	0.29534	322.74	<2e-16 ***
nsi	-1.14299	0.03238	-35.30	<2e-16 ***

Residual standard error: 55.32 on 168569 degrees of freedom
Multiple R-Squared: 0.007336, Adjusted R-squared: 0.00733
F-statistic: 1246 on 1 and 168569 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

> lm(z~nsi)

Call:

lm(formula = z ~ nsi)

Residuals:

Min 1Q Median 3Q Max
-2.2356 -0.6361 -0.1559 0.4490 9.8796

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.2002629	0.0048687	41.13	<2e-16 ***
nsi	-0.0185203	0.0005339	-34.69	<2e-16 ***

Residual standard error: 0.912 on 168569 degrees of freedom
Multiple R-Squared: 0.007089, Adjusted R-squared: 0.007083
F-statistic: 1204 on 1 and 168569 DF, p-value: < 2.2e-16

2.2.2.3. Tableau 15 (page 241)

Durée brute des phonèmes ('dp') en fonction du nombre de pieds dans l'unité intonative ('nfi')

> lm(dp~nfi)

Call:

lm(formula = dp ~ nfi)

Residuals:

Min 1Q Median 3Q Max
-71.90 -39.24 -11.90 25.76 409.76

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	92.56603	0.32364	286.01	<2e-16 ***
nfi	-1.66533	0.07509	-22.18	<2e-16 ***

Residual standard error: 55.44 on 168569 degrees of freedom
 Multiple R-Squared: 0.002909, Adjusted R-squared: 0.002903
 F-statistic: 491.8 on 1 and 168569 DF, p-value: < 2.2e-16

Durée normalisée des phonèmes ('z') en fonction du nombre de pieds dans l'unité intonative ('nfi')

```
> lm(z~nfi)
Call:
lm(formula = z ~ nfi)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.2039	-0.6338	-0.1637	0.4528	9.8887

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.178710	0.005331	33.52	<2e-16 ***
nfi	-0.032872	0.001237	-26.57	<2e-16 ***

Residual standard error: 0.9133 on 168569 degrees of freedom
 Multiple R-Squared: 0.004172, Adjusted R-squared: 0.004166
 F-statistic: 706.2 on 1 and 168569 DF, p-value: < 2.2e-16

2.2.2.4. Tableau 16 (page 242)

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans le pied ('nsf')

```
> lm(ds~nsf)
Call:
lm(formula = ds ~ nsf)
```

Residuals:

Min	1Q	Median	3Q	Max
-221.10	-71.66	-12.66	55.79	4859.90

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	269.3214	1.0279	262.0	<2e-16 ***
nsf	-29.2216	0.3917	-74.6	<2e-16 ***

Residual standard error: 106.4 on 66538 degrees of freedom
 Multiple R-Squared: 0.07718, Adjusted R-squared: 0.07716

F-statistic: 5565 on 1 and 66538 DF, p-value: < 2.2e-16

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans le pied ('nsf')

```
> lm(log(ds)~nsf)
```

Call:

```
lm(formula = log(ds) ~ nsf)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.3965	-0.3370	0.0481	0.3633	3.1960

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.476267	0.005151	1063.08	<2e-16 ***
nsf	-0.135291	0.001963	-68.92	<2e-16 ***

Residual standard error: 0.5332 on 66538 degrees of freedom

Multiple R-Squared: 0.06663, Adjusted R-squared: 0.06661

F-statistic: 4750 on 1 and 66538 DF, p-value: < 2.2e-16

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans la NRU ('nsr')

```
> lm(ds[ru=="R"]~nsr[ru=="R"])
```

Call:

```
lm(formula = ds[ru == "R"] ~ nsr[ru == "R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-255.83	-66.31	-15.31	53.69	3545.17

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	334.3575	1.2339	271.0	<2e-16 ***
nsr[ru == "R"]	-59.5240	0.6233	-95.5	<2e-16 ***

Residual standard error: 101.5 on 46210 degrees of freedom

Multiple R-Squared: 0.1648, Adjusted R-squared: 0.1648

F-statistic: 9121 on 1 and 46210 DF, p-value: < 2.2e-16

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans la NRU ('nsr')

```
> lm(log(ds[ru=="R"])~nsr[ru=="R"])
```

Call:

```
lm(formula = log(ds[ru == "R"]) ~ nsr[ru == "R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.5836	-0.2507	0.0326	0.3022	2.7199

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.801552	0.005540	1047.17	<2e-16 ***
nsr[ru == "R"]	-0.273473	0.002798	-97.73	<2e-16 ***

Residual standard error: 0.4557 on 46210 degrees of freedom

Multiple R-Squared: 0.1713, Adjusted R-squared: 0.1713

F-statistic: 9550 on 1 and 46210 DF, p-value: < 2.2e-16

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

```
> lm(ds[ru=="A"]~nsr[ru=="A"])
```

Call:

```
lm(formula = ds[ru == "A"] ~ nsr[ru == "A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-127.62	-49.12	-10.12	32.38	2833.38

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	154.1084	1.6239	94.899	< 2e-16 ***
nsr[ru == "A"]	-7.4930	0.9232	-8.116	5.23e-16 ***

Residual standard error: 76.01 on 13235 degrees of freedom

Multiple R-Squared: 0.004953, Adjusted R-squared: 0.004877

F-statistic: 65.87 on 1 and 13235 DF, p-value: 5.228e-16

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans l'anacrouse ('nsr')

```
> lm(log(ds[ru=="A"])~nsr[ru=="A"])
```

Call:

```
lm(formula = log(ds[ru == "A"]) ~ nsr[ru == "A"])
```

Residuals:

Min	1Q	Median	3Q	Max
-1.92101	-0.31179	0.04821	0.32419	3.13423

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.919296	0.010868	452.624	<2e-16 ***
nsr[ru == "A"]	-0.053846	0.006179	-8.715	<2e-16 ***

Residual standard error: 0.5087 on 13235 degrees of freedom
 Multiple R-Squared: 0.005705, Adjusted R-squared: 0.00563
 F-statistic: 75.94 on 1 and 13235 DF, p-value: < 2.2e-16

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

```
> lm(ds[ru=="S"]~nsr[ru=="S"])
Call:
lm(formula = ds[ru == "S"] ~ nsr[ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-117.94	-47.94	-12.52	31.91	4963.06

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	141.366	2.695	52.460	< 2e-16 ***
nsr[ru == "S"]	-4.424	1.377	-3.214	0.00132 **

Residual standard error: 94.71 on 7089 degrees of freedom
 Multiple R-Squared: 0.001455, Adjusted R-squared: 0.001314
 F-statistic: 10.33 on 1 and 7089 DF, p-value: 0.001317

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans l'anacrouse initiale ('nsr')

```
> lm(log(ds[ru=="S"]~nsr[ru=="S"])
Call:
lm(formula = log(ds[ru == "S"]) ~ nsr[ru == "S"])
```

Residuals:

Min	1Q	Median	3Q	Max
-1.83868	-0.29448	0.02624	0.33578	3.75388

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.804984	0.014361	334.59	< 2e-16 ***
nsr[ru == "S"]	-0.021864	0.007336	-2.98	0.00289 **

Residual standard error: 0.5047 on 7089 degrees of freedom
Multiple R-Squared: 0.001251, Adjusted R-squared: 0.001111
F-statistic: 8.882 on 1 and 7089 DF, p-value: 0.002889

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans le mot ('nsw')

```
> lm(ds~nsw)
Call:
lm(formula = ds ~ nsw)
```

Residuals:

Min	1Q	Median	3Q	Max
-192.32	-75.45	-18.88	55.55	4902.12

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	224.7537	0.8933	251.60	<2e-16 ***
nsw	-13.4356	0.4110	-32.69	<2e-16 ***

Residual standard error: 109.9 on 66538 degrees of freedom
Multiple R-Squared: 0.01581, Adjusted R-squared: 0.01579
F-statistic: 1069 on 1 and 66538 DF, p-value: < 2.2e-16

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans le mot ('nsw')

```
> lm(log(ds)~nsw)
Call:
lm(formula = log(ds) ~ nsw)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.25416	-0.33879	0.04157	0.38641	3.39058

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.250791	0.004465	1175.90	<2e-16 ***
nsw	-0.052189	0.002054	-25.40	<2e-16 ***

Residual standard error: 0.5493 on 66538 degrees of freedom
Multiple R-Squared: 0.009605, Adjusted R-squared: 0.00959
F-statistic: 645.3 on 1 and 66538 DF, p-value: < 2.2e-16

Durée brute des syllabes ('ds') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

```
> lm(ds~nsi)
```

```
Call:
```

```
lm(formula = ds ~ nsi)
```

```
Residuals:
```

```
   Min      1Q  Median      3Q     Max
-202.79 -75.43 -19.21  54.57 4878.21
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 229.1564    0.9363   244.75  <2e-16 ***
nsi          -3.6816    0.1021   -36.05  <2e-16 ***
```

```
Residual standard error: 109.7 on 66538 degrees of freedom
```

```
Multiple R-Squared: 0.01916, Adjusted R-squared: 0.01914
```

```
F-statistic: 1300 on 1 and 66538 DF, p-value: < 2.2e-16
```

Durée logarithmique des syllabes ('log(ds)') en fonction du nombre de syllabes dans l'unité intonative ('nsi')

```
> lm(log(ds)~nsi)
```

```
Call:
```

```
lm(formula = log(ds) ~ nsi)
```

```
Residuals:
```

```
   Min      1Q  Median      3Q     Max
-2.3082 -0.3418  0.0366  0.3841  3.2843
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.2856005  0.0046744  1130.7  <2e-16 ***
nsi          -0.0164684  0.0005098   -32.3  <2e-16 ***
```

```
Residual standard error: 0.5476 on 66538 degrees of freedom
```

```
Multiple R-Squared: 0.01544, Adjusted R-squared: 0.01542
```

```
F-statistic: 1043 on 1 and 66538 DF, p-value: < 2.2e-16
```

2.3. Effet du paramètre accentuel

Dans cette section des Annexes, sont présentés les résultats des analyses de covariance concernant l'effet de l'accent et de la complexité sur la durée normalisée et brute des phonèmes, puis sur la durée brute et logarithmique des syllabes.

2.3.1. Passage C01

2.3.1.1. Effet de l'accent sur les phonèmes et les syllabes (analyses de variance)

Effet sur la durée brute des phonèmes ('dp')

```
> summary(aov(C01spfilt$dp~C01spfilt$str))
      Df Sum Sq Mean Sq F value    Pr(>F)
C01spfilt$str  1  718873  718873 336.30 < 2.2e-16 ***
Residuals    17126 36607979   2138
```

Effet sur la durée normalisée des phonèmes ('z')

```
> summary(aov(C01spfilt$z~C01spfilt$str))
      Df Sum Sq Mean Sq F value    Pr(>F)
C01spfilt$str  1   49.8   49.8 50.091 1.525e-12 ***
Residuals    17126 17034.2    1.0
```

Effet sur la durée brute des syllabes ('ds')

```
> summary(aov(C01syll$ds~C01syll$str))
      Df Sum Sq Mean Sq F value    Pr(>F)
C01syll$str  1 5772404 5772404 630.87 < 2.2e-16 ***
Residuals   6785 62082486   9150
```

Effet sur la durée logarithmique des syllabes ('log(ds)')

```
> summary(aov(log(C01syll$ds)~C01syll$str))
      Df Sum Sq Mean Sq F value    Pr(>F)
C01syll$str  1  154.63  154.63 621.72 < 2.2e-16 ***
Residuals   6785 1687.50    0.25
```


2.3.1.2. Durée (normalisée et brute) des phonèmes

Effet de l'accent ('str') et du nombre de phonèmes dans la syllabe ('nps') sur la durée normalisée des phonèmes ('z')

```
> lm(z ~ nps * str, data= C01spfilt)
```

```
Call:
```

```
lm(formula = z ~ nps * str, data = C01spfilt)
```

```
Residuals:
```

```
    Min      1Q  Median      3Q     Max
-3.5884 -0.6990 -0.1548  0.5242  7.9270
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.008657	0.036871	-0.235	0.81438
nps	0.023419	0.011829	1.980	0.04774 *
strU	0.016240	0.049525	0.328	0.74298
nps:strU	-0.044714	0.016901	-2.646	0.00816 **

```
Residual standard error: 0.9972 on 17124 degrees of freedom
```

```
Multiple R-Squared: 0.003326, Adjusted R-squared: 0.003151
```

```
F-statistic: 19.05 on 3 and 17124 DF, p-value: 2.515e-12
```

Effet de l'accent ('str') et du nombre de phonèmes dans la syllabe ('nps') sur la durée brute des phonèmes ('dp')

```
> lm(dp ~ nps * str, data= C01spfilt)
```

```
Call:
```

```
lm(formula = dp ~ nps * str, data = C01spfilt)
```

```
Residuals:
```

```
    Min      1Q  Median      3Q     Max
-71.617 -35.422  -7.217  23.917 310.650
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	91.7503	1.7094	53.673	< 2e-16 ***
nps	-1.1334	0.5484	-2.067	0.0388 *
strU	-16.9157	2.2961	-7.367	1.82e-13 ***
nps:strU	1.3292	0.7835	1.696	0.0898 .

```
Residual standard error: 46.23 on 17124 degrees of freedom
```

```
Multiple R-Squared: 0.01951, Adjusted R-squared: 0.01934
```

```
F-statistic: 113.6 on 3 and 17124 DF, p-value: < 2.2e-16
```

Effet de l'accent ('str') et du nombre de phonèmes dans le pied ('npf') sur la durée normalisée des phonèmes ('z')

```
> lm(z ~ npf * str, data= C01spfilt)
Call:
lm(formula = z ~ npf * str, data = C01spfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6444	-0.7055	-0.1549	0.5284	7.9968

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.381253	0.028947	13.171	< 2e-16 ***
npf	-0.055305	0.004589	-12.051	< 2e-16 ***
strU	-0.311267	0.040332	-7.718	1.25e-14 ***
npf:strU	0.038970	0.005851	6.661	2.81e-11 ***

Residual standard error: 0.9926 on 17124 degrees of freedom

Multiple R-Squared: 0.01246, Adjusted R-squared: 0.01229

F-statistic: 72.02 on 3 and 17124 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans le pied ('npf') sur la durée brute des phonèmes ('dp')

```
> lm(dp ~ npf * str, data= C01spfilt)
Call:
lm(formula = dp ~ npf * str, data = C01spfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-83.877	-34.190	-6.781	24.162	303.515

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	110.5087	1.3352	82.76	<2e-16 ***
npf	-3.8159	0.2117	-18.03	<2e-16 ***
strU	-30.4896	1.8604	-16.39	<2e-16 ***
npf:strU	3.1682	0.2699	11.74	<2e-16 ***

Residual standard error: 45.78 on 17124 degrees of freedom

Multiple R-Squared: 0.03835, Adjusted R-squared: 0.03818

F-statistic: 227.6 on 3 and 17124 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans la nru ('npr') sur la durée normalisée des phonèmes ('z')

```
> lm(z ~ npr * str, data= C01spfilt[C01spfilt$ru=="R",])
```

Call:

```
lm(formula = z ~ npr * str, data = C01spfilt[C01spfilt$ru == "R", ])
```

Residuals:

Min	1Q	Median	3Q	Max
-3.8527	-0.7230	-0.1591	0.5246	7.8799

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.476522	0.031297	15.226	< 2e-16 ***
npr	-0.090741	0.006340	-14.313	< 2e-16 ***
strU	-0.068454	0.059390	-1.153	0.24909
npr:strU	0.031136	0.009886	3.149	0.00164 **

Residual standard error: 1.012 on 12775 degrees of freedom

Multiple R-Squared: 0.0207, Adjusted R-squared: 0.02047

F-statistic: 90.01 on 3 and 12775 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans la NRU ('npr') sur la durée brute des phonèmes ('dp')

```
> lm(dp ~ npr * str, data= C01spfilt[C01spfilt$ru=="R",])
```

Call:

```
lm(formula = dp ~ npr * str, data = C01spfilt[C01spfilt$ru == "R", ])
```

Residuals:

Min	1Q	Median	3Q	Max
-87.159	-34.190	-7.966	25.810	299.721

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	119.9189	1.4666	81.765	<2e-16 ***
npr	-6.8799	0.2971	-23.157	<2e-16 ***
strU	-25.2377	2.7832	-9.068	<2e-16 ***
npr:strU	4.4920	0.4633	9.696	<2e-16 ***

Residual standard error: 47.42 on 12775 degrees of freedom

Multiple R-Squared: 0.05139, Adjusted R-squared: 0.05117

F-statistic: 230.7 on 3 and 12775 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans les NRU de plus de trois phonèmes ('npr>3') sur la durée normalisée des phonèmes ('z')

```
> lm(z ~ npr * str, data= C01spfilt[C01spfilt$ru=="R" & C01spfilt$npr>3 ,])
```

Call:

```
lm(formula = z ~ npr * str, data = C01spfilt[C01spfilt$ru == "R" & C01spfilt$npr > 3,])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.5960	-0.7060	-0.1543	0.5136	7.8765

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.3303145	0.0512871	6.440	1.25e-10 ***
npr	-0.0673129	0.0089561	-7.516	6.14e-14 ***
strU	0.1421298	0.0744558	1.909	0.0563 .
npr:strU	-0.0009953	0.0119959	-0.083	0.9339

Residual standard error: 0.9867 on 10021 degrees of freedom

Multiple R-Squared: 0.0139, Adjusted R-squared: 0.0136

F-statistic: 47.07 on 3 and 10021 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans les NRU ('npr') dont on a retiré les deux phonèmes en position finale d'unité intonative sur la durée normalisée des phonèmes ('z')

```
> lm(z~str*npr, data=x[x$ru=="R" & x$npi-x$ipi>1,])
```

Call:

```
lm(formula = z ~ str * npr, data = x[x$ru == "R" & x$npi - x$ipi > 1, ])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.5551	-0.6316	-0.1196	0.4812	7.5869

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.198253	0.029893	6.632	3.47e-11 ***
strU	-0.049920	0.058824	-0.849	0.396
npr	-0.051078	0.005857	-8.721	< 2e-16 ***
strU:npr	0.005435	0.009527	0.571	0.568

Residual standard error: 0.89 on 10556 degrees of freedom

Multiple R-Squared: 0.01364, Adjusted R-squared: 0.01336

F-statistic: 48.65 on 3 and 10556 DF, p-value: < 2.2e-16

Même analyse pour le locuteur A01 : effet de l'accent ('str') et du nombre de phonèmes dans les NRU ('npr') dont on a retiré les deux phonèmes en position finale d'unité intonative sur la durée normalisée des phonèmes ('z')

```
> lm(z~str*npr)
```

Call:

```
lm(formula = z ~ str * npr)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.2462	-0.6854	-0.1643	0.4871	7.6218

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.100956	0.025439	3.969	7.26e-05 ***
strU	-0.052230	0.054177	-0.964	0.335
npr	-0.025497	0.005340	-4.775	1.81e-06 ***
strU:npr	-0.005912	0.009291	-0.636	0.525

Residual standard error: 0.9564 on 17473 degrees of freedom

Multiple R-Squared: 0.006259, Adjusted R-squared: 0.006088

F-statistic: 36.68 on 3 and 17473 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans le mot ('npw') sur la durée normalisée des phonèmes ('z')

```
> lm(z ~ npw * str,data= C01spfilt)
```

Call:

```
lm(formula = z ~ npw * str, data = C01spfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.5743	-0.7034	-0.1498	0.5283	7.9173

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.415814	0.026852	15.48	<2e-16 ***
npw	-0.063028	0.004315	-14.61	<2e-16 ***
strU	-0.466165	0.034581	-13.48	<2e-16 ***
npw:strU	0.063452	0.005461	11.62	<2e-16 ***

Residual standard error: 0.9912 on 17124 degrees of freedom

Multiple R-Squared: 0.01519, Adjusted R-squared: 0.01502

F-statistic: 88.03 on 3 and 17124 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans le mot ('npw') sur la durée brute des phonèmes ('dp')

```
> lm(dp ~ npw * str,data= C01spfilt)
```

Call:

```
lm(formula = dp ~ npw * str, data = C01spfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-84.682	-35.246	-5.817	24.539	304.790

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	112.0961	1.2365	90.66	<2e-16 ***
npw	-4.2072	0.1987	-21.18	<2e-16 ***
strU	-36.9571	1.5924	-23.21	<2e-16 ***
npw:strU	4.2429	0.2515	16.87	<2e-16 ***

Residual standard error: 45.64 on 17124 degrees of freedom

Multiple R-Squared: 0.04429, Adjusted R-squared: 0.04412

F-statistic: 264.5 on 3 and 17124 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans l'unité intonative ('npi') sur la durée normalisée des phonèmes ('z')

```
> lm(z ~ npi * str,data= C01spfilt)
```

Call:

```
lm(formula = z ~ npi * str, data = C01spfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6618	-0.7111	-0.1528	0.5292	7.9186

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.298728	0.027081	11.031	< 2e-16 ***
npi	-0.012632	0.001303	-9.696	< 2e-16 ***
strU	-0.223339	0.037443	-5.965	2.5e-09 ***
npi:strU	0.006472	0.001763	3.670	0.000243 ***

Residual standard error: 0.9939 on 17124 degrees of freedom

Multiple R-Squared: 0.009906, Adjusted R-squared: 0.009733

F-statistic: 57.11 on 3 and 17124 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans l'unité intonative ('npi') sur la durée brute des phonèmes ('dp')

```
> lm(dp ~ npi * str, data= C01spfilt)
```

Call:

```
lm(formula = dp ~ npi * str, data = C01spfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-79.99	-34.28	-7.28	24.35	304.99

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	100.99986	1.25449	80.511	< 2e-16 ***
npi	-0.66910	0.06035	-11.086	< 2e-16 ***
strU	-20.29273	1.73451	-11.699	< 2e-16 ***
npi:strU	0.40119	0.08168	4.912	9.11e-07 ***

Residual standard error: 46.04 on 17124 degrees of freedom

Multiple R-Squared: 0.02758, Adjusted R-squared: 0.02741

F-statistic: 161.9 on 3 and 17124 DF, p-value: < 2.2e-16

2.3.1.3. Durée (brute et logarithmique) des syllabes

Effet de l'accent ('str') et du nombre de phonèmes dans la syllabe ('nps') sur la durée brute des syllabes ('ds')

```
> lm(ds ~ nps * str, data= C01syllfilt)
```

Call:

```
lm(formula = ds ~ nps * str, data = C01syllfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-266.71	-47.01	-11.36	36.26	583.29

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	25.314	4.054	6.244	4.52e-10 ***
nps	75.350	1.440	52.319	< 2e-16 ***
strU	-15.824	5.478	-2.888	0.00388 **
nps:strU	-8.725	2.068	-4.220	2.47e-05 ***

Residual standard error: 73.35 on 6783 degrees of freedom

Multiple R-Squared: 0.4621, Adjusted R-squared: 0.4619

F-statistic: 1942 on 3 and 6783 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans la syllabe ('nps') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds)~nps*str)
Call:
lm(formula = log(ds) ~ nps * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.79424	-0.24249	0.01163	0.25337	1.29470

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.399769	0.021395	205.642	< 2e-16 ***
npsnum	0.338913	0.007601	44.590	< 2e-16 ***
strU	-0.324928	0.028913	-11.238	< 2e-16 ***
npsnum:strU	0.051617	0.010911	4.731	2.28e-06 ***

Residual standard error: 0.3871 on 6783 degrees of freedom
Multiple R-Squared: 0.4482, Adjusted R-squared: 0.4479
F-statistic: 1836 on 3 and 6783 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans le pied ('nsf') sur la durée brute des syllabes ('ds')

```
> lm(ds ~ nsf * str,data= C01syllfilt)
Call:
lm(formula = ds ~ nsf * str, data = C01syllfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-216.00	-64.00	-17.00	45.58	636.00

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	311.998	4.034	77.34	<2e-16 ***
nsf	-37.998	1.601	-23.73	<2e-16 ***
strU	-134.313	5.881	-22.84	<2e-16 ***
nsf:strU	34.149	2.098	16.28	<2e-16 ***

Residual standard error: 91.88 on 6783 degrees of freedom
Multiple R-Squared: 0.1561, Adjusted R-squared: 0.1558
F-statistic: 418.4 on 3 and 6783 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans le pied ('nsf') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds)~nsf*str)
```

Call:

```
lm(formula = log(ds) ~ nsf * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.35154	-0.30180	0.01080	0.32900	1.46364

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.688036	0.021257	267.58	<2e-16 ***
nsf	-0.170382	0.008437	-20.19	<2e-16 ***
strU	-0.634329	0.030989	-20.47	<2e-16 ***
nsf:strU	0.150196	0.011054	13.59	<2e-16 ***

Residual standard error: 0.4842 on 6783 degrees of freedom

Multiple R-Squared: 0.1369, Adjusted R-squared: 0.1365

F-statistic: 358.5 on 3 and 6783 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans la NRU ('nsr') sur la durée brute des syllabes ('ds')

```
> lm(ds ~ nsr * str,data= C01syllfilt[C01syllfilt$ru=="R",])
```

Call:

```
lm(formula = ds ~ nsr * str, data = C01syllfilt[C01syllfilt$ru == "R", ])
```

Residuals:

Min	1Q	Median	3Q	Max
-221.85	-61.85	-13.78	48.15	629.15

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	358.927	4.137	86.762	< 2e-16 ***
nsr	-78.072	2.205	-35.401	< 2e-16 ***
strU	-67.655	9.524	-7.104	1.39e-12 ***
nsr:strU	42.424	3.966	10.696	< 2e-16 ***

Residual standard error: 91.08 on 4823 degrees of freedom

Multiple R-Squared: 0.2311, Adjusted R-squared: 0.2306

F-statistic: 483.2 on 3 and 4823 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans la NRU ('nsr') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds[ru=="R"])~nsr [ru=="R"]*str[ru=="R"])
```

Call:

```
lm(formula = log(ds[ru == "R"]) ~ nsr [ru == "R"] * str[ru == "R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.19900	-0.25762	0.01837	0.29770	1.25617

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.91981	0.01958	302.322	< 2e-16 ***
nsr [ru == "R"]	-0.36254	0.01044	-34.730	< 2e-16 ***
str[ru == "R"]U	-0.28056	0.04508	-6.224	5.26e-10 ***
nsr [ru == "R"]:str[ru == "R"]U	0.18860	0.01877	10.046	< 2e-16 ***

Residual standard error: 0.4311 on 4823 degrees of freedom

Multiple R-Squared: 0.2238, Adjusted R-squared: 0.2233

F-statistic: 463.6 on 3 and 4823 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans les NRU de plus d'une syllabe ('nsr>1') sur la durée brute des syllabes ('ds')

```
> lm(ds ~ nsr * str,data= C01syllfilt[C01syllfilt$ru=="R" & C01syllfilt$nsr>1 ,])
```

Call:

```
lm(formula =ds~nsr*str, data=C01syllfilt[C01syllfilt$ru=="R" & C01syllfilt$nsr>1, ])
```

Residuals:

Min	1Q	Median	3Q	Max
-189.98	-54.27	-10.68	41.92	430.02

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	218.458	8.709	25.084	< 2e-16 ***
nsr	-22.593	3.665	-6.165	7.86e-10 ***
strU	72.813	11.379	6.399	1.77e-10 ***
nsr:strU	-13.055	4.621	-2.825	0.00475 **

Residual standard error: 77.75 on 3485 degrees of freedom

Multiple R-Squared: 0.09722, Adjusted R-squared: 0.09644

F-statistic: 125.1 on 3 and 3485 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans les NRU de plus d'une syllabe ('nsr>1') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds) ~ nsr * str, data= C01syllfilt[C01syllfilt$ru=="R" & C01syllfilt$nsr>1,])
```

Call:

```
lm(formula=log(ds)~nsr*str,data=C01syllfilt[C01syllfilt$ru=="R"&C01syllfilt$nsr>1,])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.17299	-0.25762	0.01837	0.30353	1.18560

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.39095	0.04850	111.153	< 2e-16 ***
nsr	-0.15366	0.02041	-7.529	6.46e-14 ***
strU	0.24829	0.06337	3.918	9.08e-05 ***
nsr:strU	-0.02028	0.02573	-0.788	0.431

Residual standard error: 0.433 on 3485 degrees of freedom

Multiple R-Squared: 0.08063, Adjusted R-squared: 0.07984

F-statistic: 101.9 on 3 and 3485 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans les NRU ('nsr') dont on a exclu les deux phonèmes en position finale d'unité intonative sur la durée brute des syllabes ('ds')

```
> lm(ds~str*nsr)
```

Call:

```
lm(formula = ds ~ str * nsr)
```

Residuals:

Min	1Q	Median	3Q	Max
-219.84	-62.43	-13.83	47.64	630.16

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	357.257	4.278	83.507	< 2e-16 ***
strU	-60.841	10.075	-6.039	1.68e-09 ***
nsr	-77.413	2.257	-34.292	< 2e-16 ***
strU:nsr	39.884	4.159	9.591	< 2e-16 ***

Residual standard error: 92.05 on 4516 degrees of freedom

Multiple R-Squared: 0.2301, Adjusted R-squared: 0.2296

F-statistic: 449.9 on 3 and 4516 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans les NRU ('nsr') dont on a exclu les deux phonèmes en position finale d'unité intonative sur la durée logarithmique des syllabes ('log(ds'))

```
> lm(log(ds)~str*nsr)
```

Call:

```
lm(formula = log(ds) ~ str * nsr)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.197366	-0.254243	0.008121	0.296263	1.260884

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.91202	0.02015	293.381	< 2e-16 ***
strU	-0.26578	0.04746	-5.601	2.26e-08 ***
nsr	-0.35946	0.01063	-33.805	< 2e-16 ***
strU:nsr	0.18157	0.01959	9.269	< 2e-16 ***

Residual standard error: 0.4336 on 4516 degrees of freedom

Multiple R-Squared: 0.2246, Adjusted R-squared: 0.2241

F-statistic: 436 on 3 and 4516 DF, p-value: < 2.2e-16

Effet de l'accent ('str') sur la durée des syllabes ('ds') dans les nru de deux syllabes ('nru=2')

```
>aov(C01syll$ds[C01syll$ru=="R"&C01syll$nsr=="2"]~C01syll$str[C01syll$ru=="R"&C01syll$nsr=="2"])
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
C01syll\$str[...]	1	1282471	1282471	197.74	< 2.2e-16 ***
Residuals	2186	14177804	6486		

Effet de l'accent ('str') sur la durée des syllabes ('ds') dans les nru de trois syllabes ('nru=3')

```
>aov(C01syll$ds[C01syll$ru=="R"&C01syll$nsr=="3"]~C01syll$str[C01syll$ru=="R"&C01syll$nsr=="3"])
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
C01syll\$str[...]	1	281199	281199	53.626	4.675e-13 ***
Residuals	1101	5773366	5244		

Effet de l'accent ('str') sur la durée des syllabes ('ds') dans les nru de quatre syllabes ('nru=4')

```
>aov(C01syll$ds[C01syll$ru=="R"&C01syll$nsr=="4"]~C01syll$str[C01syll$ru=="R"&C01syll$nsr=="4"])
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
C01syll\$str[...]	1	3707	3707	0.8847	0.3481
Residuals	196	821199	4190		

Effet de l'accent ('str') et du nombre de syllabes dans le mot ('nsw') sur la durée brute des syllabes ('ds')

```
> lm(ds ~ nsw * str, data= C01syllfilt)
```

Call:

```
lm(formula = ds ~ nsw * str, data = C01syllfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-214.87	-63.23	-15.00	45.00	636.13

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	313.446	3.585	87.42	<2e-16 ***
nsw	-39.574	1.422	-27.83	<2e-16 ***
strU	-167.052	4.759	-35.10	<2e-16 ***
nsw:strU	48.185	1.861	25.89	<2e-16 ***

Residual standard error: 90.33 on 6783 degrees of freedom

Multiple R-Squared: 0.1844, Adjusted R-squared: 0.184

F-statistic: 511.1 on 3 and 6783 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans le mot ('nsw') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds)~nsw*str)
```

Call:

```
lm(formula = log(ds) ~ nsw * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.21485	-0.29176	0.01422	0.31522	1.65719

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.702892	0.018916	301.49	<2e-16 ***
nsw	-0.181201	0.007503	-24.15	<2e-16 ***
strU	-0.826237	0.025106	-32.91	<2e-16 ***
nsw:strU	0.231966	0.009817	23.63	<2e-16 ***

Residual standard error: 0.4766 on 6783 degrees of freedom

Multiple R-Squared: 0.1638, Adjusted R-squared: 0.1634

F-statistic: 442.8 on 3 and 6783 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans l'unité intonative ('nsi') sur la durée brute des syllabes ('ds')

```
> lm(ds ~ nsi * str, data= C01syllfilt)
```

Call:

```
lm(formula = ds ~ nsi * str, data = C01syllfilt)
```

Residuals:

Min	1Q	Median	3Q	Max
-204.16	-65.19	-17.17	45.53	644.85

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	271.1493	4.1566	65.233	< 2e-16 ***
nsi	-5.9979	0.4920	-12.192	< 2e-16 ***
strU	-93.0188	5.6898	-16.348	< 2e-16 ***
nsi:strU	4.5324	0.6614	6.852	7.91e-12 ***

Residual standard error: 94.56 on 6783 degrees of freedom

Multiple R-Squared: 0.1061, Adjusted R-squared: 0.1057

F-statistic: 268.4 on 3 and 6783 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans l'unité intonative ('nsi') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds)~nsi*str)
```

Call:

```
lm(formula = log(ds) ~ nsi * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.24764	-0.31132	0.02413	0.33706	1.47430

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.487281	0.021759	252.181	< 2e-16 ***
nsi	-0.024600	0.002575	-9.552	< 2e-16 ***
strU	-0.428026	0.029785	-14.371	< 2e-16 ***
nsi:strU	0.016517	0.003463	4.770	1.88e-06 ***

Residual standard error: 0.495 on 6783 degrees of freedom

Multiple R-Squared: 0.0977, Adjusted R-squared: 0.0973

F-statistic: 244.8 on 3 and 6783 DF, p-value: < 2.2e-16

2.3.2. Totalité des données

2.3.2.1. Effet de l'accent sur les phonèmes et les syllabes (analyses de variance)

Effet sur la durée brute des phonèmes ('dp')

```
> aov(allspfilt$dp~allspfilt$str)
              Df Sum Sq Mean Sq F value Pr(>F)
allspfilt$str  1 12304683 12304683  4088 < 2.2e-16 ***
Residuals    168569 507388157   3010
```

Effet sur la durée normalisée des phonèmes ('z')

```
> aov(allspfilt$z~allspfilt$str)
              Df Sum Sq Mean Sq F value Pr(>F)
allspfilt$str  1   788    788 946.25 < 2.2e-16 ***
Residuals    168569 140406      1
```

Effet sur la durée brute des syllabes ('ds')

```
> aov(allsyllfilt$ds~allsyllfilt$str)
              Df Sum Sq Mean Sq F value Pr(>F)
allsyllfilt$str  1 98668286 98668286 9148.4 < 2.2e-16 ***
Residuals      66538 717634617  10785
```

Effet sur la durée logarithmique des syllabes ('log(ds)')

```
> aov(log(allsyllfilt$ds)~allsyllfilt$str)
              Df Sum Sq Mean Sq F value Pr(>F)
allsyllfilt$str  1 2771.1 2771.1 10537 < 2.2e-16 ***
Residuals      66538 17497.7    0.3
```

2.3.2.2. Durée (normalisée et brute) des phonèmes

Effet de l'accent ('str') et du nombre de phonèmes dans la syllabe ('nps') sur la durée normalisée des phonèmes ('z')

```
> lm(z~ nps*str)
Call:
lm(formula = z ~ nps * str)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-2.2458 -0.6265 -0.1527  0.4427  9.9961
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.070827	0.010891	6.503	7.88e-11 ***
nps	0.016806	0.003523	4.771	1.83e-06 ***
strU	-0.028380	0.015007	-1.891	0.0586 .
nps:strU	-0.040056	0.005253	-7.626	2.44e-14 ***

Residual standard error: 0.9125 on 168567 degrees of freedom

Multiple R-Squared: 0.005926, Adjusted R-squared: 0.005909

F-statistic: 335 on 3 and 168567 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans la syllabe ('nps') sur la durée brute des phonèmes ('dp')

> lm(dp~ nps*str)

Call:

lm(formula = dp ~ nps * str)

Residuals:

Min	1Q	Median	3Q	Max
-84.66	-39.60	-10.16	25.34	410.40

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	108.1638	0.6538	165.428	<2e-16 ***
nps	-4.5017	0.2115	-21.286	<2e-16 ***
strU	-26.4009	0.9009	-29.304	<2e-16 ***
nps:strU	2.9189	0.3154	9.256	<2e-16 ***

Residual standard error: 54.78 on 168567 degrees of freedom

Multiple R-Squared: 0.02656, Adjusted R-squared: 0.02654

F-statistic: 1533 on 3 and 168567 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans le pied ('npf') sur la durée normalisée des phonèmes ('z')

> lm(z~ npf*str)

Call:

lm(formula = z ~ npf * str)

Residuals:

Min	1Q	Median	3Q	Max
-2.3414	-0.6252	-0.1500	0.4384	10.0028

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.423536	0.008099	52.29	<2e-16 ***
npf	-0.057687	0.001418	-40.67	<2e-16 ***

strU	-0.359200	0.011663	-30.80	<2e-16 ***
npf:strU	0.045385	0.001852	24.51	<2e-16 ***

Residual standard error: 0.9079 on 168567 degrees of freedom
Multiple R-Squared: 0.01586, Adjusted R-squared: 0.01585
F-statistic: 905.8 on 3 and 168567 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans le pied ('npf') sur la durée brute des phonèmes ('dp')

```
> lm(dp~ npf*str)
Call:
lm(formula = dp ~ npf * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-95.66	-39.07	-10.69	26.81	410.09

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	119.31796	0.48492	246.06	<2e-16 ***
npf	-4.65639	0.08491	-54.84	<2e-16 ***
strU	-36.05390	0.69826	-51.63	<2e-16 ***
npf:strU	3.81717	0.11087	34.43	<2e-16 ***

Residual standard error: 54.36 on 168567 degrees of freedom
Multiple R-Squared: 0.04156, Adjusted R-squared: 0.04155
F-statistic: 2437 on 3 and 168567 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans la NRU ('npr') sur la durée normalisée des phonèmes ('z')

```
> lm(z[ru=="R"]~ npr[ru=="R"]*str[ru=="R"])
Call:
lm(formula = z[ru == "R"] ~ npr[ru == "R"] * str[ru == "R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.3546	-0.6327	-0.1452	0.4633	8.8490

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.446831	0.008953	49.91	<2e-16 ***
npr[ru == "R"]	-0.077350	0.001982	-39.03	<2e-16 ***
str[ru == "R"]U	-0.199253	0.019041	-10.47	<2e-16 ***
npr[ru == "R"]:str[ru == "R"]U	0.042385	0.003365	12.60	<2e-16 ***

Residual standard error: 0.9151 on 123710 degrees of freedom
 Multiple R-Squared: 0.01518, Adjusted R-squared: 0.01516
 F-statistic: 635.8 on 3 and 123710 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans la NRU ('npr') sur la durée brute des phonèmes ('dp')

```
> lm(dp[ru=="R"]~ npr[ru=="R"]*str[ru=="R"])
Call:
lm(formula = dp[ru == "R"] ~ npr[ru == "R"] * str[ru == "R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-98.62	-41.27	-11.05	27.73	417.95

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	124.6932	0.5496	226.87	<2e-16 ***
npr[ru == "R"]	-7.0720	0.1217	-58.13	<2e-16 ***
str[ru == "R"]U	-32.3732	1.1690	-27.69	<2e-16 ***
npr[ru == "R"]:str[ru == "R"]U	5.2310	0.2066	25.32	<2e-16 ***

Residual standard error: 56.18 on 123710 degrees of freedom
 Multiple R-Squared: 0.03963, Adjusted R-squared: 0.0396
 F-statistic: 1701 on 3 and 123710 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans la NRU de plus de trois phonèmes ('npr>3') sur la durée normalisée des phonèmes ('z')

```
> lm(z[ru=="R" & npr>3]~npr [ru=="R" & npr>3]*str[ru=="R" & npr>3]))
Call:
lm(formula = z[ru=="R" & npr>3] ~ npr [ru=="R" & npr>3] * str[ru=="R" & npr>3])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.1819	-0.6118	-0.1396	0.4412	8.5473

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.295372	0.015621	18.909	< 2e-16 ***
npr[ru=="R"&npr>3]	-0.051093	0.002914	-17.533	< 2e-16 ***
str[ru=="R"&npr>3]U	-0.033329	0.023374	-1.426	0.153898
npr[ru=="R"&npr>3]:str[ru=="R"&npr>3]U	0.014023	0.004026	3.483	0.000496 ***

Residual standard error: 0.8804 on 89842 degrees of freedom
 Multiple R-Squared: 0.00539, Adjusted R-squared: 0.005357
 F-statistic: 162.3 on 3 and 89842 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans la NRU de plus de quatre phonèmes ('npr>4') sur la durée normalisée des phonèmes ('z')

```
> lm(z[ru=="R" & npr>4]~npr [ru=="R" & npr>4]*str[ru=="R" & npr>4]))
```

Call:

```
lm(formula=z[ru=="R" & npr>4] ~ npr [ru=="R" & npr>4] * str[ru=="R" & npr>4])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.1668	-0.6100	-0.1437	0.4290	8.3387

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.230235	0.025823	8.916	<2e-16 ***
npr[ru=="R" & npr>4]	-0.041445	0.004204	-9.860	<2e-16 ***
str[ru=="R" & npr>4]U	0.029317	0.033923	0.864	0.387
npr[ru=="R"&npr>4]:str[ru=="R"&npr>4]U	0.004718	0.005364	0.880	0.379

Residual standard error: 0.8667 on 62971 degrees of freedom

Multiple R-Squared: 0.004039, Adjusted R-squared: 0.003992

F-statistic: 85.13 on 3 and 62971 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans la NRU dont on a retiré les deux phonèmes en position finale d'unité intonative ('npr') sur la durée normalisée des phonèmes ('z')

```
> lm(z~str*npr)
```

Call:

```
lm(formula = z ~ str * npr)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.1899	-0.5720	-0.1205	0.4322	8.9752

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.204905	0.008961	22.865	< 2e-16 ***
strU	-0.137659	0.019772	-6.962	3.37e-12 ***
npr	-0.038772	0.001919	-20.203	< 2e-16 ***
strU:npr	0.013439	0.003404	3.948	7.90e-05 ***

Residual standard error: 0.8413 on 103127 degrees of freedom

Multiple R-Squared: 0.009076, Adjusted R-squared: 0.009048

F-statistic: 314.9 on 3 and 103127 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans le mot ('npw') sur la durée normalisée des phonèmes ('z')

```
> lm(z~ npw*str))
```

```
Call:
```

```
lm(formula = z ~ npw * str)
```

Residuals:

```
   Min     1Q  Median     3Q    Max
-2.3162 -0.6264 -0.1550  0.4335  9.9908
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.397506	0.007750	51.29	<2e-16 ***
npw	-0.057407	0.001465	-39.19	<2e-16 ***
strU	-0.428841	0.009990	-42.92	<2e-16 ***
npw:strU	0.060504	0.001853	32.66	<2e-16 ***

Residual standard error: 0.9085 on 168567 degrees of freedom

Multiple R-Squared: 0.0146, Adjusted R-squared: 0.01459

F-statistic: 832.8 on 3 and 168567 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans le mot ('npw') sur la durée brute des phonèmes ('dp')

```
> lm(dp~ npw*str)
```

```
Call:
```

```
lm(formula = dp ~ npw * str)
```

Residuals:

```
   Min     1Q  Median     3Q    Max
-93.766 -38.721 -9.488  25.598 413.167
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	117.44707	0.46410	253.07	<2e-16 ***
npw	-4.68156	0.08772	-53.37	<2e-16 ***
strU	-40.85545	0.59825	-68.29	<2e-16 ***
npw:strU	4.92292	0.11094	44.38	<2e-16 ***

Residual standard error: 54.4 on 168567 degrees of freedom

Multiple R-Squared: 0.03997, Adjusted R-squared: 0.03995

F-statistic: 2339 on 3 and 168567 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans l'unité intonative ('npi') sur la durée normalisée des phonèmes ('z')

```
> lm(z~ npi*str)
```

```
Call:
```

```
lm(formula = z ~ npi * str)
```

Residuals:

```
   Min     1Q  Median     3Q    Max
-2.3236 -0.6266 -0.1530  0.4448  9.9764
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.3102710	0.0069115	44.89	<2e-16 ***
npi	-0.0096920	0.0003134	-30.93	<2e-16 ***
strU	-0.2209908	0.0098323	-22.48	<2e-16 ***
npi:strU	0.0045890	0.0004361	10.52	<2e-16 ***

Residual standard error: 0.9093 on 168567 degrees of freedom

Multiple R-Squared: 0.01284, Adjusted R-squared: 0.01282

F-statistic: 731 on 3 and 168567 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de phonèmes dans l'unité intonative ('npi') sur la durée brute des phonèmes ('dp')

```
> lm(dp~ npi*str)
```

```
Call:
```

```
lm(formula = dp ~ npi * str)
```

Residuals:

```
   Min     1Q  Median     3Q    Max
-86.35 -39.38 -10.91  26.62 413.09
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	106.54296	0.41545	256.45	<2e-16 ***
npi	-0.59680	0.01884	-31.68	<2e-16 ***
strU	-22.69415	0.59102	-38.40	<2e-16 ***
npi:strU	0.30266	0.02621	11.55	<2e-16 ***

Residual standard error: 54.66 on 168567 degrees of freedom

Multiple R-Squared: 0.03094, Adjusted R-squared: 0.03093

F-statistic: 1794 on 3 and 168567 DF, p-value: < 2.2e-16

2.3.2.3. Durée (brute et logarithmique) des syllabes

Effet de l'accent ('str') et du nombre de phonèmes dans la syllabe ('nps') sur la durée brute des syllabes ('ds')

```
> lm(ds~nps*str)
```

```
Call:
```

```
lm(formula = ds ~ nps * str)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-285.85	-55.54	-14.43	38.10	4954.46

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	58.4821	1.6040	36.461	<2e-16 ***
nps	67.4729	0.5658	119.261	<2e-16 ***
strU	-32.5897	2.2075	-14.763	<2e-16 ***
nps:strU	-7.6514	0.8460	-9.044	<2e-16 ***

```
Residual standard error: 89.39 on 66536 degrees of freedom
```

```
Multiple R-Squared: 0.3486, Adjusted R-squared: 0.3486
```

```
F-statistic: 1.187e+04 on 3 and 66536 DF, p-value: < 2.2e-16
```

Effet de l'accent ('str') et du nombre de phonèmes dans la syllabe ('nps') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds)~nps*str)
```

```
Call:
```

```
lm(formula = log(ds) ~ nps * str)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-1.91129	-0.25354	0.01885	0.28124	3.68833

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.552694	0.007704	590.93	<2e-16 ***
nps	0.303033	0.002718	111.51	<2e-16 ***
strU	-0.466818	0.010603	-44.02	<2e-16 ***
nps:strU	0.078360	0.004064	19.28	<2e-16 ***

```
Residual standard error: 0.4294 on 66536 degrees of freedom
```

```
Multiple R-Squared: 0.3948, Adjusted R-squared: 0.3947
```

```
F-statistic: 1.447e+04 on 3 and 66536 DF, p-value: < 2.2e-16
```

Effet de l'accent ('str') et du nombre de syllabes dans le pied ('nsf') sur la durée brute des syllabes ('ds')

```
> lm(ds~nsf*str)
```

Call:

```
lm(formula = ds ~ nsf * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-264.18	-64.18	-14.18	48.06	4931.73

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	324.0537	1.3295	243.75	<2e-16 ***
nsf	-40.8765	0.5849	-69.89	<2e-16 ***
strU	-152.6263	2.0077	-76.02	<2e-16 ***
nsf:strU	37.7149	0.7819	48.24	<2e-16 ***

Residual standard error: 100.2 on 66536 degrees of freedom

Multiple R-Squared: 0.1814, Adjusted R-squared: 0.1814

F-statistic: 4915 on 3 and 66536 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans le pied ('nsf') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds)~nsf*str)
```

Call:

```
lm(formula = log(ds) ~ nsf * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.60685	-0.28814	0.03764	0.33204	3.55751

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.724872	0.006628	863.75	<2e-16 ***
nsf	-0.173587	0.002916	-59.53	<2e-16 ***
strU	-0.733467	0.010009	-73.28	<2e-16 ***
nsf:strU	0.161667	0.003898	41.48	<2e-16 ***

Residual standard error: 0.4996 on 66536 degrees of freedom

Multiple R-Squared: 0.1806, Adjusted R-squared: 0.1806

F-statistic: 4889 on 3 and 66536 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans la NRU ('nsr') sur la durée brute des syllabes ('ds')

```
> lm(ds[ru=="R"]~nsr [ru=="R"]*str[ru=="R"])
```

Call:

```
lm(formula = ds[ru == "R"] ~ nsr [ru == "R"] * str[ru == "R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-262.21	-67.61	-12.21	51.39	3538.79

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	354.8122	1.4233	249.29	<2e-16 ***
nsr [ru == "R"]	-73.6033	0.8364	-88.00	<2e-16 ***
str[ru == "R"]U	-95.1858	3.7623	-25.30	<2e-16 ***
nsr [ru == "R"]:str[ru == "R"]U	46.6413	1.6438	28.37	<2e-16 ***

Residual standard error: 100.6 on 46208 degrees of freedom

Multiple R-Squared: 0.1797, Adjusted R-squared: 0.1796

F-statistic: 3373 on 3 and 46208 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans la NRU ('nsr') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds[ru=="R"]~nsr [ru=="R"]*str[ru=="R"])
```

Call:

```
lm(formula = log(ds[ru == "R"]) ~ nsr [ru == "R"] * str[ru == "R"])
```

Residuals:

Min	1Q	Median	3Q	Max
-2.60840	-0.27815	0.02723	0.30167	2.69517

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.883105	0.006403	918.76	<2e-16 ***
nsr [ru == "R"]	-0.330270	0.003763	-87.77	<2e-16 ***
str[ru == "R"]U	-0.369184	0.016926	-21.81	<2e-16 ***
nsr [ru == "R"]:str[ru == "R"]U	0.183207	0.007395	24.77	<2e-16 ***

Residual standard error: 0.4526 on 46208 degrees of freedom

Multiple R-Squared: 0.1827, Adjusted R-squared: 0.1826

F-statistic: 3443 on 3 and 46208 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans la NRU de plus d'une syllabe ('nsr>1') sur la durée brute des syllabes ('ds')

> lm(ds[ru=="R"&nsr>1]~nsr [ru=="R"&nsr>1]*str[ru=="R"&nsr>1])

Call:

lm(formula = ds[ru=="R" & nsr>1] ~ nsr [ru=="R" & nsr>1] * str[ru=="R" & nsr>1])

Residuals:

Min	1Q	Median	3Q	Max
-186.70	-58.41	-8.74	44.30	1294.30

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	230.215	3.481	66.136	< 2e-16 ***
nsr [ru=="R" & nsr>1]	-21.403	1.539	-13.912	< 2e-16 ***
str[ru=="R" & nsr>1]U	29.412	4.537	6.482	9.19e-11 ***
nsr [ru=="R" & nsr>1]:str[ru=="R" & nsr>1]U	-5.559	1.941	-2.865	0.00418 **

Residual standard error: 84.07 on 29343 degrees of freedom

Multiple R-Squared: 0.02886, Adjusted R-squared: 0.02876

F-statistic: 290.6 on 3 and 29343 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans la NRU de plus de deux syllabes ('nsr>2') sur la durée brute des syllabes ('ds')

> lm(ds[ru=="R"&nsr>2]~nsr [ru=="R"&nsr>2]*str[ru=="R"&nsr>2]))

Call:

lm(formula = ds[ru=="R" & nsr>2] ~ nsr [ru=="R" & nsr>2] * str[ru=="R" & nsr>2])

Residuals:

Min	1Q	Median	3Q	Max
-156.19	-55.19	-13.22	38.48	844.81

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	149.777	15.500	9.663	< 2e-16 ***
nsr[ru=="R" & nsr>2]	3.916	4.983	0.786	0.431962
str[ru=="R" & nsr>2]U	64.326	18.365	3.503	0.000463 ***
nsr[ru=="R"&nsr>2]:str[ru=="R"&nsr>2]U	-16.888	5.875	-2.874	0.004060 **

Residual standard error: 78.9 on 7952 degrees of freedom

Multiple R-Squared: 0.007014, Adjusted R-squared: 0.00664

F-statistic: 18.72 on 3 and 7952 DF, p-value: 4.248e-12

Effet de l'accent ('str') et du nombre de syllabes dans la NRU de plus de trois syllabes ('nsr>3') sur la durée brute des syllabes ('ds')

```
> lm(ds[ru=="R"&nsr>3]~nsr [ru=="R"&nsr>3]*str[ru=="R"&nsr>3]))
```

Call:

```
lm(formula = ds[ru=="R" & nsr>3] ~ nsr [ru == "R" & nsr>3] * str[ru=="R" & nsr>3])
```

Residuals:

Min	1Q	Median	3Q	Max
-141.723	-51.723	-6.182	38.277	267.277

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	352.91	191.27	1.845	0.0653 .
nsr[ru=="R"&nsr>3]	-46.68	47.72	-0.978	0.3282
str[ru=="R"&nsr>3]U	-285.10	226.28	-1.260	0.2080
nsr[ru=="R"&nsr>3]:str[ru=="R"&nsr>3]U	70.16	56.46	1.243	0.2143

Residual standard error: 67.23 on 1001 degrees of freedom

Multiple R-Squared: 0.00224, Adjusted R-squared: -0.0007502

F-statistic: 0.7491 on 3 and 1001 DF, p-value: 0.5229

Effet de l'accent ('str') et du nombre de syllabes dans la NRU dont on a exclu les deux phonèmes en position finale d'unité intonative ('nsr') sur la durée brute des syllabes ('ds')

```
> lm(ds~str*nsr)
```

Call:

```
lm(formula = ds ~ str * nsr)
```

Residuals:

Min	1Q	Median	3Q	Max
-260.46	-67.12	-14.78	51.88	3540.54

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	351.7961	1.4709	239.16	<2e-16 ***
strU	-87.2823	4.0026	-21.81	<2e-16 ***
nsr	-72.3386	0.8536	-84.74	<2e-16 ***
strU:nsr	43.4850	1.7285	25.16	<2e-16 ***

Residual standard error: 101.3 on 43092 degrees of freedom

Multiple R-Squared: 0.1749, Adjusted R-squared: 0.1748

F-statistic: 3045 on 3 and 43092 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans la NRU dont on a exclu les deux phonèmes en position finale d'unité intonative ('nsr') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds)~str*nsr)
```

Call:

```
lm(formula = log(ds) ~ str * nsr)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.60127	-0.27892	0.02646	0.30090	2.70230

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.870855	0.006597	889.89	<2e-16 ***
strU	-0.335609	0.017952	-18.70	<2e-16 ***
nsr	-0.325145	0.003828	-84.93	<2e-16 ***
strU:nsr	0.168106	0.007752	21.68	<2e-16 ***

Residual standard error: 0.4543 on 43092 degrees of freedom

Multiple R-Squared: 0.1807, Adjusted R-squared: 0.1806

F-statistic: 3167 on 3 and 43092 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans le mot ('nwr') sur la durée brute des syllabes ('ds')

```
> lm(ds~nsw*str)
```

Call:

```
lm(formula = ds ~ nsw * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-257.22	-64.12	-14.12	46.32	4936.70

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	318.3609	1.2107	262.96	<2e-16 ***
nsw	-42.1409	0.5763	-73.12	<2e-16 ***
strU	-175.4352	1.6318	-107.51	<2e-16 ***
nsw:strU	52.3302	0.7560	69.22	<2e-16 ***

Residual standard error: 99.62 on 66536 degrees of freedom

Multiple R-Squared: 0.1911, Adjusted R-squared: 0.1911

F-statistic: 5241 on 3 and 66536 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans le mot ('nwr') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds)~nsw*str)
```

Call:

```
lm(formula = log(ds) ~ nsw * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.58719	-0.29360	0.04208	0.32994	3.57513

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.722090	0.006006	952.75	<2e-16 ***
nsw	-0.190463	0.002859	-66.62	<2e-16 ***
strU	-0.886423	0.008095	-109.50	<2e-16 ***
nsw:strU	0.253563	0.003750	67.61	<2e-16 ***

Residual standard error: 0.4942 on 66536 degrees of freedom

Multiple R-Squared: 0.1983, Adjusted R-squared: 0.1983

F-statistic: 5487 on 3 and 66536 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans l'unité intonative ('nwi') sur la durée brute des syllabes ('ds')

```
> lm(ds~nsi*str)
```

Call:

```
lm(formula = ds ~ nsi * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-248.19	-66.44	-16.09	49.61	4925.31

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	276.3865	1.2474	221.56	<2e-16 ***
nsi	-4.5992	0.1395	-32.97	<2e-16 ***
strU	-97.9934	1.7570	-55.77	<2e-16 ***
nsi:strU	2.7490	0.1922	14.31	<2e-16 ***

Residual standard error: 102.9 on 66536 degrees of freedom

Multiple R-Squared: 0.1375, Adjusted R-squared: 0.1375

F-statistic: 3536 on 3 and 66536 DF, p-value: < 2.2e-16

Effet de l'accent ('str') et du nombre de syllabes dans l'unité intonative ('nwi') sur la durée logarithmique des syllabes ('log(ds)')

```
> lm(log(ds)~nsi*str)
```

Call:

```
lm(formula = log(ds) ~ nsi * str)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.53101	-0.30781	0.03535	0.34744	3.51673

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.5118267	0.0061769	892.333	<2e-16 ***
nsi	-0.0181883	0.0006908	-26.328	<2e-16 ***
strU	-0.4724403	0.0086999	-54.304	<2e-16 ***
nsi:strU	0.0086262	0.0009515	9.066	<2e-16 ***

Residual standard error: 0.5094 on 66536 degrees of freedom

Multiple R-Squared: 0.1483, Adjusted R-squared: 0.1483

F-statistic: 3862 on 3 and 66536 DF, p-value: < 2.2e-16

3. Annexes 3. Syllabification en anglais britannique

3.1. Résultats du test *t* (passage C01)

Comparaison de la durée normalisée ('z') des attaques ('O') et des consonnes ambisyllabiques ('A')

```
> t.test(C01c$z[C01c$ss2=="O"],C01c$z[C01c$ss2=="A"],var.equal=T,paired=F)
```

Two Sample t-test

data: C01c\$z[C01c\$ss2 == "O"] and C01c\$z[C01c\$ss2 == "A"]

t = -1.3492, df = 5937, p-value = 0.1773

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.07936163 0.01465402

sample estimates:

mean of x mean of y

-0.04319554 -0.01084174

Comparaison de la durée brute ('dp') des attaques ('O') et des consonnes ambisyllabiques ('A')

```
> t.test(C01c$dp[C01c$ss2=="O"],C01c$dp[C01c$ss2=="A"],var.equal=T,paired=F)
```

Two Sample t-test

data: C01c\$dp[C01c\$ss2 == "O"] and C01c\$dp[C01c\$ss2 == "A"]

t = -6.6682, df = 5937, p-value = 2.823e-11

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-8.466931 -4.619669

sample estimates:

mean of x mean of y

67.96679 74.51009

Comparaison de la durée normalisée ('z') des attaques ('O') et des codas ('C')

```
> t.test(C01c$z[C01c$ss2=="O"],C01c$z[C01c$ss2=="C"],var.equal=T,paired=F)
```

Two Sample t-test

data: C01c\$z[C01c\$ss2 == "O"] and C01c\$z[C01c\$ss2 == "C"]

t = -3.781, df = 6331, p-value = 0.0001577

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.1518168 -0.0481425

sample estimates:

mean of x mean of y

-0.04319554 0.05678413

Comparaison de la durée brute ('dp') des attaques ('O') et des codas ('C')

```
> t.test(C01c$dp[C01c$ss2=="O"],C01c$dp[C01c$ss2=="C"],var.equal=T,paired=F)
```

Two Sample t-test

data: C01c\$dp[C01c\$ss2 == "O"] and C01c\$dp[C01c\$ss2 == "C"]

t = -13.3175, df = 6331, p-value = < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-16.97952 -12.62215

sample estimates:

mean of x mean of y

67.96679 82.76763

Comparaison de la durée normalisée ('z') des attaques ('O') et des codas ('C') en excluant les codas en position finale d'unité intonative

```
> t.test(C01b$z[C01b$ss2=="O"],C01b$z[C01b$ss2=="C"],var.equal=T,paired=F)
```

Two Sample t-test

data: C01b\$z[C01b\$ss2 == "O"] and C01b\$z[C01b\$ss2 == "C"]

t = 1.0074, df = 6150, p-value = 0.3138

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.02524030 0.07860307

sample estimates:

mean of x mean of y

-0.00796676 -0.03464814

Comparaison de la durée brute ('dp') des attaques ('O') et des codas ('C') en excluant les codas en position finale d'unité intonative

```
> t.test(C01b$dp[C01b$ss2=="O"],C01b$dp[C01b$ss2=="C"],var.equal=T,paired=F)
```

Two Sample t-test

data: C01b\$dp[C01b\$ss2 == "O"] and C01b\$dp[C01b\$ss2 == "C"]

t = -5.2007, df = 6150, p-value = 2.050e-07

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-7.357451 -3.329220
sample estimates:
mean of x mean of y
67.84171 73.18504

Comparaison de la durée normalisée ('z') de toutes les voyelles ('N') et des voyelles qui précèdent une consonne ambisyllabique ('N2')

```
> t.test(C01b$z[C01b$ss2=="N"],C01b$z[C01b$ss2=="N2"],var.equal=T,paired=F)
```

Two Sample t-test

data: C01b\$z[C01b\$ss2 == "N"] and C01b\$z[C01b\$ss2 == "N2"]
t = 13.22, df = 6508, p-value = < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.2781516 0.3750052
sample estimates:
mean of x mean of y
0.1377549 -0.1888235

Comparaison de la durée brute ('dp') de toutes les voyelles ('N') et des voyelles qui précèdent une consonne ambisyllabique ('N2')

```
> t.test(C01b$dp[C01b$ss2=="N"],C01b$dp[C01b$ss2=="N2"],var.equal=T,paired=F)
```

Two Sample t-test

data: C01b\$dp[C01b\$ss2 == "N"] and C01b\$dp[C01b\$ss2 == "N2"]
t = 12.5719, df = 6510, p-value = < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
12.57075 17.21528
sample estimates:
mean of x mean of y
89.36318 74.47016

3.2. Résultats du test t (totalité des données)

Comparaison de la durée normalisée ('z') des attaques ('O') et des consonnes ambisyllabiques ('A')

```
> t.test(allc$new.zp[allc$ss2=="O"],allc$new.zp[allc$ss2=="A"],var.equal=T,paired=F)
```

Two Sample t-test

data: allc\$new.zp[allc\$ss2 == "O"] and allc\$new.zp[allc\$ss2 == "A"]

t = 14.0064, df = 57075, p-value = < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

0.08924155 0.11828154

sample estimates:

mean of x mean of y

0.08870036 -0.01506119

Comparaison de la durée brute ('dp') des attaques ('O') et des consonnes ambisyllabiques ('A')

```
> t.test(allc$dp[allc$ss2=="O"],allc$dp[allc$ss2=="A"],var.equal=T,paired=F)
```

Two Sample t-test

data: allc\$dp[allc\$ss2 == "O"] and allc\$dp[allc\$ss2 == "A"]

t = 7.0068, df = 57075, p-value = 2.465e-12

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

1.886565 3.351913

sample estimates:

mean of x mean of y

70.98934 68.37010

Comparaison de la durée normalisée ('z') des attaques ('O') et des codas ('C')

```
> t.test(allc$z[allc$ss2=="O"],allc$z[allc$ss2=="C"],var.equal=T,paired=F)
```

Two Sample t-test

data: allc\$z[allc\$ss2 == "O"] and allc\$z[allc\$ss2 == "C"]

t = -3.2241, df = 67837, p-value = 0.001264

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.036587787 -0.008921436

sample estimates:

mean of x mean of y
0.08870036 0.11145497

Comparaison de la durée brute ('dp') des attaques ('O') et des codas ('C')

```
> t.test(allc$dp[allc$ss2=="O"],allc$dp[allc$ss2=="C"],var.equal=T,paired=F)
```

Two Sample t-test

data: allc\$dp[allc\$ss2 == "O"] and allc\$dp[allc\$ss2 == "C"]
t = -36.4155, df = 67837, p-value = < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-14.56094 -13.07356
sample estimates:
mean of x mean of y
70.98934 84.80659

Comparaison de la durée normalisée ('z') des attaques ('O') et des codas ('C') en excluant les codas en position finale d'unité intonative

```
> t.test(all$z[all$ss2=="O"],all$z[all$ss2=="C"],var.equal=T,paired=F)
```

Two Sample t-test

data: all\$z[all\$ss2 == "O"] and all\$z[all\$ss2 == "C"]
t = 13.2042, df = 61009, p-value = < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.08149318 0.10990376
sample estimates:
mean of x mean of y
0.088636276 -0.007062192

Comparaison de la durée normalisée ('z') de toutes les voyelles ('N') et des voyelles précédant une consonne ambisyllabique ('N2')

```
> t.test(all$z[all$ss2=="N"],all$z[all$ss2=="N2"],var.equal=T,paired=F)
```

Two Sample t-test

data: all\$z[all\$ss2 == "N"] and all\$z[all\$ss2 == "N2"]
t = 19.6667, df = 61475, p-value = < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.1327737 0.1621680
sample estimates:

mean of x mean of y
0.009037856 -0.138432984

Comparaison de la durée brute ('dp') de toutes les voyelles ('N') et des voyelles précédant une consonne ambisyllabique ('N2')

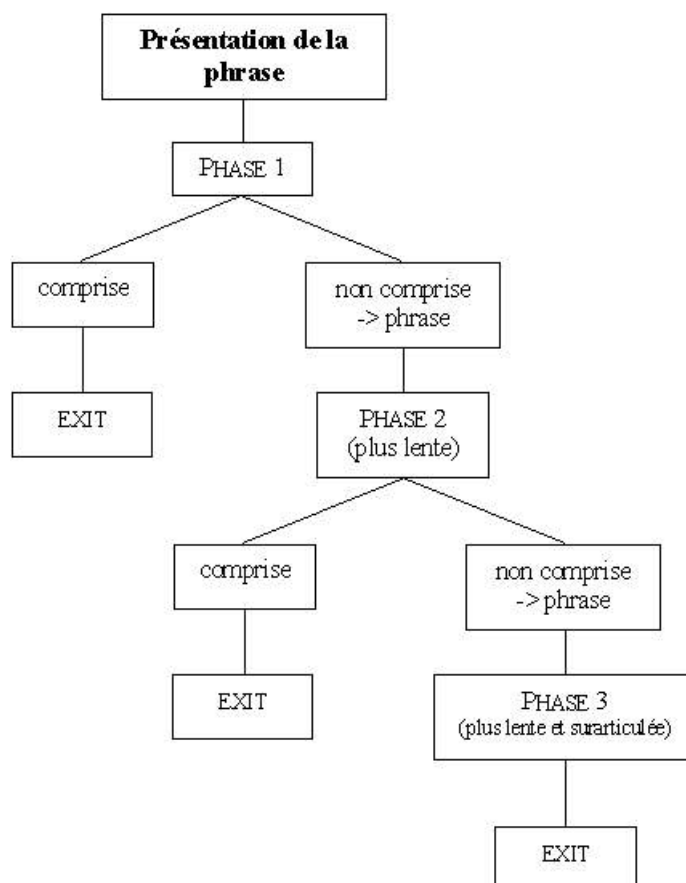
```
> t.test(all$dp[all$ss2=="N"],all$dp[all$ss2=="N2"],var.equal=T,paired=F)
```

Two Sample t-test

data: all\$dp[all\$ss2 == "N"] and all\$dp[all\$ss2 == "N2"]
t = 18.5754, df = 61475, p-value = < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
8.092334 10.001524
sample estimates:
mean of x mean of y
97.35271 88.30578

3.3. Fonctionnement du test de production

La figure ci-après illustre les différentes étapes du test de production : la Phase 1 (lecture normale), la Phase 2 (lecture plus lente) et la Phase 3 (lecture encore plus lente et surarticulée).



3.4. *Phrases de l'expérience en production et différentes phases de lecture*

3.4.1. **Première partie de l'expérience**

1. Why didn't you say so? (Phase 1)
2. Kevin's mother often reminds him to take his camera. (Phases 1, 2 et 3)
3. « Hello! I'd like to inquire about accommodation. » (Phases 1 et 2)
4. Let's focus on this absolutely amazing scenery. (Phases 1, 2 et 3)
5. Come and get it! (Phase 1)
6. There was an assumption that he suffocated her with a pillow. (Phases 1, 2 et 3)
7. Does anybody around here fancy some jacket potatoes? (Phases 1, 2 et 3)
8. He said that he would do it today. (Phase 1)
9. For our annual Halloween party, we'll allow any costume. (Phases 1 et 2)
10. Yesterday, we eventually managed to watch Harry Potter and Braveheart. (Phases 1, 2 et 3)
11. Mary's Mummy is singing a lullaby to her baby. (Phases 1 et 2)
12. Have you ever been to Kenya? (Phase 1)
13. The passenger picked up some lilies, daisies and daffodils. (Phases 1, 2 et 3)
14. The enormous animal vanished in the forest. (Phases 1 et 2)
15. Send Katy an email! (Phase 1)
16. Everyone congratulated the Canary population for their peaceful independence. (Phases 1, 2 et 3)
17. He apologised for his unacceptable behaviour. (Phases 1 et 2)
18. Which restaurant do you prefer, Chinese or Thai? (Phase 1)
19. Attention everyone, don't leave your luggage unattended or police officers will remove it. (Phases 1, 2 et 3)
20. The city hall was occupied by terrorists. (Phases 1 et 2)
21. How could you get there on time? (Phase 1)
22. « Hello! I'd like to inquire about accommodation. » (Phases 1, 2 et 3)
23. The happy fishermen were selling a great amount of anchovy. (Phases 1, 2 et 3)
24. His name is Richard. (Phase 1)
25. They were judging Rachel for illegally selling tobacco. (Phases 1, 2 et 3)
26. What's your family name? (Phase 1)
27. It's got no added sugar in it. (Phases 1 et 2)

28. They visited Fountain's Abbey and they thoroughly enjoyed it. (Phase 1, 2 et 3)
29. How did this happen? (Phase 1)
30. He apologised for his unacceptable behaviour. (Phases 1, 2 et 3)
31. The famous novelist managed to finish his book about pottery. (Phases 1 et 2)
32. Tomorrow, it's John's birthday. (Phase 1)
33. Let's now focus on this absolutely amazing scenery. (Phases 1 et 2)
34. There's nothing to be ashamed of. (Phases 1, 2 et 3)
35. Are you going to do it? (Phase 1)
36. Catherine doesn't like cabbage anymore. (Phases 1 et 2)
37. For our annual Halloween party, we'll allow all costumes. (Phases 1, 2 et 3)
38. Everybody enjoyed the party last night. (Phase 1)
39. Mary's Mummy sang a lullaby to her baby. (Phases 1, 2 et 3)
40. The city hall was occupied by terrorists. (Phases 1, 2 et 3)

3.4.2. Deuxième partie de l'expérience

1. Kevin's mother often reminds him to take his camera.
2. « Hello! I'd like to inquire about accommodation. »
3. Let's focus on this absolutely amazing scenery.
4. There was an assumption that he suffocated her with a pillow.
5. Does anybody around here fancy some jacket potatoes?
6. For our annual Halloween party, we'll allow any costume.
7. Yesterday, we eventually managed to watch Harry Potter and Braveheart.
8. Mary's Mummy is singing a lullaby to her baby.
9. The passenger picked up some lilies, daisies and daffodils.
10. The enormous animal vanished in the forest.
11. Everyone congratulated the Canary population for their peaceful independence.
12. He apologised for his unacceptable behaviour.
13. Attention everyone, don't leave your luggage unattended or police officers will remove it.
14. The city hall was occupied by terrorists.
15. The happy fishermen were selling a great amount of anchovy.
16. They were judging Rachel for illegally selling tobacco.

17. It's got no added sugar in it.
 18. They visited Fountain's Abbey and they thoroughly enjoyed it.
 19. The famous novelist managed to finish his book about pottery.
 20. Catherine doesn't like cabbage anymore.
 21. There's nothing to be ashamed of.

3.5. *Matrices des effectifs*

3.5.1. Matrices des effectifs pour les regroupements ambisyllabiques

- Type de consonne intervocalique

	Ambisyllabiques	Non ambisyllabiques	TOTAL
Sonorantes	64	1434	1498
Obstruantes	22	2232	2254
TOTAL	86	3666	3752

- Patron accentuel

Matrice des occurrences en position pré- et post-accentuelles (en tenant compte uniquement des accents primaires)

	Ambisyllabiques	Non ambisyllabiques	TOTAL
1-0	56	2002	2058
0-1	12	940	952
TOTAL	86	4030	4116

Matrice des occurrences en position pré- et post-accentuelles (en incluant les accents secondaires)

	Ambisyllabiques	Non ambisyllabiques	TOTAL
Post-accentuel	67	968	1035
Pré-accentuel	12	2187	2199
TOTAL	79	3155	3234

- **Complexité orthographique de la consonne intervocalique**

	Ambisyllabiques	Non ambisyllabiques	TOTAL
C simple	31	2153	2184
C double	51	843	894
TOTAL	82	2996	3078

- **Voyelle précédente**

	Ambisyllabiques	Non ambisyllabiques	TOTAL
V courte	82	3516	3598
V longue	4	612	616
TOTAL	86	4128	4214

3.5.2. Matrices des effectifs pour les regroupements codaïques

- **Type de consonne intervocalique**

	Codaïques	Non codaïques	TOTAL
Sonorantes	183	1315	1498
Obstruantes	204	2050	2254
TOTAL	387	3365	1752

- **Patron accentuel**

Matrice des occurrences en position pré- et post-accentuelles (en tenant compte uniquement des accents primaires)

	Codaïques	Non codaïques	TOTAL
1-0	253	1805	2058
0-1	75	877	952
TOTAL	387	3729	4116

Matrice des occurrences en position pré- et post-accentuelles (en incluant les accents secondaires)

	Codaïques	Non codaïques	TOTAL
Post-accentuel	274	1980	2254
Pré-accentuel	80	900	980
TOTAL	354	2880	3234

- **Complexité orthographique de la consonne intervocalique**

	Codaïques	Non codaïques	TOTAL
C simple	171	2013	2184
C double	170	724	894
C mixte	46	878	924
TOTAL	387	3615	4002

- **Voyelle précédente**

	Codaïques	Non codaïques	TOTAL
V courte	363	3235	3598
V longue	24	592	616
TOTAL	387	3827	4214

3.6. Résultats des χ^2

3.6.1. Regroupements ambisyllabiques

- **Type de consonne intervocalique : sonorantes vs obstruantes**

```
> x<-matrix(c(64,22,1434,2232),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction
data: x
X-squared = 42.2018, df = 1, p-value = 8.232e-11

- **Patron accentuel**

Opposition "1-0" vs "0-1" (uniquement les accents primaires)

```
> x<-matrix(c(56,12,2002,940),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction
data: x
X-squared = 5.6445, df = 1, p-value = 0.01751

Opposition pré et post-accentuel en incluant les accents secondaires

```
> x<-matrix(c(12,67,2187,968),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction
data: x
X-squared = 101.2936, df = 1, p-value = < 2.2e-16

- **Complexité orthographique de la consonne intervocalique**

Opposition C simple vs C double

```
> x<-matrix(c(31,51,2153,843),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction
data: x
X-squared = 43.285, df = 1, p-value = 4.732e-11

Trois types de complexité orthographique (simple, double, mixte)

```
> x<-matrix(c(31,51,4,2153,843,920),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test
data: x
X-squared = 72.2224, df = 2, p-value = < 2.2e-16

- **Voyelle précédente : voyelle courte vs voyelle longue**

```
> x<-matrix(c(82,4,3516,612),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction
data: x
X-squared = 6.1959, df = 1, p-value = 0.01280

3.6.2. Regroupements codaïques

- Type de consonne intervocalique : sonorantes vs obstruantes

```
> x<-matrix(c(183,204,1315,2050),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction
data: x
X-squared = 9.4101, df = 1, p-value = 0.002158

- Patron accentuel

Opposition "1-0" vs "0-1" (uniquement les accents primaires)

```
> x<-matrix(c(253,75,1805,877),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction
data: x
X-squared = 12.6183, df = 1, p-value = 0.000382

Opposition pré et post-accentuel en incluant les accents secondaires

```
> x<-matrix(c(80,274,900,1980),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction
data: x
X-squared = 10.7654, df = 1, p-value = 0.001034

- Complexité orthographique de la consonne intervocalique

Opposition C simple vs C double

```
> x<-matrix(c(171,170,2013,724),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction
data: x
X-squared = 79.4395, df = 1, p-value = < 2.2e-16

Trois types de complexité orthographique (simple, double, mixte)

```
> x<-matrix(c(171,170,46,2013,724,878),nc=2)
> chisq.test(x)
```

Pearson's Chi-squared test
data: x
X-squared = 121.1426, df = 2, p-value = < 2.2e-16

- **Voyelle précédente : voyelle courte vs voyelle longue**

```
> x<-matrix(c(363,24,3235,592),nc=2)
```

```
> chisq.test(x)
```

Pearson's Chi-squared test with Yates' continuity correction

data: x

X-squared = 23.4482, df = 1, p-value = 1.283e-06

CONTENU DES CD-ROM







CD-ROM 1. La base de données Aix-MARSEC

Audio

408 fichiers-son du corpus Aix-MARSEC
















Dictionnaires

Dictionnaires utilisés pour la phonétisation :

-  Dictionnaires
 -  ALD.txt (dictionnaire général)
 -  mots_compl.txt (liste de formes inconnues complémentaires)
 -  formes_réduites.txt (liste des formes réduites)
 -  formes_pb.txt (liste des formes problématiques)
 -  mots_inacc.txt (liste des mots toujours inaccentués)

Scripts

Scripts de traitement d'Aix-MARSEC et de mise en forme des données tabulaires :

-  Scripts
 -  01_mtomultitxtgrid_v3.pl
 -  02_txtgridtomatrix_v4.pl
 -  03_mtx+pron_inacc_v6.pl
 -  04_mtx+prontopho_v16.pl
 -  05_phototxtgrid_inacc.pl
 -  06_a_add_Aber.pl
 -  06_b_add_Jassem.pl
 -  07_mergetiers.praat
 -  08_to_txt.pl
 -  09_re_align_v6.pl
 -  10_momel_mergetiers.praat
 -  11_add_text.pl.
 -  mise-en-forme.pl
 -  addindices.pl

TextGrid

408 fichiers-étiquette du corpus Aix-MARSEC

données.zip

Données tabulaires zippées à partir desquelles les tests sont effectués.





thèse.pdf

Version pdf de la thèse.

CD-ROM 2. Expérience en production











Données

Tableaux (Excel) de données :

-  Données
 -  Tableau1(global).xls
 -  Tableau2(ambi).xls
 -  Tableau3(coda).xls

Enregistrements







Fichiers-son et fichiers-étiquette classés par locuteur (locuteurs A-I) :


-  Son
 -  part1
 -  part2
-  TextGrid
 -  TextGrid_auto
 -  part1
 -  part2
 -  TextGrid_man
 -  part1
 -  part2

Le dossier ‘TextGrid_auto’ regroupe les fichiers-étiquette en sortie du script de détection automatique des pauses. Le dossier ‘TextGrid_man’ contient les fichiers-étiquette manuels.

Présentations

Deux présentations (PowerPoint) pour l’anglais (présentées aux sujets) et une pour le français (phase test) :











-  Présentations
 -  anglais
 -  part1.ppt
 -  part2.ppt
 -  français
 -  test_français.ppt

 ***get_pauses_v2.6.praat (script de détection des pauses)***

CD-ROM 3. Expérience en production (suite)

Enregistrements

Fichiers-son et fichiers-étiquettes classés par locuteur (locuteurs J-N) :

-  Son
 -  part1
 -  part2
-  TextGrid
 -  TextGrid_auto
 -  part1
 -  part2
 -  TextGrid_man
 -  part1
 -  part2