



ÉVALUATION QUALITATIVE DE LA PROSODIE D'APPRENANTS FRANÇAIS: APPORT DE PARAMÉTRISATION PROSODIQUES

Anne Tortel

► To cite this version:

Anne Tortel. ÉVALUATION QUALITATIVE DE LA PROSODIE D'APPRENANTS FRANÇAIS: APPORT DE PARAMÉTRISATION PROSODIQUES. Linguistique. Université de Provence - Aix-Marseille I, 2009. Français. NNT: . tel-00455248

HAL Id: tel-00455248

<https://theses.hal.science/tel-00455248>

Submitted on 9 Feb 2010

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

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

DOCTORAT Aix-Marseille Université – Université de Provence

U.F.R. LACS - Formation Doctorale : Cognition, Langage, Education

Discipline : PHONETIQUE ANGLAISE

N° bibliothèque :

THESE
pour obtenir le grade de
DOCTEUR D'AIX-MARSEILLE UNIVERSITÉ

Présentée et soutenue publiquement le 7 décembre 2009 par

Anne TORTEL

**EVALUATION QUALITATIVE
DE LA PROSODIE D'APPRENANTS FRANCAIS :
APPORT DE PARAMETRISATIONS PROSODIQUES**

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

JURY

RAPPORTEURS :

Monsieur Philip Carr, Professeur, Université de Montpellier III

Monsieur Jean-Yves Dommergues, Professeur, Université de Paris VIII, St Denis

EXAMINATEURS :

Monsieur Nicolas Ballier, Professeur, Université de Paris VII

Madame Nadine Herry-Benit, Maître de conférences, Université de Paris VIII, St Denis

Monsieur Laurence White, Research fellow, Université de Bristol, UK

DOCTORAT Aix-Marseille Université - Université de Provence

U.F.R. LACS - Formation Doctorale : LANGAGE ET PAROLE
Discipline : PHONETIQUE ANGLAISE

THESE
pour obtenir le grade de
DOCTEUR D'AIX-MARSEILLE UNIVERSITE

Présentée et soutenue publiquement le 7 décembre 2009 par

Anne TORTEL

**EVALUATION QUALITATIVE
DE LA PROSODIE D'APPRENANTS FRANCAIS :
APPORT DE PARAMETRISATIONS PROSODIQUES**

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

JURY

RAPPORTEURS :

Monsieur Philip Carr, Professeur, Université de Montpellier III
Monsieur Jean-Yves Dommergues, Professeur, Université de Paris VIII, St Denis

EXAMINATEURS :

Monsieur Nicolas Ballier, Professeur, Université de Paris VII,
Madame Nadine Herry-Benit, Maître de conférences, Université de Paris VIII, St Denis
Monsieur Laurence White, Research fellow, Université de Bristol, UK

A mon oncle, qui aurait sûrement aimé « voir ça »,

A mes parents,

A Chacha...

Remerciements

Je tiens, parmi toutes les personnes qui ont contribué à l'aboutissement de ce travail de thèse, à remercier tout particulièrement :

Les membres du jury,

Jean-Yves Dommergues et Phil Carr, qui ont accepté, en tant que rapporteurs, de lire et de juger ce travail dans un court délai, en dépit du surplus de travail que cela implique ;

Mes examinateurs, Nadine Herry-Bénit, Nicolas Ballier et Laurence White, pour avoir manifesté leur intérêt pour ce travail, et l'avoir enrichi par de précieux conseils ;

Mon directeur, Daniel Hirst, dont la recherche a inspiré ce travail. Merci pour tous tes conseils, ton soutien, ton écoute et ton humour irremplaçable qui m'ont accompagnée pendant ces années de recherche ;

Mes collègues,

Membres du LPL et leur directeur Philippe Blache, pour m'avoir accueillie au sein du laboratoire en m'offrant des conditions de travail plus qu'idéales. Merci tout particulièrement à : Alain, Stéphane, Thierry, Robert, Isabelle, Joëlle, Armelle, Bernard, pour votre aide et votre disponibilité de tous les instants ;

Cyril et Sébastien, pour m'avoir supportée pendant toutes ces heures d'enregistrements et pour avoir pris le risque de me laisser manipuler le matériel toute seule ! Merci pour l'aide indispensable que vous m'avez apportée et le temps que vous m'avez consacré ;

La bande de joyeux lurons, mes chers doctorants, Amandine, Céline, Francesco, Cécile, Léo, Pauline, Vincent, Céline, Haydee, Jean-Phi, sans qui ce travail n'aurait pas la même saveur ; et mes anglophones préférées, Kizzi et Oriana, un grand merci pour leur soutien ;

Mes sujets, pour l'abnégation avec laquelle ils ont prêté leur corps à la science ;

Mes collègues du B106 ! Valérie, Sophie, Mireille, Gabor, Isabelle, Peter, Laetitia, Quelle belle expérience vous m'avez offerte ! Vous m'avez permis de réaliser ma passion tout en me faisant confiance ;

Les secrétaires du département d'anglais, Monique, Sylvie, Sandrine, Jocelyne et leur boss Jean-François pour avoir accepté très courageusement de participer à mes tests d'anglais ;

Mes nouveaux collègues de Paris 8, pour l'accueil chaleureux et l'attention que j'ai reçus « débarquant de ma Provence ».

Laetitia et Lisa-Marie, les brunettes corses incontournables de mon parcours. Merci pour votre présence (même à distance !) et votre soutien ;

Mireille, pour ton écoute, ton attention, et ces innombrables papotages... merci pour ces nombreuses heures de sport qui nous ont permis de nous défouler avec acharnement;

Poupchou, miss script, pour ton aide à toute épreuve, tes bonnes soupes, et ces innombrables jus de fruits, rendez-vous immanquables de notre quotidien !

Amandine, mon allophone ! Il y en aurait tellement à dire ! Et ce n'est que le début ! ...fous rires, potins, colères, déprimés, joies, merci pour tous ces moments partagés ;

Mes amis, que j'ai grandement négligés dernièrement ;

Didou, sans toi cette thèse n'existerait pas. Tu es l'essence de ce travail, m'ayant envoutée de phonétique ! Merci pour ton aide inestimable, tes conseils, ton soutien et pour m'avoir transmis tellement de force et d'ondes positives. Tu as toujours gardé un œil sur moi ; tu as suivi chacun de mes pas depuis ces dix dernières années. Merci pour ces échanges si riches. Merci pour avoir toujours été là quand il le fallait ;

Et enfin, Mes frères & Mes parents, qui ont toujours cru en moi, qui ont toujours tout mis en œuvre pour me permettre de réussir. Votre dévotion et votre réconfort m'ont portée jusqu'au bout de ce travail. Vous avez fait bien plus que ce qui était possible pour que j'en sois ici aujourd'hui.

Table des matières

REMERCIEMENTS.....	5
TABLE DES MATIERES	7
TABLE DES FIGURES.....	13
TABLE DES TABLEAUX.....	15
AVANT-PROPOS	16
INTRODUCTION	18
CHAPITRE 1. INTRODUCTION A LA NOTION DE RYTHME & CLASSIFICATION	
RYTHMIQUE	26
1.1 PRESENTATION GENERALE	28
1.1.1 INTRODUCTION ET DEFINITION.....	28
1.1.2 RYTHME : SYSTEME ORGANISATEUR DE LA PROSODIE	30
1.1.3 LE RYTHME : LA BETE NOIRE DE L'APPRENANT ?.....	32
1.2 LE RYTHME DE LA PAROLE	34
1.2.1 ISOCHRONIE	35
... <i>plus ou moins isochrone</i>	36
<i>D'une isochronie pure à une isochronie subjective</i>	37
1.2.2 CLASSIFICATION DES LANGUES : EVOLUTION DU CONCEPT	40
<i>Théorie des langues isoaccentuelles/isosyllabiques</i>	40
<i>Variantes de la classification.</i>	43
... <i>vers un continuum rythmique</i>	44
<i>Coexistence rythmique : études inter-dialectales</i>	47
1.2.3 PROPOSITIONS DE METRIQUES RYTHMIQUES	48
<i>Ramus & al. (1999)</i>	48
<i>Low & al. (2000)</i>	50
<i>Remarques comparatives et limites des deux études proposées</i>	55

<i>Métriques rythmiques et tempo: nouvelles propositions</i>	<i>56</i>
<i>Application des métriques: discrimination L1/L2</i>	<i>58</i>
SYNTHESE.....	62
CHAPITRE 2. ACQUISITION D'UNE LANGUE	65
2.1 INTRODUCTION GENERALE.....	67
2.1.1 LANGAGE : DEFINITIONS	67
2.1.2 ORIGINE DU RYTHME DANS LE LANGAGE : HORLOGE INTERNE	68
2.2 ACQUISITION DE L'ANGLAIS LANGUE MATERNELLE (L1).....	69
2.2.1 INTRODUCTION GENERALE : LE STATUT DE LA PROSODIE DANS L'ACQUISITION DU LANGAGE	69
2.2.2 ACQUISITION DU RYTHME DE LA LANGUE MATERNELLE.....	70
<i>Perception rythmique du fœtus</i>	<i>71</i>
▪ Eléments neurophysiologiques	71
▪ Approche expérimentale	72
<i>Emergences de structures sonores</i>	<i>73</i>
▪ Rythme dans les cris infantiles.....	73
▪ Perception du nouveau né.....	74
<i>Vocalisation et babillage.....</i>	<i>77</i>
▪ Renforcement de la structure rythmique	77
▪ Prosodie : structure d'accueil.....	79
<i>Acquisition de la structure rythmique</i>	<i>79</i>
▪ Acquisition de l'accent	80
▪ Indices prosodiques à la structuration des énoncés	81
2.3 ACQUISITION DE L'ANGLAIS LANGUE SECONDE (L2) CHEZ L'ADULTE.....	82
2.3.1 INTRODUCTION GENERALE : LE STATUT DE LA PROSODIE DANS L'APPRENTISSAGE D'UNE LANGUE SECONDE	82
2.4 ACQUISITION DU RYTHME DE LA LANGUE SECONDE.....	86

<i>Analyse et constat.....</i>	86
<i>Approches pédagogiques dans l'enseignement L2</i>	87
▪ Approche académique	87
▪ Approche quasi –traditionnelle.....	87
▪ Travail de la prononciation par le rythme	88
▪ Approches non conventionnelles	89
2.5 PROBLEMES PARTICULIERS LIES A L'APPRENANT.....	90
<i>Contraintes extra-linguistiques</i>	90
<i>Contraintes physiologiques</i>	91
<i>L1 source d'influence de la L2</i>	92
▪ Crible et Interférence	92
▪ Transfert et Interlangue.....	93
▪ Interlangue et transfert prosodique	96
<i>Approches électroniques non conventionnelles : la perception au cœur du problème.....</i>	98
▪ SUVAGs : Système Universel-Verbo-Auditif-Guberina	99
▪ « L'oreille électronique ».....	100
▪ I-LANG : approche neurolinguistique	101
SYNTHESE.....	105
CHAPITRE 3. PROSODIE ET MULTIMEDIA : PROBLEMES D'EVALUATION	107
3.1 LES LOGICIELS D'APPRENTISSAGE DE LA PROSODIE DE L'ANGLAIS	109
3.1.1 INTRODUCTION	109
3.1.2 CONSTAT	110
<i>Historique sommaire de l'évolution de la visualisation de l'intonation.....</i>	110
▪ Premières tentatives	110
▪ Visualisation de la courbe mélodique	111

▪ Nouvelles technologies : phonétique et multimédia.....	111
3.2 APPORTS DES LOGICIELS DE LANGUES.....	124
<i>Utilité dans l'apprentissage de la prosodie.....</i>	<i>124</i>
<i>L'apprenant au cœur de l'outil.....</i>	<i>125</i>
<i>Feedback pédagogique.....</i>	<i>126</i>
<i>Limites de l'outil multimédia.....</i>	<i>128</i>
3.3 L'EVALUATION PROSODIQUE EN L2	129
<i>Définitions et concepts</i>	<i>129</i>
<i>... propositions de critères ?.....</i>	<i>131</i>
<i>La synthèse de parole</i>	<i>136</i>
3.4 CONCLUSION	136
SYNTHESE.....	138
CHAPITRE 4. CONSTITUTION ET EXPLOITATION D'UNE BASE DE DONNEES : ENGLISH	143
4.1 REMARQUES PRELIMINAIRES.....	145
4.2 ORIGINE DU CORPUS ENGLISH.....	148
4.2.1 CORPUS DE BASE : EUROM 1	148
4.2.2 DESCRIPTION DU CORPUS EUROM 1	148
<i>Eurom 1 : corpus anglais</i>	<i>149</i>
4.2.3 ENRICHISSEMENT DU CORPUS EUROM 1 (CORPUS ANGLAIS)	149
<i>Etat des lieux</i>	<i>149</i>
<i>Contributions personnelles : extension du corpus EUROM 1.....</i>	<i>150</i>
4.4 ENGLISH : CREATION D'UN CORPUS ANGLAIS L1 & L2.....	152
4.4.1 CHOIX, REFLEXIONS ET EXIGENCES	152
<i>ENGLISH : corpus, base de données ?</i>	<i>152</i>
<i>Parole de laboratoire vs parole spontanée vers parole naturelle</i>	<i>154</i>
<i>Focus sur l'oral</i>	<i>155</i>

<i>Variétés des genres.....</i>	<i>156</i>
<i>Variétés des niveaux d'apprentissage des locuteurs</i>	<i>156</i>
<i>Qualité technique du corpus.....</i>	<i>156</i>
4.4.2 ENGLISH : DESCRIPTION DU CORPUS	157
<i>De Eurom 1 à ENGLISH : choix des textes et préparation des stimuli</i>	<i>157</i>
<i>Contenu de la base de données.....</i>	<i>159</i>
▪ <i>Lecture.....</i>	<i>159</i>
▪ <i>Répétition</i>	<i>160</i>
▪ <i>Monologue.....</i>	<i>161</i>
<i>Locuteurs.....</i>	<i>162</i>
<i>Codage du corpus.....</i>	<i>163</i>
<i>Accessibilité du corpus ENGLISH.....</i>	<i>164</i>
4.4.3 TRANSCRIPTIONS DU CORPUS.....	164
<i>Transcription orthographique</i>	<i>164</i>
<i>Transcription phonémique.....</i>	<i>165</i>
<i>Alignement acoustique manuel.....</i>	<i>166</i>
4.5 PERSPECTIVES.....	168
SYNTHESE.....	170
CHAPITRE 5. ANALYSES EXPERIMENTALES	172
5.1 PROBLEMATIQUE ET OBJECTIFS.....	173
5.2 METHODOLOGIE.....	175
5.2.1 BASE DE DONNEES	175
5.2.2 CALCUL ET OBTENTION DES METRIQUES.....	177
5.3 ANALYSES STATISTIQUES.....	179
5.3.1 CARACTERISATION DES GROUPES DE 3 NIVEAUX	179
5.3.2 ANALYSES DES METRIQUES RYTHMIQUES : DISTINCTION L1/L2	182

▪ ΔV , ΔC , %V (Ramus & al., 1999).....	183
▪ PVI (Grabe & Low, 2002).....	186
▪ cvC,cvV (Dellwo, 2006 ; White & Mattys, 2007).	189
5.3.3 ESSAI DE PARAMETRISATION RYTHMIQUE : RECHERCHE DE COMBINAISONS PAR SVM	193
▪ Analyse à un facteur	194
▪ Analyse combinatoire à deux facteurs	195
▪ Analyse combinatoire à trois facteurs	196
▪ Analyse combinatoire à quatre facteurs	197
5.4 DISCUSSION ET CONCLUSION	198
5.4.1 INFLUENCE DE LA L1 SUR LA L2 ?	198
5.4.2 DISCRIMINATION A PARTIR DES CRITERES RYTHMIQUES	200
CONCLUSION ET PERSPECTIVES.....	205
RÉFÉRENCES	211

Table des figures

Figure 1. Axe représentant le degré d'importance de l'accent dans différentes langues (Dauer, 1983:60)	45
Figure 2. Valeurs du PVI pour les 18 langues (tiré de Grabe & Low, 2002:22)	54
Figure 3. Illustration des appareils SUVAGs en 1961 puis 2000	99
Figure 4. Illustration des modalités de la communication et fonctionnement cérébral.....	102
Figure 5. Exemple de représentation graphique proposée dans le logiciel SaRP	113
Figure 6. Proposition de feedback par affichage de l'oscillogramme (Voicebook, 1998)	116
Figure 7. Proposition de feedback par affichage de la courbe d'amplitude (Wincecil, 1997) ..	116
Figure 8. Feedback visuel des mots « 'present vs pre'sent » avec comparaison des productions natif/apprenant selon les contours intonatifs.....	119
Figure 9. Feedback visuel des mots « ' <i>present</i> vs <i>pre'sent</i> » avec comparaison des productions natif/apprenant (rythme/intensité).....	119
Figure 10. Illustration du feedback explicatif des points saillants à travailler par l'apprenant.	120
Figure 11. Exemple de feedback visuel du logiciel Winpitch LTL,.....	121
Figure 12. Exemple de feedback de la production d'un apprenant (en bas) comparée à celle d'un natif (en haut), http://www.lpl.univ-aix.fr/~lpldev/prosodia	122
Figure 13. Illustration d'une fenêtre du logiciel SWAMS.....	123
Figure 14. Exemple de TextGrid d'annotation extrait d'ANGLISTH effectué dans PRAAT ..	167
Figure 15. Exemple de segmentation en CVC.....	177
Figure 16. Représentation graphique de l'analyse discriminante des 1260 phrases produites par les 63 locuteurs	180

Figure 17. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan (ΔV , ΔC)	180
Figure 18. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan ($\%V$, ΔC)	182
Figure 19. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan ($rPVI_V$, $rPVI_C$)	184
Figure 20. Visualisation des points de dispersion (gauche) et de erreurs-types (droite) selon le plan ($nPVI_V$, $nPVI_C$).....	185
Figure 21. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan ($rPVI_C$, $nPVI_V$).....	186
Figure 22. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan (cvV , cvC).....	187
Figure 23. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan ($\%V$, cvV).....	188
Figure 24. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan ($\%V$, cvC).....	189
Figure 25. Visualisation des points de dispersions (gauche) et des erreurs-types (droite) selon le plan (ΔC , cvC).....	192

Table des tableaux

Tableau 1. Récapitulation des principales différences des propriétés structurelles rythmiques	34
Tableau 2. Classement des différentes langues selon leur classe rythmique.....	44
Tableau 3. Langues étudiées dans les travaux de Grabe & Low (tiré de Grabe & Low (2002: 13): « Traditional Rhythmic classifications of languages investigated in the present study ».....	51
Tableau 4. Répartition des enregistrements des passages par locuteur (corpus EUROM 1) ..	149
Tableau 5. Récapitulation des différentes métriques utilisées pour notre étude.....	174
Tableau 6. Méthode de calcul des 9 paramètres rythmiques	178
Tableau 7. Extrait des données de sortie obtenues par le script "calculate_rhythm_metrics"	179
Tableau 8. Présentation des résultats des prédictions des différentes classes (FR1/FR2/GB), (en haut, valeurs brutes ; en bas, valeurs exprimées en pourcentage)	181
Tableau 9. Présentation des données de sortie d'une analyse SVM - une discrimination à un paramètre.....	194
Tableau 10. Présentation des résultats des meilleures combinaisons à deux facteurs	195
Tableau 11. Présentation des résultats des meilleures combinaisons à trois facteurs	196
Tableau 12. Présentation des résultats des meilleures combinaisons à quatre facteurs	197
Tableau 13. Récapitulatif des différentes métriques présentées en fonction de la tâche de discrimination visée	201
Tableau 14. Récapitulatif des meilleures prédictions pour chacune des combinaisons factorielles	202
Tableau 15. Présentation des prédictions obtenues par l'analyse SVM des métriques exposées dans la littérature	203

AVANT-PROPOS

Nous faisons ici une première remarque concernant les conventions d'écriture puis par la suite nous éclaircissons quelques notions clés utilisées tout au long de ce travail.

Au cours de cette rédaction, nous avons choisi, pour une question de lisibilité et de respect des conventions, de ne citer que le premier auteur d'un travail dès lors que celui-ci impliquait plus de deux collaborateurs. La référence se présente alors dans le texte sous la forme de « 1^{er} auteur et al. » (date) [...].

La seconde démarche de cet avant-propos consiste à clarifier la terminologie communément utilisée dans ce travail.

- Acquisition/ apprentissage

L'origine latine du mot « acquérir » renvoie à « *acquaerere* » qui signifie « *ajouter à ce qui existe déjà* » (Le Robert, 2000 :28). Krashen (1981) distingue deux champs notionnels qu'il désigne comme « *second-language acquisition* » et « *second-language learning* ». Il définit l'acquisition comme le processus inconscient qui mène à l'utilisation spontanée de la langue, et l'apprentissage comme le processus conscient qui conduit à la connaissance des règles d'une langue. A partir de cette distinction, nous utilisons le terme « acquisition » lorsque le contexte est celui de la langue maternelle ; en revanche, dans le cadre de l'apprentissage d'une langue étrangère, nous utilisons respectivement les deux termes car nous pensons que les deux notions se rejoignent au niveau du résultat obtenu, dans la mesure où l'apprentissage (guidé ou non guidé) représente le moyen d'atteindre le stade d'acquisition.

- Langue étrangère/ langue seconde

Dans ce travail de thèse, les deux vocables sont utilisés pareillement. Néanmoins, notons que les termes « langue étrangère » et « langue seconde » se distinguent l'un de l'autre et s'utilisent selon un contexte précis. Cuq (2003) définit la langue seconde comme une langue parlée à l'étranger et possédant un statut particulier. Par exemple, pour l'anglais, on pourra citer le cas des pays africains où l'anglais n'est pas la langue maternelle, mais son usage a une fonction principalement administrative, l'anglais n'étant

ni la langue maternelle ni une langue étrangère comme en France. Il en est de même, par exemple, pour la langue française dans les pays du Maghreb. La langue seconde est utilisée comme langue d'enseignement à partir d'un certain niveau et permet l'accession à un niveau social plus élevé. Par contraste, la langue étrangère est une langue non maternelle, apprise en milieu scolaire. Appliqué à la situation des apprenants francophones adultes, l'apprentissage est à la fois guidé (cadre scolaire, universitaire, associatif) et libre (discussions avec des natifs, visionnage de films en version originale etc.). La distinction entre ces deux termes n'est pas pertinente.

- L1/L2

Nous utilisons les abréviations L1 et L2. L1 correspond à la langue maternelle ou la langue première, tandis que L2 est utilisée pour référer à la langue seconde, langue étrangère ou encore la langue cible.

- Intelligibilité/compréhensibilité

Nous distinguons les deux termes de la manière suivante : l'intelligibilité représente la capacité à reconnaître, distinguer un mot, une phrase, tandis que la compréhensibilité indique la capacité à capter la signification de ce mot ou de cette phrase. Ainsi selon Field (2005), l'intelligibilité est perçue comme une qualité première d'une production d'un apprenant L2, nécessaire afin d'atteindre un niveau aisé de compréhensibilité du discours par l'interlocuteur. Bien souvent, ces deux termes ne sont pas distingués et apparaissent comme synonymes. Dans cette étude, nous considérons que le terme intelligibilité est un facteur fondamental qui facilite la compréhensibilité du discours.

Introduction

Le but premier pour tout apprenant d'une L2 est, bien évidemment, de pouvoir s'exprimer, mais surtout de se faire comprendre par les locuteurs natifs de cette langue. Or un accent étranger peut quelquefois faire entrave à cet objectif. Plusieurs études (Anderson-Hsieh *et al.*, 1992 ; Munro et Derwing, 1995, 1998 ; Jilka, 2000 ; Hardison, 2004) ont montré que la prosodie jouait un rôle important dans l'intelligibilité et la compréhensibilité du discours non natif. Cela ne semble pas surprenant car, comme le montre Konopczynski (1999) par son modèle d'Intonologie Développementale Interactive (I.D.I.)¹, toute imprégnation naturelle avec une langue étrangère se fait d'abord par ses éléments prosodiques qui constituent une véritable « structure d'accueil de tous les autres éléments de la parole » (Konopczynski, 1999 :68). Le système phonologique vient donc s'organiser à l'intérieur des schémas rythmiques et intonatifs. Cependant, avec l'effet de l'âge, il devient difficile pour l'apprenant que ses productions L2 ne soient pas influencées par sa langue maternelle. La L1 apparaît comme source d'influence et provoque des interférences auditives avec la L2. Le contact (Lado, 1957) entre les deux langues (L1 & L2) est rendu responsable des erreurs observables chez les apprenants, en raison des transferts qu'ils opèrent de la langue maternelle sur la langue cible. Dès lors, plus les mauvaises habitudes sont ancrées, plus la tâche d'apprentissage de la prosodie devient ardue. Ces phénomènes étant inconscients, il est important de les extérioriser et de les rendre observables et accessibles pour l'apprenant. Les technologies multimédia pourraient être d'une grande utilité dans la rééducation prosodique ; elles semblent être un outil adapté pour un travail de l'oral, alliant son et image, permettant à l'utilisateur de prendre conscience de ses différences de production avec celles d'un natif. Bongaerts (1999) note plusieurs facteurs pouvant aider à l'amélioration des productions des locuteurs L2 :

- (i) une grande motivation pour réaliser 'naturellement' les accents lexicaux
- (ii) un accès illimité à l'écoute de la langue cible

¹ Dérivé du concept d'Intonologie Développementale introduit par Crystal (1976), IDI introduit les aspects psycho-phonologiques et prosodiques de la conversation adulte-enfant dans une perspective développementale. C'est une analyse qui prend en compte les productions de l'enfant, de l'adulte et leurs effets sur le comportement vocal et verbal de l'enfant et sur son évolution.

(iii) un entraînement intense à la perception et production de la L2.

Ces caractéristiques conduisent de nouveau à penser que les technologies multimédia constituent un moyen idéal à l'entraînement et à l'amélioration des productions L2 (Hincks & Edlund, 2009 ; Tanner & Landon, 2009).

Depuis quelques années maintenant, les logiciels de langue sont en nombre croissant. Un intérêt grandissant a été accordé à ces nouvelles méthodes multimédia grâce aux progrès importants réalisés dans les technologies de la parole. Nombreux sont les logiciels qui se focalisent sur l'entraînement de la prononciation au niveau segmental, mais un petit nombre accorde une place aux caractéristiques prosodiques ainsi qu'à leur évaluation. Les tâches à réaliser se concentrent sur la syllabe, le syntagme, la phrase étant l'unité maximale proposée. La composante qu'est l'évaluation représente le point central des critiques de ces nouveaux outils multimédia et peu d'études ont été réalisées à ce sujet. En effet, lorsqu'on s'intéresse de près à ces logiciels, le premier constat est le manque évident de feedback, basé sur des critères acoustiques objectifs, qui pourrait aider l'utilisateur à progresser et corriger ses erreurs. Leur intérêt pédagogique est alors vivement remis en question (Cazade, 1999 ; Santacroce, 2002), mettant en avant le manque d'évaluation proposée à l'apprenant qui se voit confronté à l'interprétation d'oscillogrammes ou de courbes mélodiques difficilement interprétables sans l'aide d'un spécialiste. Neri *et al.* (2002) observent qu'une majeure partie du problème vient du fait que, bien souvent, beaucoup de logiciels, produits commerciaux ont favorisé et opté pour les nouveautés technologiques au détriment de la valeur pédagogique. De nombreuses perspectives de recherche sont donc à envisager afin de pallier l'absence de feedback.

Problématique :

C'est ce manque d'études portant sur l'évaluation de critères prosodiques objectifs dans l'apprentissage de l'anglais L2 par des apprenants français qui a motivé le choix de cette recherche. Par ailleurs, une seconde motivation à l'origine de ces travaux repose sur l'idée que la prosodie doit être à la base de l'enseignement d'une langue seconde, et non en arrière-plan, celle-ci étant le pivot du langage verbal. L'enseignement des structures

prosodiques doit se faire simultanément à celui des traits segmentaux et ne devrait pas être « optionnel », contrairement à ce qui se pratique à l'heure actuelle.

Le but de la présente thèse est l'élaboration d'un système d'évaluation des traits prosodiques des francophones qui parlent anglais afin d'en dégager des paramètres acoustiques objectifs intégrables dans un logiciel de langue. Si ce champ d'études suscite de plus en plus d'intérêt (Dodane & Konopczynski, (2001) pour le FLE, Minematsu & al., (1999), entre autres pour les apprenants japonais), peu d'études (Herry & Hirst, 2002 ; Poquin, 2009) ont été menées concernant les francophones dont la L2 est l'anglais. Ce domaine ouvre la voie à de nombreux travaux. Les travaux proposés dans cette recherche tentent de relier les recherches menées en prosodie avec l'enseignement de celle-ci en L2.

Afin de cibler les traits prosodiques erronés des apprenants, il est indispensable d'exploiter et d'appuyer nos travaux sur une base de données orales comparatives de francophones parlant anglais et de natifs anglophones. Or, à notre connaissance, une telle base de données ne semble pas exister ou est difficilement accessible librement par la communauté. Un premier problème se pose alors à nous : constituer une telle base de données, suffisamment conséquente, afin de pouvoir par la suite, à partir des données recueillies, élaborer des critères d'évaluation prosodique.

Objectifs :

Les objectifs de cette recherche vont donc être (i) en amont de propositions de critères évaluatifs, l'élaboration d'une importante base de données d'anglais L1 & L2, (ii) à partir de laquelle pourra être établie la recherche de paramètres acoustiques prosodiques.

Le cadre d'étude pour cette recherche concerne donc la langue maternelle, le français, et la langue en apprentissage, l'anglais, dans une situation où (i) le déséquilibre, entre le système maîtrisé et le système en construction, reste conséquent (groupe d'adultes de niveau faible) et (ii) l'écart entre les deux systèmes n'est pas aussi important (groupe d'apprenants universitaires anglicistes). La méthode utilisée est une méthode comparative des traits prosodiques par la constitution d'un groupe contrôle de natifs anglophones. Le sujet d'étude cible donc l'apprenant francophone en parcours universitaire, ainsi que des français adultes, désireux de « se remettre » à l'anglais. L'apprenant se situe au cœur de cette recherche, car, sans lui ce travail n'aurait pas lieu d'être, mais c'est aussi pour lui que

ce travail existe. En tant qu'enseignant, il est aisé d'observer la difficulté face à laquelle l'apprenant français est confronté lorsqu'il parle dans une autre langue. La gêne et la honte qui sont éprouvées à l'idée de s'exprimer dans une langue étrangère empêchent bien souvent d'établir un échange. La tâche de l'apprenant est d'entrer dans une organisation prosodique étrangère et d'être capable de se créer une infrastructure différente de celle utilisée dans sa langue maternelle.

Enfin, l'objet de cette recherche est la prosodie², et plus particulièrement le rythme des productions de francophones en anglais. L'étude des éléments suprasegmentaux de la réalisation des apprenants permettra de cibler les infractions des traits prosodiques qui entraînent une distorsion caractéristique d'un accent étranger, ou qui sont à l'origine d'une incompréhension dans le discours. Le choix des paramètres acoustiques rythmiques retenus, basés sur le développement de nouvelles métriques, est explicité dans le chapitre 1. Enfin, l'absence de prise en compte des traits intonatifs dans ce travail ne signifie en aucun cas qu'ils sont moins importants et ou qu'ils sont tout simplement ignorés. Au contraire, nous pensons que les structures intonatives et rythmiques (ainsi que les autres composants tels que les pauses, le tempo etc.) sont profondément associées pour former un tout qu'est la prosodie. Cependant, l'intonation fait appel à un autre niveau de structuration qui ne fait pas l'objet de cette recherche mais nécessite d'être également exploitée pour compléter les paramètres évaluatifs observables.

Plan de la thèse :

Ce travail de recherche se scinde en deux grandes parties : la section 1, qui traite des aspects théoriques de notre sujet et la section 2, qui présente les études expérimentales menées ainsi que les résultats obtenus.

² La prosodie est définie comme une branche de la linguistique consacrée à la description (aspect phonétique) et à la représentation formelle (aspect phonologique) des éléments de l'expression orale tels que les accents, les tons, l'intonation et la quantité, dont la manifestation concrète, dans la production de la parole, est associée aux variations de la fréquence fondamentale (f0), de la durée et de l'intensité (paramètres prosodiques physiques), ces variations étant perçues par l'auditeur comme des changements de hauteur (ou de mélodie), de longueur et de sonie (paramètres prosodiques subjectifs). La prosodie permet de distinguer l'oral de l'écrit (Di Cristo, 2000).

Le chapitre 1 est consacré à la notion de rythme, dans lequel sont proposées une définition et une description de cette notion. Puis, au travers des deux langues étudiées, sont exposées une présentation de l'historique de la classification des langues ainsi que les dernières études menées dans ce cadre. Cela nous permet d'étudier le fonctionnement des deux langues en question, mettant en avant les principales différences qui les opposent et les caractéristiques propres à chaque langue ; nous pourrions ainsi fonder les hypothèses énoncées dans l'analyse expérimentale.

Le chapitre 2 présente les différentes étapes de l'acquisition du rythme de la langue maternelle dans la toute première période de vie du nourrisson, puis, dans une seconde partie nous aborderons les différents moyens mis en œuvre dans le cadre de l'enseignement de la prosodie, et particulièrement du rythme. Ce chapitre permet d'exposer l'acquisition du système prosodique de la langue maternelle et ses implications pour l'apprentissage d'une L2, et ainsi de comprendre les difficultés auxquelles l'apprenant est confronté lors de l'apprentissage d'une autre langue.

Le chapitre 3 propose un état des lieux des différents outils multimédia disponibles actuellement, incluant les systèmes d'évaluation proposés lors de l'utilisation de ces logiciels. Sont également présentés les derniers travaux sur l'évaluation de l'anglais, récemment entrepris chez les apprenants non francophones. C'est à partir du constat négatif d'un manque évident de feedback que la proposition des critères évaluatifs objectifs est nécessaire.

Le chapitre 4 détaille les différentes étapes qui ont abouti à la constitution d'une base de données orales de l'anglais L1 & L2, nommée ANGLISH. Sont exposés le protocole d'expérimentation ainsi que les différentes annotations et segmentations apportées aux données. Ce corpus d'apprenants nous permet d'évaluer les différences rythmiques entre les apprenants français et les natifs anglophones.

Le chapitre 5 définit le protocole expérimental des analyses des facteurs rythmiques basées sur les différentes métriques (exposées dans le chapitre 1) et présente les différents résultats obtenus pour l'élaboration de paramètres évaluatifs.

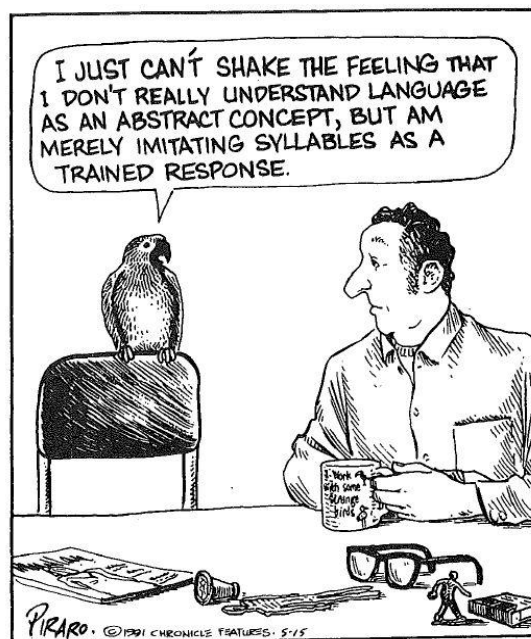
Section 1.

Aspects théoriques

CHAPITRE 1. INTRODUCTION A LA NOTION DE RYTHME & CLASSIFICATION RYTHMIQUE

*« Seul le rythme provoque le court-circuit poétique
et transmue le cuivre en or, la parole en verbe »*

*Ethiopiennes, Léopold Sédar Senghor
(post-face, p.111, Seuil)*



Copyright © 1991 by Chronicle Features.

1991 by Chronicle Features, from Fromkin & Rodman (1998)

Ce premier chapitre est consacré au cadre général dans lequel s'inscrit ce travail. La première partie de ce chapitre propose une familiarisation à la notion de rythme et tente de dégager une définition du rythme (la tâche n'étant pas aisée étant donné les divergences de points de vue et la complexité du phénomène à rendre compte) en indiquant par ailleurs sa place et fonction au sein de la prosodie. Dans un second temps, nous abordons le concept d'isochronie qui conduit au débat de la classification des langues en fonction de leurs caractéristiques rythmiques, fortement remis en question dans la littérature. Les approches alternatives récentes sont alors décrites et serviront de point d'ancrage à la partie expérimentale.

1.1 Présentation générale

Pour aborder cette notion de rythme, il est proposé dans cette partie une présentation générale en essayant de proposer une définition, de resituer la place du rythme au sein de la prosodie et enfin de dégager les principales fonctions de cet élément.

1.1.1 Introduction et définition

Le terme « rythme » est souvent attribué à la notion de musicalité. Ainsi, il est fréquent d'entendre « tu n'es pas en rythme » ou encore l'expression « avoir le rythme dans la peau ». Au-delà de ces expressions et bien plus que la relation rythme/musique, il est facilement constatable que la notion de rythme est présente tout autour de nous. En effet, le rythme se manifeste dans des multitudes d'activités chez l'être humain. Notre quotidien est rythmé par le lever, une prise de nourriture trois fois par jour, le coucher ; organisation elle-même rythmée par un cycle régulier du jour et de la nuit. Tous ces phénomènes, et bien d'autres encore, sont structurés par un rythme, que ce soit au niveau artistique (musique, danse, poésie), biologique (rythme cardiaque, respiratoire) ou même linguistique. Handel (1989:383) remarque que le rythme existe à de nombreux niveaux :

« Rhythmic organisation is an inherent part of all human activity. Rhythm energizes structures, creates, and expresses temporal quality. It is an important in hammering a nail, or playing basketball as it is in speech, music, or dancing. Time is not merely the passive medium within which events occur. Rather, time acts to shape and determine all phenomena. Moreover, rhythmic organisation exists at all levels of activity ».

Il vient d'être mis en évidence que le rythme est présent sur diverses activités ou phénomènes comme si, sans rythme, l'être humain ne pouvait se réaliser. Mais qu'entend-on par rythme, comment définir le rythme et tout particulièrement dans le domaine de la linguistique ?

Etymologiquement parlant (Le Robert, Dictionnaire Historique de la Langue Française, 2000), le terme rythme vient du latin « *rythmus* » (mouvement, battement régulier, mesure, cadence), terme lui-même emprunté au grec « *rhuthmos* » qui fait référence à des valeurs analogiques de proportions régulières, d'ordonnement symétrique, et par extension, de manière d'être, s'appliquant à la forme d'une chose, à une espèce. Le mot est dérivé de « *rhein* » qui signifie « couler » et se rapporte à ce qui est

mouvant, modifiable ou qui correspondrait à un « arrangement des parties ». Après un emploi isolé en musique jusqu'au XVI^{ème} siècle, le rythme est défini comme « *un retour à intervalles réguliers des temps forts et des temps faibles dans un vers* » (Larousse, 2003). Au sens figuré, il est décrit comme une « *alternance, succession plus ou moins régulière que présentent certains événements : le rythme des saisons* ». Rapporté à la musique, le rythme est défini comme l'effet obtenu par la succession des temps forts et temps faibles ».

D'un point de vue biologique, l'exemple du rythme respiratoire peut être cité : il est défini comme « *une succession des inspirations et expirations* » (Larousse, 2003).

A partir de ces définitions, il devient alors possible de parler « des rythmes » et non plus « du rythme » étant donné son rôle organisateur des divers phénomènes de la vie quotidienne. Un point commun se dégage entre toutes ces définitions : le rythme semble reposer sur une alternance d'un phénomène qui se répète selon le schéma fort/faible.

Lorsque l'on se penche sur l'aspect linguistique, cette même caractéristique de rythmicité binaire se retrouve dans le langage. D'après Liberman & Prince (1977), la parole serait rythmée selon la caractéristique du schéma : un battement fort suivi d'un battement faible (ou deux maximum) entre deux battements forts. C'est ce qu'ils appellent le Principe d'Alternance Rythmique (« Principle of Rhythmic Alternation »). Cependant, si nous nous en tenons à cette définition, il faut alors conclure que le rythme de la parole est une simple alternance de syllabes accentuées et inaccentuées. Or, la notion d'« intervalles réguliers » n'est pas à négliger car elle introduit une caractéristique temporelle importante. Platon lui-même définissait le rythme comme « l'ordre dans le mouvement » et soutenait que le rythme était une qualité intrinsèque à l'être humain, le différenciant de l'animal :

« [...] alors que le sens de l'ordre en ces deux domaines échappe complètement aux autres êtres vivants, seule la nature humaine peut l'acquérir » (Platon: 138)

Ainsi Caelen-Haumont (1983) définit-elle le rythme comme une structuration des durées. Au travers des différentes propositions de définitions exposées dans la littérature, il est à noter que les dimensions temporelles et accentuelles sont prises en compte. Dès lors, le rythme se base sur une structure temporelle sur laquelle se développent les variations intensives et tonales (Fraisse, 1956). Malgré tout, certains auteurs tels que Dell (1984), expliquent que la structure rythmique est constituée seulement de schèmes accentuels. Pour l'auteur, le schème accentuel représente une structure « abstraite » qui sert de « balisage »

de la chaîne syllabique et dont le statut est similaire à celui des barres de mesures de notation musicale.

Cette dimension accentuelle amène à s'interroger sur le lien rythme/accentuation ; en effet, est-ce le rythme qui entraîne la production d'accents ou sont-ce les accents qui imposent le rythme ? Selon Bolinger (1981), suivant la proposition de Pike (1945), il existe deux types de rythme : une rythmicité syllabique et une rythmicité accentuelle (durée, f_0 , intensité). Selon l'auteur, le rythme apparaît comme procédé accentuel. Ce n'est plus l'accentuation qui est procédé rythmique, contrairement à la définition de Di Cristo (2002) qui parle du rythme comme étant « l'organisation temporelle des syllabes métriquement fortes et des syllabes métriquement faibles ». Le phénomène accentuel est alors directement lié au rythme dans lequel le concept de métrique est introduit. Di Cristo (2002) décrit l'architecture prosodique comme « un supra-système phonique constitué de trois sous-systèmes » appelés métrique, tonal et temporel. L'auteur distingue alors mètre et rythme. L'organisation métrique correspond au niveau profond, c'est-à-dire à la structuration théorique des énoncés qui est représentée mentalement par le rythme.

Dans ce travail, nous adoptons la définition proposée par Di Cristo & Hirst (1997), selon laquelle le rythme est décrit comme « l'organisation temporelle des proéminences ». Ces proéminences seraient les éléments mis en relief par des moyens tonals, intensifs et temporels. Par cette définition, procédés accentuel et temporel sont réunis et le rythme devient révélateur de proéminences. Le rythme est donc un phénomène prosodique complexe qui agit comme principe générateur de structures ou de groupements plus ou moins récurrents, par la régularité ou l'alternance d'éléments saillants (Di Cristo & Hirst, 1993).

1.1.2 Rythme : système organisateur de la prosodie

La prosodie se compose de plusieurs éléments, souvent qualifiés de suprasegmentaux, du fait qu'ils s'étendent sur une durée qui dépasse celle des segments (phonématiques). Ces éléments englobent, entre autres, l'intonation, l'accentuation, le tempo, les pauses ; le rythme en fait partie. Ce dernier est en relation étroite avec de nombreux paramètres tels que le timbre, le débit, etc. et est lié à la syntaxe, la sémantique et la pragmatique (relations qui ont suscité de nombreux débats mais dont nous ne traiterons pas ici, cela n'étant pas l'objet de cette thèse). Astésano (2001) souligne que le

rythme participe à l'agencement d'unités dans le temps, qui peuvent être phonémiques, syllabiques, pausales, accentuelles ou bien encore mélodiques (etc.).

Il semble que la fonction majeure du rythme soit celle de la structuration de la chaîne parlée. Imaginez des instruments qui jouent la même harmonie, la même suite de notes, mais espacées au goût de chacun. Le résultat est une cacophonie. Si ces musiciens ajoutent la notion de temporalité, nous obtenons une musique. Il en va de même pour le langage. Fraisse (1956 :9) met en avant cette caractéristique :

« [...] l'organisation temporelle des proéminences qui participent de la structuration prosodique des énoncés ».

Handel (1983 :383) insiste sur la notion de temporalité, qui a pour fonction de créer et mettre en forme :

« Rhythm energizes structures, creates, and expresses temporal quality [...] rather than, time acts to shape and determine all phenomena ».

C'est donc le rythme qui va structurer et permettre, avec l'ensemble des éléments prosodiques, de mener à la musicalité du langage. Le rythme met donc en forme ce qui pourrait être « cacophonie » et amène à l'harmonisation des énoncés. Myers (1984) considère le rythme comme une capacité du système cognitif à « former » et à structurer l'information perçue. Cette idée de structuration se retrouve également chez Roach (1991 :10) qui met en évidence le rôle primordial de la structuration rythmique pour le langage et la communication :

« [it] 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 ».

Bolinger (1972) donne au rythme une fonction d'aide à l'accès au sens et à l'information. L'auteur met en relation accent et information : dans un énoncé proposé, le mot portant l'accent le plus fort porte l'information principale que le locuteur veut faire passer. Pour Beckman (1992), c'est par l'aspect métrique du rythme que va s'organiser la parole, en segmentant dans le continuum verbal des unités permettant d'accéder au sens.

Le rythme constitue donc un élément à part entière de la prosodie. Un rôle d'organisateur et d'accès au sens lui est attribué. Qu'il soit défini d'un point de vue de la

structure ou de la régularité, il peut être convenu que l'élément commun qui s'en dégage est bien l'occurrence de séquences d'événements.

1.1.3 Le rythme : la bête noire de l'apprenant ?

Lorsque l'on se penche sur les caractéristiques rythmiques du français et de l'anglais, un certain nombre de différences peuvent être relevées et pourraient expliquer les difficultés auxquelles les francophones sont confrontés lors de productions orales en anglais. Adams (1979) étudie auditivement et instrumentalement les productions en langue anglaise d'apprenants de diverses nationalités et montre la difficulté d'acquérir le rythme de la L2. Elle met en avant différents facteurs à l'origine de ces difficultés, tels que l'insuffisance de différence de durées entre les syllabes accentuées et inaccentuées, la non réduction des voyelles inaccentuées, le manque de liaisons dans l'enchaînement du discours, les pauses inappropriées, les déplacements d'accent de mots. L'auteur montre par le biais de ces facteurs que les productions du rythme des apprenants sont influencées par leur capacité à acquérir les schémas rythmiques de la L2.

Si l'on applique les facteurs d'Adams aux productions des apprenants, on constate rapidement qu'ils sont tout à fait vérifiés et que les problèmes de réalisations correctes du rythme de la langue cible, l'anglais, peuvent s'expliquer par plusieurs différences majeures de propriétés structurelles rythmiques entre les deux langues. De fait, Delattre (1966) propose une liste comparative du fonctionnement des deux langues. Selon l'auteur, les caractéristiques articulatoires du français influencent la production du rythme. Le mode tendu du français justifie de la stabilité du timbre des sons au cours de l'articulation et donne un rythme perçu avec peu d'inégalité entre les syllabes, tandis que le mode relâché explique la diphtongaison des voyelles et donc les glissements intra-vocaliques. Delattre note que l'accentuation de l'anglais est marquée par une augmentation de l'intensité syllabique, alors que celle du français est caractérisée par une augmentation de la durée syllabique. Enfin, le français a une structure syllabique simple (CV), l'anglais possède une syllabification fermée, une structure syllabique complexe (CVC).

Wenk (1985) rend compte des différences majeures au niveau de l'accentuation. En français, l'accent est placé en fin de groupe rythmique tandis qu'en anglais il est généralement en tête de groupe rythmique. La réalisation de l'accent est observée par une augmentation de l'intensité en anglais contrairement à une augmentation de durée en français. Au niveau articulatoire, le français garde une tension dans la production des

syllabes inaccentuées, *a contrario* de l'anglais dont la tension est relâchée, ce qui permet une centralisation vocalique. Cela entraîne l'absence de distinction voyelles courtes/voyelles longues en français, caractéristique vocalique majeure de l'anglais. Par ailleurs, il est à noter au niveau de l'accent lexical que le français aura tendance à accentuer la dernière syllabe du mot tandis que l'anglais est une langue à accent libre, l'accent peut être sur n'importe quelle syllabe du mot (ce qui lui attribue une fonction distinctive dans la mesure où la place de l'accent permettra de distinguer une paire de mots, exemple : '*record* (disque) /*re*'cord (enregistrer)).

Bolinger (1981) explique que la qualité vocalique est à l'origine d'une rythmicité différente entre les deux langues. L'anglais a une tendance à la réduction vocalique des syllabes inaccentuées, ce qui entraîne un certain type de rythme, tandis que le français présente une qualité constante des voyelles pleines, donnant l'impression d'un battement syllabique. Bolinger fait remarquer que c'est bien la non réduction vocalique du français qui entraîne un caractère isosyllabique de la langue, responsable de cet aspect rythmique. Il illustre ce propos par l'exemple de l'anglais hawaïen qui est caractérisé par l'absence de réduction vocalique, ce qui entraîne une rythmicité syllabique.

A partir de ces différences structurelles, la langue maternelle va alors créer des interférences rythmiques dans la production de la L2. Nicaise & Gray (1998) en listent quelques-unes, similaires aux critères exposés par Adams (précédemment cités) :

- trop de force dans les syllabes accentuées
- syllabes inaccentuées insuffisamment réduites
- déplacement d'accent sur la dernière syllabe du dernier mot du groupe rythmique
- déplacement d'accent lexical sur la dernière syllabe d'un mot pluri-syllabique
- absence de neutralisation vocalique des syllabes inaccentuées

Ces facteurs peuvent être illustrés par l'exemple suivant : / maRgA'REt ta'tS9R/

Toutes ces différences (répertoriées succinctement dans le tableau présenté ci-dessous) vont entraîner un certain nombre de distorsions rythmiques dans la production des apprenants.

ANGLAIS	FRANCAIS
➤ structure syllabique complexe (CVC)	➤ structure syllabique Simple (CV)
➤ distinction quantitative voyelles courtes vs. longues	➤ absence d'opposition quantitative des voyelles
➤ réduction vocalique	➤ absence de réduction vocalique
➤ langue à accent libre	➤ langue à accent fixe
➤ isoaccentuelle	➤ isosyllabique

**Tableau 1. Récapitulation des principales différences des propriétés structurelles rythmiques
-anglais vs français-**

Les facteurs évoqués de différenciation de durées entre syllabes accentuées/inaccentuées, entre voyelles pleines/réduites, entre structures syllabiques simples/complexes amènent à une différence traditionnelle entre les deux langues qui est celle de la classification des langues selon un caractère isosyllabique (français) ou isoaccentuel (anglais), que l'on peut noter à la lecture de ce tableau en dernière position. En effet, cette notion d'égalité/d'inégalité, de régularité/d'irrégularité du rythme ouvre le débat vers une représentation traditionnelle des deux langues dont les tendances rythmiques établies sont complètement opposées. Cette thématique, qui est le point central de départ de notre analyse expérimentale, fait l'objet d'une présentation diachronique dans la section suivante.

1.2 Le rythme de la parole

Il serait surprenant que cette notion de tendance générale à la régularité qui entoure notre univers ne se reflète pas dans le langage, car c'est à partir de cette régularité rythmique qu'il semble intéressant de confronter les deux langues choisies (anglais et français), selon

leurs caractéristiques rythmiques. Par ailleurs, Cruttenden (1986) suggère que le rythme est à la base de tout système prosodique apportant une contribution déterminante à la fonction d'identification des langues (Di Cristo & Hirst, 1993). C'est sur cette régularité du rythme que va alors se baser une catégorisation scalaire des langues, mettant en opposition l'anglais et le français. C'est au travers de cette typologie opposant les deux langues en question que peuvent être expliquées, en partie, les erreurs prosodiques produites par les apprenants francophones.

1.2.1 Isochronie

Une caractéristique de la structuration rythmique, se situant au cœur de la naissance d'une dichotomisation des langues, est le concept d'isochronie des unités rythmiques.

Si l'on s'en réfère à une définition traditionnelle « populaire » de l'isochronie (du grec *isochronos*), on trouvera plutôt les termes d'« isochronisme » ou encore d'« isochrone » ayant pour définition « qui s'effectue dans des intervalles de temps égaux » (Larousse, 2003). En application dans le domaine du langage, le principe d'isochronie peut se définir comme une segmentation de la chaîne parlée en un certain nombre d'unités de durée égale. Deux types d'isochronie se distinguent³ : l'isochronie syllabique (*syllable-timing*) -retour à intervalles réguliers des syllabes- et l'isochronie accentuelle (*stress-timing*) -retour à intervalles réguliers des accents. Cette dualité serait née de la traditionnelle comparaison faite entre l'anglais et le français, langues utilisées comme « représentant » de l'une ou l'autre tendance rythmique (Astésano, 2001).

Cette notion d'isochronie introduite par Steele (1779) est reprise par Classe (1939) qui en questionnait l'existence. Classe analyse les durées entre les syllabes accentuées de l'anglais au moyen d'un kymographe. Il remarque une isochronie dans la parole, toutefois selon certaines conditions, c'est-à-dire avec des structures phonémique et grammaticale similaires. Il conclut qu'une isochronie stricte serait rare en anglais dans la mesure où

³ On a également décrit un troisième type nommé « isomoraïque » (*mora-timed*) mais nous ne développons pas cette catégorie rythmique qui correspond aux langues à tons et n'entre donc pas dans notre étude.

plusieurs facteurs entrent en jeu et qu'elle serait le résultat d'un ensemble de « coïncidences » compliquées.

Le terme réapparaît dans la définition du rythme donnée par Pike (1945) – reprise par Abercrombie (1967) et Halliday (1967). L'auteur parle de récurrence isochrone d'unités ; on évoque alors l'isochronie accentuelle ou d'isochronie syllabique, ce qui permettrait d'attribuer aux langues un statut de langue « isoaccentuel » ou « isosyllabique ».

Ainsi, selon Pike (1945) et Abercrombie (1967), l'anglais possède des intervalles entre les pieds qui sont de longueur à peu près égale. Cette langue a donc une tendance à l'isochronie accentuelle, c'est-à-dire que les syllabes accentuées (indépendamment du nombre de syllabes formant l'unité rythmique) sont égales en durée. Cette isochronie est possible seulement si la durée de chaque syllabe n'est pas égale. Cette définition suppose que les syllabes se trouvant dans les intervalles sont soit étirées, soit compressées pour pouvoir obtenir au final une durée constante de l'intervalle accentuel.

Concernant le cas du français, ce dernier est décrit comme une langue à isochronie syllabique, c'est-à-dire que les syllabes se produisent à intervalles de temps réguliers.

Cependant, le principe d'isochronie appelle à de nombreuses études car si les travaux menés ne rejettent pas cette notion, ils tentent plutôt d'en apporter une justification par des preuves scientifiques, ce qui ne va pas forcément dans le sens d'une isochronie parfaite. Les études de Faure, Hirst & Chafcouloff (1980), Roach (1982), Dauer (1983) sur l'anglais montrent que c'est la perception des syllabes accentuées et inaccentuées qui mène à une impression de régularité des intervalles accentuels. Les travaux de Wenk & Wioland (1982) sur le français montrent que la durée moyenne des syllabes semble diminuer quand deux syllabes accentuées successives présentent un intervalle trop important. Cependant leurs travaux montrent aussi que les durées syllabiques ne sont pas constantes en français. Ces auteurs soulignent que l'allongement de la syllabe finale du groupe rythmique n'est pas pris en compte par l'isochronie syllabique. Toutes ces études concluent que l'isochronie n'est pas réelle.

...plus ou moins isochrone

Que ce soit pour l'anglais ou pour le français, le principe d'isochronie ne s'est pas révélé convaincant. S'il est difficile de rassembler des preuves objectives de la tendance à l'isochronie, cette notion n'est pas fondamentalement remise en cause. De plus, nous nous devons de souligner que, jusqu'ici, l'isochronie a été décrite comme un phénomène

« rigide » avec des unités égales en durée. Cependant, il est à noter que les auteurs (Pike, 1945 ; Halliday, 1967) soutenant le principe d'isochronie parlent de durée d'intervalles à peu près égale. En voici quelques exemples :

- Pike (1945 :33) « [...] somewhat uniform »
(1945:35) « [...] more or less evenly recurrent interval »
- Halliday (1967:23) « [...] roughly regular »

L'expérience de Halliday (1967) montre qu'il existe seulement une tendance vers une certaine régularité. L'auteur a souhaité mesurer le rapport de durées des pieds d'une, deux et trois syllabes. Les résultats du rapport ne présentent pas des chiffres égaux mais plutôt approximatifs (5 :6 :7), ce qui n'empêche pas une isochronie relative.

Kohler (1991) remarque, qu'en allemand, la compression des syllabes se fait seulement dans les intervalles d'une et deux syllabes. A partir de trois syllabes, il n'y a plus de compression. L'auteur note aussi une tendance des intervalles accentuels à être de durée régulière, car la durée de la syllabe inaccentuée semble s'ajuster selon la durée de la syllabe accentuée précédente.

Si une isochronie pure ne puise pas son existence sur des données objectives, si les recherches menées tout au long de ces années n'ont pas mené à des résultats concluants, nous pouvons supposer, au vu des travaux (Bolinger, 1965 ; O'Connor et Tooley (1961), l'existence d'une certaine régularité ; le principe de compression cité ci-dessus abonde dans le sens d'une recherche de régularité dans les énoncés. Mais cette absence de données objectives dans la quête de trace d'isochronie par des analyses acoustiques mène vers l'hypothèse selon laquelle la régularité rythmique n'a d'existence qu'à un niveau perceptif et non à un niveau physique. Des études (Lehiste, 1977 ; Couper-Kuhlen, 1993) ont montré que, de manière générale, l'individu perçoit la parole comme régulière, alors qu'acoustiquement elle ne l'est pas.

D'une isochronie pure à une isochronie subjective

Nombreux sont les chercheurs qui sont arrivés à la conclusion que l'isochronie ou la rythmicité devait être considérée comme un phénomène perceptif. Certains auteurs, tels que Meumann (1894), pensaient déjà que le rythme était le fruit d'une activité créatrice de la personne :

« Ce qui est essentiel dans le rythme, consiste en des actes intellectuels dont l'occasion est fournie dans certaines conditions par une succession rapide de sensations déterminées » (1894 : 272).

Des décennies plus tard, l'idée est toujours présente. Selon le psycholinguiste Fraisse (1956), la régularité rythmique ne serait réelle qu'au niveau perceptif. Pour l'auteur, c'est le sujet lui-même qui forme ou crée des groupements ; ainsi il y a perception du rythme lorsque le sujet forme des groupements de façon subjective ; ceux-ci seraient organisés en groupes de deux, trois ou quatre lorsqu'il y a une succession régulière de sons identiques.

« À vrai dire, il n'y a jamais groupement, ou forme, que pour un sujet qui perçoit ce groupement et, comme nous allons le voir plus loin, nous ne percevons de groupements que dans des limites psychophysiologiques inhérentes au mode de fonctionnement de nos sens et de nos centres nerveux. En ce sens toute rythmicité est subjective » (1956 : 9).

Le groupement est donc relatif à un sujet percevant. De plus, l'auteur ajoute que les structures rythmiques sont caractérisées par deux principes régis par des processus perceptifs: (i) le principe d'assimilation et (ii) le principe de distinction. Le sujet réorganise les structures en utilisant ces deux principes pour obtenir une rythmicité ; Fraisse (1956) mène une expérience par une technique de reproduction de frappe à quatre tapes (soit trois temps) selon la structure suivante : • • • •

Le sujet réajuste la rythmicité de l'énoncé soit en sous-estimant la valeur entre deux temps (assimilation), soit en surestimant le temps (distinction). Fraisse remarque dans le cas présenté ci-dessus que lorsque le premier temps prend des valeurs croissantes, il est légèrement sous-estimé, en même temps que les deux temps longs sont de plus en plus surestimés. Fraisse conclut :

« il y a en effet une interdépendance des temps telle que la modification de la durée d'un des temps entraîne une réorganisation de tous. [...] Les réorganisations se font suivant deux principes. L'un, de distinction, accuse les différences insuffisantes. L'autre, d'assimilation, c'est-à-dire de relative égalisation, tend à supprimer ou tout au moins à réduire les différences minimales » (1956 : 58).

Ces principes rappellent l'illusion de Delboeuf (Delorme & Fluckiger, 2003) (cf. annexes, section 2): celui-ci explique qu'un cercle placé dans un autre cercle concentrique plus grand est lui-même surestimé (principe d'assimilation) ; en revanche, si le cercle à

l'extérieur est bien plus grand, alors le cercle se trouvant à l'intérieur est sous-estimé (principe de distinction). Ce même type de phénomène est observé par Allen (1975) dans une expérience qui démontre que les sujets sous-estiment la durée des intervalles longs et surestiment la durée des intervalles courts, d'où une impression de régularité. Le paramètre de l'accentuation est aussi pris en compte dans l'étude de Fraisse sur la « rythmicité subjective ». Les caractéristiques de l'accentuation auraient des incidences sur le rythme ; Fraisse s'appuie sur des considérations perceptives. L'auteur cite l'expérience de Schmidt (1939) : les sujets doivent reproduire les groupes rythmiques (en les frappant) comme ils les perçoivent. Les résultats font ressortir deux dominantes de schéma d'accentuation selon lesquelles les sujets frappent plus fortement le premier ou le dernier élément du groupe rythmique. Deux décennies plus tard, Fraisse (1974), par une série d'expériences faites avec des sujets de groupes linguistiques divers, montre que la position perçue de l'élément accentué varie selon l'intensité et la hauteur de l'accent. Ainsi, si l'accent produit subit une augmentation d'intensité ou de hauteur, il sera perçu comme débutant le groupement ; par contre, si l'élément qui porte l'accent apparaît plus long que les autres, il sera perçu comme fin de groupe.

Fraisse n'est pas le seul à prôner une rythmicité apparaissant comme phénomène perceptif. Ce point de vue est aussi vivement exprimé par Benguerel & D'Arcy (1986). Selon ces auteurs, une quelconque isochronie ou rythmicité ne se produirait qu'à un niveau perceptif et non au niveau de la réalisation acoustique⁴ :

« it seems surprising that anyone interested in finding evidence for perceptual regularity would expect measurements (no matter how careful) of the acoustic signal to provide such evidence... In our view, if there exists any regularity or rhythmicity in speech, it is at the perceptual level, and possibly at the pre-production level, but not at the acoustic level » (1986: 244).

⁴ Cependant notons qu'une expérience menée par Cummins & Fort (1998) démontre qu'une rythmicité accentuelle est tout à fait observable à partir de mesures acoustiques : ces auteurs définissent le rythme comme une organisation hiérarchique des unités prosodiques temporelles.

Ces auteurs soutiennent qu'il y a un contraste entre la phase de pré-production et la phase de production : une déviation entre le patron régulier sous-jacent et la production irrégulière résulte en partie de l'influence de diverses caractéristiques intrinsèques des contraintes articulatoires, rythmiques, contextuelles, des phonèmes (etc.). Lors de la phase de production, une déformation se produit ; celle-ci engendre des énoncés qui ne sont pas isochrones.

Diverses expériences du type de celles de Fraisse sont menées afin de démontrer le phénomène perceptif de l'isochronie. Au cours d'une expérience, Lehiste (1977) propose à des sujets d'identifier (par stimuli) les intervalles les plus courts et les plus longs. Les résultats montrent qu'il faut diminuer de 30 à 100 ms un intervalle pour qu'il soit jugé plus court et vice versa. Les sujets semblent plus ou moins compenser les changements de durée. L'auteur conclut qu'ils perçoivent une certaine rythmicité dans la parole qui n'a pas forcément réalité d'être et met en avant l'incapacité des individus à identifier les différences existantes.

Le principe d'isochronie a donc ouvert la voie à de nombreux débats. Illusion perceptive ou réalité acoustique, la notion d'isochronie est directement associée à la classification des langues. En effet, il semblerait que le rythme de la parole soit un bon moyen pour caractériser et classer les langues dans des catégories. C'est à partir de cette hypothèse d'isochronie qu'il serait possible, en mesurant les durées entre soit les syllabes, soit les accents, soit les mores, de déterminer l'appartenance à une classe rythmique d'une langue. C'est ce que nous nous proposons de présenter dans la partie suivante au travers d'un bref historique.

1.2.2 Classification des langues : évolution du concept

Théorie des langues isoaccentuelles/isosyllabiques

La catégorisation des langues est née d'après une impression (auditive) de régularité dans la rythmicité des langues – une régularité qui se retrouve soit dans la syllabe, soit dans l'accent. Lloyd James (1940) compare le son produit par certaines langues (prenons ici l'anglais comme exemple) au rythme produit par un message codé en morse « morse code rhythm languages » et le son produit par les autres langues (comme le français) au rythme produit par le son d'une mitraillette « machine-gun rhythm ». C'est sur

cette distinction observée par Lloyd James que Pike (1945) a puisé son inspiration pour renommer ces deux différences rythmiques en deux catégories de langues montrant une certaine tendance rythmique. Ainsi, les premières hypothèses concernant le rythme des langues révèlent une dichotomie qui permet de classer les langues selon deux catégories :

(i) les langues dites à rythmicité syllabique (ou « syllable-timed ») : le rythme se base sur les syllabes, c'est-à-dire que ce sont les syllabes qui reviennent à intervalles réguliers.

Pike explique que les syllabes sont de durée quasi égale et que la durée des intervalles est proportionnelle au nombre de syllabes :

« French syllables tend to come at more or less evenly recurrent intervals – so that, as a result, phrases with extra syllables take proportionately more time » (1945: 35)

(ii) les langues dites à rythmicité accentuelle (ou « stress-timed ») : le rythme se base sur les accents, c'est-à-dire que ce sont les accents qui reviennent à intervalles réguliers.

Pike décrit ces langues comme ayant une durée des intervalles accentuels quasi égale mais ne dépendant pas du nombre de syllabes dans l'intervalle ; ce dernier est caractérisé par un certain degré de compression des syllabes.

Abercrombie (1967) ajoute que toutes les langues possèdent un rythme et ont obligatoirement leur place dans l'une ou l'autre des catégories. L'auteur réintroduit la notion d'isochronie (proposée auparavant par Pike (1945): selon lui, le rythme des langues est donc basé soit sur une isochronie syllabique soit sur une isochronie accentuelle, l'isochronie s'appuyant sur un principe de durée égale des unités rythmiques. C'est donc par ce principe d'isochronie que ces auteurs justifient une classification des langues. Cette isochronie s'accompagne d'un mouvement régulier de contractions et relâchements des muscles respiratoires, qui intervient soit sur la syllabe soit sur l'accent produit. Abercrombie fonde ses « hypothèses » sur des observations physiologiques. Selon lui, le rythme est une affaire de rythme musculaire concernant plus particulièrement les muscles respiratoires. L'auteur choisit deux groupes de trois langues et considère que pour les

langues à rythmicité syllabique telles que le français, le télougou⁵, et le yoruba⁶, chaque syllabe correspond à des pulsations pulmonaires et que pour les langues à rythmicité accentuelle, telles que l'anglais, le russe et l'arabe, les pulsations pulmonaires interviennent à chaque accent produit :

“...known as a syllable-timed rhythm, the periodic recurrence of movement is supplied by the syllable-producing process: the chest-pulses, and hence the syllables, recur at equal intervals of time-they are isochronous [...] known as a stress-timed rhythm, the periodic recurrence of movement is supplied by the stress-producing process: the stress-pulses, and hence the stressed syllables, are isochronous [...]” (1967: 97).

Nous nous retrouvons ainsi avec deux catégories strictes de langues que ces auteurs ont nommées : les langues isosyllabiques et les langues isoaccentuelles.

Dans cette classification, seulement deux catégories rythmiques nous sont proposées. Or nous nous devons de souligner qu'il existe une troisième catégorie de langues, dont fait partie notamment le japonais, que l'on appelle les langues à rythmicité moraique. Bloch (1950) introduit cette troisième catégorie de langues et explique que ce type de rythme repose sur la more⁷. Dans ce type de catégorie, c'est au niveau des mores que nous retrouvons une certaine isochronie.

Par ailleurs, il est à noter que par cette catégorisation, l'anglais et le français sont formellement opposés, et semblent, selon nos auteurs, avoir des caractéristiques rythmiques très différentes. Ces deux langues apparaissent comme les prototypes de l'une et l'autre classe rythmique, très souvent pris comme exemple pour illustrer l'une ou l'autre des catégories rythmiques. Très rapidement, ce principe de catégorisation des langues suscite l'intérêt de nombreux linguistes et va entraîner de nombreux travaux à ce sujet.

⁵ Langue dravidienne (famille de langues du sud de l'Inde) parlée dans l'état d'Andhra Pradesh (Inde).

⁶ Langue originaire du sud ouest du Nigeria (Bénin) et d'une petite partie du Togo, parlée aussi au Brésil, Cuba, Amérique Latine, Caraïbes.

⁷ sous unité des syllabes suivant le schéma CV (ma) ou (C)VV (aa) ou CVC (mam)

Variantes de la classification.

Wenk & Wioland (1982) adoptent une position plus modérée en utilisant les termes de « trailer-timed » (ou encore « regulated group-finally ») pour parler des langues dites isosyllabiques et de « leader-timed » (ou « regulated group-initially ») pour les langues dites isoaccentuelles. Cependant, le même type d'opposition est présenté : d'un côté les langues « codachrones » (trailer-timed languages) - comme le français - dont la syllabe accentuée se situe en finale de groupe rythmique, et de l'autre les langues « capochrones » (leader-timed languages) – comme l'anglais - dont la syllabe accentuée se situe en initiale de groupe rythmique. Le même principe se retrouve dans les travaux de Di Cristo & Hirst (1998) sauf que ces auteurs ne classent pas les langues dans une certaine catégorie mais constatent que les groupes rythmiques du français se caractérisent par une structure de « tête à droite » et les groupes rythmiques de l'anglais reposent sur une structure de « tête à gauche ».

Liberman & Prince (1977) se basent sur les différents niveaux de l'organisation rythmique des langues pour mettre en évidence la catégorisation des langues. Pour cela, les auteurs créent une grille métrique utilisant un système d'astérisques placés au-dessus des énoncés représentant différents niveaux d'accentuation. Trois niveaux sont proposés : les syllabes, les éléments proéminents, l'accent principal (du mot).

Beckman (1992) propose de justifier cette dichotomie des langues par le biais de la « périodicité ». Elle oppose « les langues accentuelles » comme l'anglais « aux langues de frontière » comme le français. Selon l'auteur, l'anglais possède une périodicité marquée par l'accentuation, contrairement au français qui présente une périodicité ponctuée par des coupures temporelles.

Toutes ces études proposent donc des catégories bien distinctes de langues, en fonction du rythme. Trois catégories classiques sont proposées : les langues à rythme accentuel, syllabique ou moraique ; une classification classique que nous pouvons illustrer par le tableau suivant (représentation et classement des langues selon nos diverses lectures) :

langues accentuelles	langues syllabiques	langues moraïques
anglais allemand arabe néerlandais russe polonais thaï	français espagnol italien grec catalan telougou/ yoruba hindi	japonais tamoul

Tableau 2. Classement des différentes langues selon leur classe rythmique

Cependant, cette catégorisation classique des langues en fonction du rythme ne reçoit pas l'accord de tous les linguistes et se voit fortement remise en question.

...vers un continuum rythmique

Dauer (1983) tente de justifier les conclusions d'Abercrombie en y apportant des mesures acoustiques. Elle procède à une analyse en mesurant les variabilités des durées des intervalles entre accents, en choisissant l' « onset » (attaque) comme début d'intervalle. Dauer obtient des résultats insatisfaisants pour valider l'hypothèse d'Abercrombie. Les durées des intervalles ne sont pas égales. Dauer explique que les différences entre les langues résident plutôt dans la structure interne de l'intervalle, reposant sur des propriétés plus phonologiques, notamment :

- la complexité des compositions syllabiques
- l'absence ou non de réduction vocalique
- l'importance de l'accent

Ainsi, selon les caractéristiques proposées, nous pouvons décrire les deux types de langue de la façon suivante :

(i) Les langues dites syllabiques (comme le français) se composent de syllabes simples (la plus courante étant CV), et n'admettent pas de réduction vocalique. Les syllabes sont donc de taille relativement stable et semblent donner un rythme régulier.

(ii) Dans les langues accentuelles (comme l'anglais), la composition des syllabes est beaucoup plus complexe et variée et porte généralement sur l'accent tonique ; les syllabes plus simples font l'objet d'une réduction vocalique, accentuant la distinction entre syllabes

accentuées et inaccentuées (fortes/faibles). L'effet de l'accent a tendance à être plus marqué dans ce type de langue car il s'accompagne d'une combinaison d'allongement, de modification du contour mélodique, de changement de hauteur et de qualité ; il est donc plus perceptible en anglais.

Dauer (1983) établit alors un continuum rythmique sur lequel elle place les langues selon une certaine tendance plus ou moins accentuelle. Elle utilise un axe qui représente le degré d'importance de l'accent.

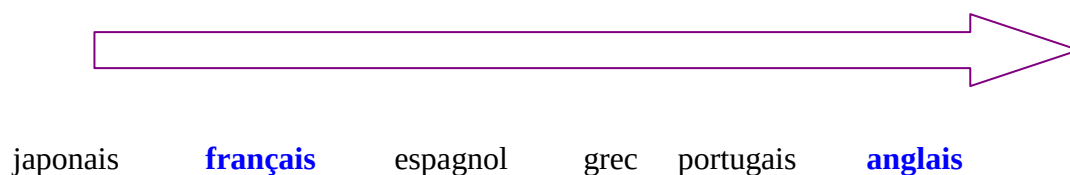


Figure 1. Axe représentant le degré d'importance de l'accent dans différentes langues (Dauer, 1983:60)

L'axe représenté par une flèche permet d'observer un continuum allant des langues moins accentuelles (partant du japonais) aux langues plus accentuelles (l'anglais) ; on retrouve à l'extrême gauche les langues à rythmicité moraique (japonais), à l'extrême droite les langues à rythmicité accentuelle (l'anglais) et sur l'axe les langues à rythmicité syllabique comme le français.

Les travaux de Dauer vont inspirer de nombreuses autres recherches ouvrant la voie à une seconde école qui préfère opter pour l'idée d'un continuum rythmique, remarque qui avait été faite quelques années auparavant par Delattre (1966), préconisant une répartition des langues sur un continuum de valeurs contrastées et non dans deux catégories absolues. Ainsi d'autres critères vont s'ajouter au fur et à mesure, suivant l'idée de ce continuum: Hoequist (1983, cité par Herry, 2001) ajoute la caractéristique d'effet d'adjacence: les syllabes accentuées subissent l'influence des syllabes inaccentuées adjacentes ; le continuum va donc de l'absence d'effet d'adjacence à fort effet d'adjacence à l'autre extrême. Miller *et al.* (1984) argumente dans le sens de critères phonologiques (tels ceux énoncés par Dauer) ; Eriksson (1991) souhaite s'intéresser à la composition des intervalles plutôt qu'à leur durée. Dauer (1987) établit, quelques années après l'élaboration de l'axe scalaire, une classification selon certains traits phonologiques ; l'auteur propose un éventail de traits. Elle établit un système de « cotation » selon lequel chaque trait se voit

assigner de la valeur « + » ou d'un « - ». De cette façon, plus une langue obtient de « + » plus il est probable, selon Dauer, que cette langue en question ait de « forts accents » et donc ait tendance à être accentuelle. Bertinetto (1989) propose une orientation scalaire en remplacement de la dichotomie traditionnelle. Bertinetto explique qu'il n'existerait qu'un type de langues unique mais qui varie en fonction de déviations scalaires (par rapport à une isochronie pure). Selon lui, toutes les langues ont une base rythmique commune. L'auteur présente une synthèse des facteurs phonétiques et phonologiques (dont certains se retrouvent chez Dauer) 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.

L'auteur conclut qu'une langue possède un certain degré des différents traits proposés ci-dessus, ce qui contribue à l'orienter vers un pôle rythmique plutôt qu'un autre :

« iso-accentual and iso-syllabic languages do not contrast because of some radically different feature they are supposed to possess (or, alternatively, not to possess), but rather because of the varying degree to which they seem to exhibit the various features, phonetical and phonological, which are supposed to contribute to orient a language more towards one or towards the other pole of rhythmical behaviour » (1987: 111).

Les langues sont alors caractérisées par différents facteurs ; elles possèdent chacun de ces facteurs selon un certain degré d'importance. Les diverses langues sont localisées selon une scalarisation, et non plus classées selon une dichotomie stricte. Elles sont plus ou moins isoaccentuelles ou plus ou moins isosyllabiques.

Il semble que ce continuum ait inspiré de nouvelles études. Tout d'abord, certains linguistes ont affirmé qu'une langue pouvait présenter les deux tendances rythmiques à la fois. Parallèlement à cela, il semble que ce continuum conduit à de nouveaux travaux. Des chercheurs s'interrogent sur l'hypothèse selon laquelle il puisse exister des variabilités à l'intérieur même d'une langue, celles-ci classant les dialectes sur des points différents du continuum.

Coexistence rythmique : études inter-dialectales

Ce principe d'obtenir un continuum de langues plus ou moins syllabiques ou accentuelles va développer de nouvelles études afin de comparer les variabilités à l'intérieur d'une même langue.

La majorité des études inter-dialectales porte sur la variabilité de l'anglais. Bolinger (1981) montre par une expérience, que l'anglais d'Hawaï aurait une tendance à la rythmicité syllabique. Deterding (2001) s'intéresse aux variabilités dialectales entre l'anglais britannique standard et l'anglais de Singapour. Il choisit de mesurer la durée des syllabes. Ces travaux montrent que les durées syllabiques varient considérablement entre les deux variétés : l'anglais de Singapour semble avoir des durées syllabiques plutôt égales, contrairement à celles de l'anglais britannique standard. De plus, les voyelles réduites semblent plus présentes en anglais britannique standard. L'auteur conclut donc que l'anglais de Singapour a une tendance à l'isosyllabité, tandis que l'anglais britannique standard serait plutôt isoaccentuel. Grabe & Low (2002) étudient la même comparaison et abondent dans le même sens que Deterding. Elles observent aussi que les voyelles réduites de l'anglais de Singapour sont aussi moins centralisées et plus longues que celles de l'anglais britannique standard.

D'autres langues ont été étudiées en ce sens : Fernandez-Cruz (2000) montre des différences rythmiques pour les dialectes portugais. Elle analyse un rythme accentuel pour le portugais européen, syllabique pour le portugais brésilien et moraique pour le portugais afro-brésilien. L'étude de Barkat-Defradas, Hamdi, Pellegrino (2004) montre différentes rythmicités de l'arabe allant d'est en ouest.

De plus en plus de travaux se sont développés, autant concernant les différences à l'intérieur des langues se penchant sur la comparaison de caractéristiques des variétés dialectales, que, de manière générale, des études entre les langues du monde. Les dernières études (Roach, 1982 ; Dauer, 1983) rejettent le principe d'isochronie reposant sur une répétition d'unités semblables (pied, syllabe ou more) classant les langues dans trois catégories distinctes. Elles penchent plutôt pour un continuum. Ces derniers travaux ne reposent plus sur une isochronie accentuelle ou syllabique (ou moraique) mais sur des caractéristiques plus phonologiques, comme la complexité des syllabes, la corrélation entre poids syllabique et accent, l'absence ou non de réduction vocalique. Récemment de nouvelles approches, inspirées de ces derniers travaux, se sont développées.

1.2.3 Propositions de métriques rythmiques

Les travaux présentés ci-dessous ont pour but de tester expérimentalement, par le développement de métriques reposant sur des données acoustiques, la classification des langues proposées auparavant.

Ramus & al. (1999)

Ramus, Nespor & Mehler (1999) et Ramus (1999) proposent une nouvelle approche phonologique de la caractérisation des langues. Dans ces travaux sur la discrimination des langues, Ramus (1999) fonde son approche sur des bases phonologiques du rythme qu'il décrit ainsi :

« Notre approche repose sur une conception du rythme de parole, non plus comme primitive de l'organisation temporelle des langues, mais comme conséquence de certaines de leurs propriétés phonologiques [...] ». (1999:2).

Alors que les travaux précédents n'ont pas abouti à démontrer le classement rythmique des langues, Ramus choisit de résoudre ce problème en calculant l'intervalle entre consonnes et voyelles. Ramus se base sur deux principales caractéristiques (que l'on retrouve chez Dauer (1983) : la complexité de la structure syllabique et la réduction des voyelles inaccentuées). Selon l'auteur, l'hypothèse est qu'une analyse de la complexité syllabique d'une langue devrait permettre de déterminer sa classe rythmique. L'argument consiste à dire que plus une langue a de syllabes complexes, plus elle serait de type accentuel, et inversement, plus une langue a de syllabes simples, plus elle serait de type syllabique (ou moraique). Ramus se propose de quantifier ces propriétés phonologiques en mesurant la complexité syllabique par une segmentation en consonnes/voyelles, sur un corpus qu'il a constitué à partir d'un corpus multilingue du LSCP. Huit langues ont été choisies : l'anglais, le néerlandais, le polonais, le français, l'espagnol, l'italien, le catalan, le japonais. Quatre locutrices natives (de chaque langue) prononçaient 5 phrases, ce qui donne une somme de 20 phrases par langue, et donc un corpus de 160 phrases. Ce corpus

est segmenté manuellement en intervalles consonantiques⁸ et vocaliques⁹. Ramus introduit ainsi deux nouveaux corrélats ; ici, on ne mesure plus des durées d'unités rythmiques, ni de durées de phonèmes, mais on lance un calcul statistique de durées d'intervalles consonantiques et vocaliques suivant ces deux corrélats:

- la proportion de durées des intervalles vocaliques: %V (l'auteur note que la proportion d'intervalles consonantiques étant égale à $1 - \%V$, elle n'apporte aucune information supplémentaire).
- l'écart type des durées des intervalles consonantiques : ΔC

Ces deux corrélats acoustiques font partie de trois variables choisies par Ramus (la troisième variable étant l'écart type des durées d'intervalles vocaliques (ΔV)): l'auteur part du principe que le rythme est une propriété linguistique détectable dans le signal de la parole et a donc transposé les trois classes rythmiques (accentuelle, syllabique, moraïque) en trois variables. Les deux paramètres acoustiques sont liés à la structure syllabique des langues selon leur catégorie : en effet, les syllabes complexes comportent plus de consonnes que les syllabes simples. Le raisonnement est donc le suivant: plus une langue a de syllabes complexes, plus la proportion moyenne de consonnes augmente et plus le %V diminue. Ces deux paramètres acoustiques devraient donc pouvoir permettre de distinguer les variations entre les langues par une analyse de variance. L'auteur explique que ce sont les deux corrélats retenus qui lui ont permis de démontrer que les langues choisies se regroupent dans les classes traditionnelles rythmiques, suivant les trois catégories classiques : l'anglais, le néerlandais, le polonais appartiennent aux langues isoaccentuelles, et le français, l'espagnol, l'italien et le catalan sont isosyllabiques ; le japonais, lui, appartient à la catégorie isomoraïque. Ces résultats vont dans le sens de ceux établis par Dauer (1983). Cependant, l'auteur tient à préciser que ces indices ne présupposent pas forcément que les langues se divisent en deux groupes distincts. Les résultats obtenus

⁸ Intervalle consonantique = suite de consonnes, c'est-à-dire, toute séquence ininterrompue de consonnes (de la 1^{ère} consonne à la dernière de cette séquence)

⁹ Intervalle vocalique = suite de voyelles, c'est-à-dire, toute séquence de voyelles (de la 1^{ère} voyelle de la séquence à la dernière).

montrent donc que les langues à structure syllabique plus complexe ont un %V plus faible et un ΔC plus élevé que les langues à structure syllabique simple. L'auteur conclut que les variables %V et ΔC sont significatives et représentent le meilleur corrélat acoustique, puisqu'elles semblent rendre compte de la notion traditionnelle de classe rythmique. Notre intérêt portant plus particulièrement sur l'anglais et le français, nous notons les résultats obtenus : l'anglais qui a des voyelles pleines et réduites a un %V inférieur à celui du français (40.1% pour l'anglais contre 43.6% pour le français), langue qui n'a pas de réduction vocalique. En revanche, ΔC est plus important en anglais et reflète une structure syllabique complexe (46.4 en anglais contre 37.8 en français).

Low & al. (2000)

Une seconde étude présentant des paramètres différents est proposée par Grabe & Low (2002). Ces travaux s'appuient sur l'hypothèse faite par Low, Grabe & Nolan (2000), qu'une combinaison de l'équation $nPVI_V$ (décrite ci-dessous) avec des mesures de variabilité d'intervalles intervocaliques rendrait un meilleur indicateur de classe rythmique que le $nPVI$ seul. Le but de l'étude de Grabe & Low (2002) est d'apporter des preuves objectives et acoustiques à la classification des langues à partir de mesures de durées, tout en vérifiant l'hypothèse émise par Low, Grabe & Nolan (2000). Ces travaux diffèrent du précédent (Ramus, 1999) par deux nouveautés. Tout d'abord, les auteurs utilisent une approche phonétique puisque les unités phonologiques ne sont pas mises en relation avec le rythme. Cette étude ne s'appuie pas sur des mesures de durées des intervalles entre accents ou des syllabes mais propose de calculer le degré de variabilité dans les mesures successives des intervalles acoustico-phonétiques. Les calculs s'effectuent sur la durée des voyelles et des intervalles entre les voyelles. La deuxième particularité de cette étude est que les auteurs ont choisi d'étendre leur corpus sur 18 langues. Grabe & Low ont repris 11 langues déjà étudiées et ont introduit 7 nouvelles langues (un des objectifs de ces travaux étant d'observer des langues non classées) que l'on peut visualiser dans le tableau récapitulatif proposé ci-dessous:

Language	Classification
British English	Stress-timed (Classe, 1939, Pike, 1946, Abercrombie 1967)
German	Stress-timed (Kohler, 1982)
Dutch	Stress-timed (Ladefoged, 1975, Smith, 1976)
Thai	Stress-timed (Luangthongkum, 1977)
Tamil	Syllable-timed (Corder, 1973, Asher, 1985)
Spanish	Syllable-timed (Pike, 1946, Hockett, 1958)
French	Syllable-timed (Abercrombie, 1967, Catford, 1977)
Singapore English	Syllable-timed (Tongue, 1979, Platt and Weber, 1980)
Japanese	Mora-timed (Blos, 1942, Han, 1962)
Polish	Mixed (Dauer, 1987, Nespors, 1990)
Catalan	Mixed (Dauer, 1983, Nespors, 1990)
Estonian	Unclassified
Greek	Unclassified
Luxembourg	Unclassified
Malay	Unclassified
Mandarin	Unclassified
Rumanian	Unclassified
Welsh	Unclassified

Tableau 3. Langues étudiées dans les travaux de Grabe & Low (tiré de Grabe & Low (2002: 13): « Traditional Rhythmic classifications of languages investigated in the present study ».

Grabe & Low ont créé leur corpus en utilisant des passages comparables de parole lue. Chaque langue est représentée par un seul locuteur qui avait pour tâche de lire « The North Wind and The Sun » à leur rythme. Les enregistrements se sont déroulés en chambre sourde. Le corpus a été segmenté manuellement. Les auteurs précisent qu’elles ont mesuré la durée des voyelles qui étaient repérables sur le signal acoustique, c’est-à-dire les voyelles voisées, les voyelles dévoisées ne possédant pas les caractéristiques formantiques d’une voyelle voisée, car elles ne sont pas observables sur le signal. Les intervalles vocaliques sont mesurés de l’attaque (« onset ») de la première voyelle de l’intervalle jusqu’à la fin (« offset ») de la dernière voyelle ; pour les intervalles consonantiques, la mesure est prise depuis la fin de la dernière voyelle de l’intervalle (précédent) jusqu’à

l'attaque de la prochaine voyelle, sans porter attention au nombre de consonnes dans l'intervalle. Enfin, les analyses acoustiques ne tiennent pas compte des hésitations, ni des pauses. Afin de mener à bien leur étude, des mesures de durées vocaliques et intervocaliques sont calculées puis analysées par une équation appelée « Pairwise Variability Index » (PVI). Cette équation va permettre d'observer les variations, en calculant le degré de variation entre mesures successives. Les auteurs proposent en fait deux équations utilisables selon le « type » d'intervalles (consonantiques ou vocaliques).

La première équation est appelée « raw Pairwise Variability Index » (rPVI) :

$$PVI = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m-1) \right]$$

Dans cette équation, m représente le nombre d'intervalles (vocaliques ou intervocaliques), d représente la durée de l'intervalle. L'équation n'a pas été normalisée afin de ne pas supprimer les différences de structures syllabiques propres à chaque langue. Cette métrique permet de calculer les intervalles consonantiques et vocaliques ; pour l'étude de Grabe & Low (2002), ce calcul est appliqué uniquement aux intervalles consonantiques.

La seconde équation, qui est normalisée, permet de calculer également les deux types d'intervalles vocaliques et consonantiques. Grabe & Low (2002) l'appliquent uniquement aux intervalles vocaliques:

$$PVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1}) / 2} \right| / (m-1) \right]$$

Dans cette équation m représente le nombre d'items dans la phrase et d la durée de chaque item. L'équation est normalisée afin de neutraliser les effets du débit de parole. On calcule la différence de durée entre chaque couple successif de mesures, on prend la valeur

absolue de la différence que l'on divise par la durée moyenne du couple. Les différences sont ajoutées, puis divisées par le nombre de différences. Le résultat obtenu est multiplié par 100 car la normalisation produit des valeurs fractionnelles.

A partir de ces calculs, Grabe & Low émettent deux principales hypothèses :

1- les langues dites isoaccentuelles (anglais) auront des taux élevés de variabilités vocaliques (nPVI_V) et intervocaliques (rPVI_C). Inversement, les langues dites isosyllabiques (français) auront des taux de nPVI_V et rPVI_C bas.

2- les deux langues dites « mixtes » : le catalan aura un taux élevé de nPVI_V avec un taux bas de rPVI_C, car il possède une structure syllabique simple et aussi des réductions vocaliques ; le polonais (langue à structure syllabique complexe et pas de réduction de voyelles) aura un taux plus bas de nPVI_V que les langues isoaccentuelles mais un taux relativement haut de rPVI_C.

Aucune hypothèse n'a été formulée pour les nouvelles langues entrantes.

Les résultats obtenus après calculs vont dans le sens des prédictions des auteurs. Les valeurs obtenues s'accordent avec la classification de l'anglais, néerlandais et allemand comme langues isoaccentuelles et le français, l'espagnol comme langues isosyllabiques : les variabilités de durée sont beaucoup plus importantes pour les langues isoaccentuelles que les autres. Mais les auteurs ne prônent tout de même pas une catégorisation de ces deux classes. Les valeurs obtenues pour le japonais sont similaires à celle des langues isosyllabiques. Les résultats du catalan et du polonais soutiennent l'hypothèse énoncée. Notons cependant que le catalan a un taux de nPVI_V équivalent à celui du français qui n'a pas de réduction vocalique. Le thaïlandais est classé isoaccentuel comme prédit ; et enfin le tamil ne serait pas isosyllabique car cette langue obtient des taux de nPVI_V et rPVI_C trop élevés. En ce qui concerne les nouvelles langues ajoutées, elles ne conviennent à aucune des trois classes ; leurs valeurs se chevauchent.

L'étude ne s'arrête pas là. Conscientes du manque irrévocable de sujets dans la composition de leur corpus, Grabe & Low testent la stabilité des valeurs obtenues par une analyse statistique. Elles concluent à des résultats satisfaisants et à l'efficacité de l'équation PVI. Les auteurs signalent que si la plupart des résultats favorisent l'aspect de catégorisation des langues, en aucun cas elles ne soutiennent l'idée de dichotomie:

« *This finding supports the rhythmic classification suggested by Pike (1946) and Abercrombie (1967), even if the evidence does not come from isochronous interstress-intervals or syllable-durations. There is no support, however, for a strict categorical distinction between languages with high vocalic and intervocalic PVI values and languages with low vocalic and intervocalic PVI values. Rather, it appears that languages can be more or less ‘stress-timed’ or syllable-timed’* » (2002: 22).

Grabe & Low préfèrent donc parler de langues plus ou moins isoaccentuelles (ou syllabiques). Elles concluent à un chevauchement des langues:

« [...] *the results show that there is overlap between the stress-timed and the syllable-timed group and unclassified languages [...]* » (2002: 31).

A l’aide du graphique tiré de l’étude de Grabe & Low (2002) comparons de plus près les résultats obtenus pour nos deux langues élues prototypes parmi les 18 langues et qui nous intéressent pour notre propre étude: l’anglais et le français.

légende : ○ = stress-timed ● =syllable-timed ■ =mora-timed □ =unclassified

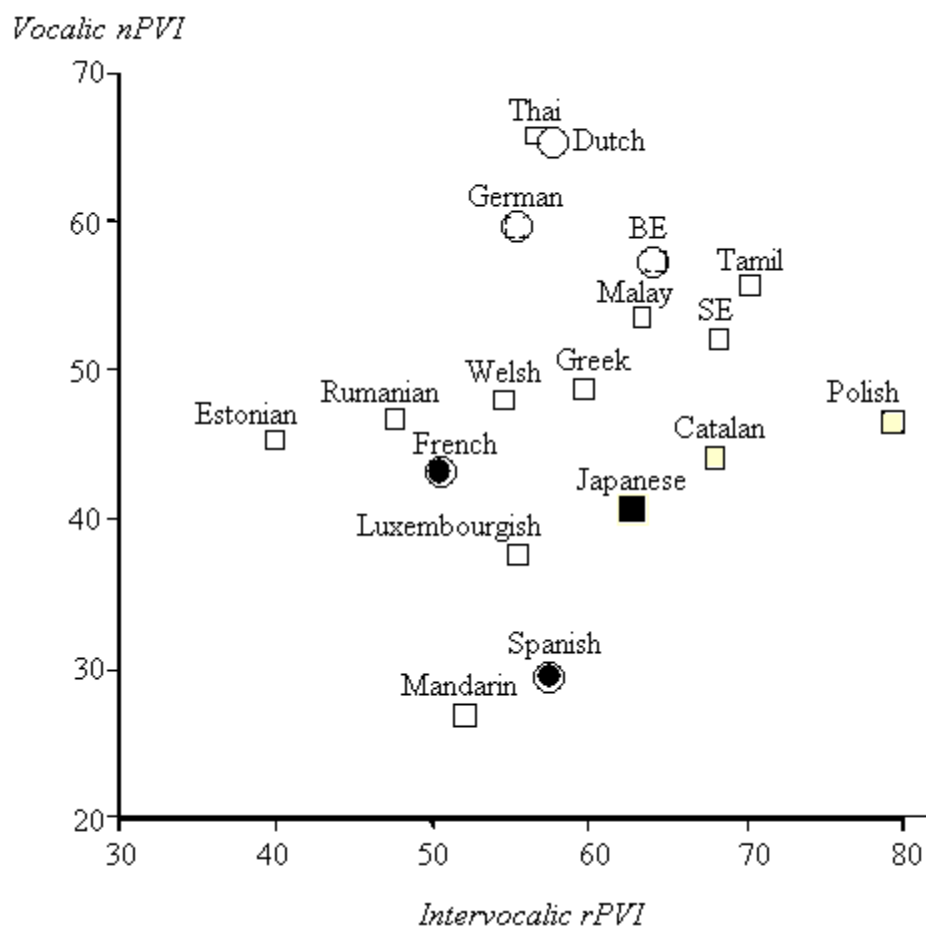


Figure 2. Valeurs du PVI pour les 18 langues (tiré de Grabe & Low, 2002:22)

Il est à noter que les valeurs obtenues par le PVI fournissent des preuves acoustiques évidentes pour les différences rythmiques que l'on retrouve de nouveau entre nos deux langues. D'après les résultats de Grabe & Low (2002) nous relevons que l'anglais, langue dite isoaccentuelle, présente un taux de variabilité vocalique nPVI important qui s'élève à 57.2, ce qui la fait figurer en haut du graphique. Quant au français, langue dite isosyllabique, on lui note un taux de variabilité vocalique (nPVI) beaucoup plus bas (43.5).

Même si l'anglais et le français n'appartiennent pas à une dichotomie stricte des langues, que ce soit sous forme de catégorie, de continuum, de traits, de variabilité, ces deux langues sont dans tous ces cas « opposées » ou, de toute façon, ne se présentent pas sur la même lignée. Que ce soit au niveau de l'organisation temporelle ou des études sur la classification des langues, l'anglais et le français présentent des diversités rythmiques évidentes qui ne peuvent les unir dans une même catégorie.

Remarques comparatives et limites des deux études proposées

Les deux dernières études proposent une approche différente de celles proposées auparavant. Il apparaît à la vue des paragraphes précédents qu'une dichotomie des langues se voit « voler en éclats ». Par contre, une diversité rythmique des langues est prouvée par des données acoustiques objectives.

Cependant, les résultats des deux études présentées ci-dessus ne se rejoignent qu'au niveau des langues traditionnelles (c'est-à-dire situées aux extrêmes de la figure) ; pour ce qui est des langues situées au centre du graphique, les résultats sont différents selon les calculs : avec les valeurs de %V (Ramus), la proportion des durées de voyelles est plus basse pour le grec, le catalan, le gallois, le luxembourgeois, et le japonais que pour l'allemand. Donc, selon l'hypothèse de Ramus, ces langues devraient être plus isoaccentuelles que l'allemand. Par contre, le taux de nPVI_V de Grabe & Low montre le contraire, puisqu'il est plus bas pour ces mêmes langues que l'allemand ; suivant l'hypothèse de ces auteurs, ces langues devraient donc être moins isoaccentuelles que l'allemand. Malgré ces différences de résultats, ces travaux représentent une grande avancée dans la mesure où ils fournissent des preuves acoustiques d'une variation rythmique des langues.

Néanmoins, divers problèmes font l'objet de débats dans la littérature, avec deux grands bémols concernant ces études : le nombre limité de sujets dans les corpus utilisés et la méthode de calcul utilisée qui est directement associée au tempo.

Wagner & Dellwo (2004) expliquent que la combinaison %V- ΔC (Ramus, 1999) rend plutôt compte de la structure syllabique que le rythme lui-même et que ces mesures ne sont pas résistantes au tempo. Ils préconisent trois points essentiels afin de mesurer le rythme :

- (i) les mesures doivent prendre en compte plus de paramètres que la structure syllabique,
- (ii) la métrique doit être résistante aux changements de tempo dans la mesure où ces fluctuations ne sont pas en lien avec la structure rythmique,
- (iii) les paramètres choisis doivent capturer la nature séquentielle des schémas rythmiques.

Barry & al. (2003) commentent et étudient le même problème exposé précédemment, à savoir le lien entre les mesures rythmiques et le tempo de la parole. Ils montrent que ΔC et ΔV sont directement associés au paramètre du tempo. Par ailleurs, une seconde analyse menée par Russo & Barry (2008) vient confirmer l'étude de 2003 et montre que le %V est un paramètre résistant au changement de tempo.

Hirst (2009) montre que les paramètres acoustiques proposés par Ramus ne pourraient être réellement considérés comme acoustiques. Les mesures proposées par Ramus reposent sur une segmentation linguistique préalable du signal, en phonèmes, et sur une classification des phonèmes en consonnes (C) ou en voyelles (V). L'auteur explique qu'aucune de ces opérations n'est acoustique et que le même résultat peut être obtenu en mesurant le pourcentage du nombre de phonèmes vocaliques, par rapport au nombre total de phonèmes et l'écart type du nombre de phonèmes consonantiques dans les différents intervalles consonantiques.

Métriques rythmiques et tempo: nouvelles propositions

Des recherches récentes tentent d'améliorer l'efficacité des métriques (Hirst, 2000 ; Dellwo 2006, White & Mattys, 2007) et apportent de nouvelles propositions s'appuyant sur une méthode de calcul différente.

Hirst (2000) propose d'utiliser un nouveau paramètre s'appuyant sur le coefficient de variation (noté désormais 'cv'), une mesure de dispersion relative, calculée selon la formule:

$$c_v = \frac{\sigma}{\mu}$$

L'auteur explique que le cv est indépendant de l'unité de mesure, ainsi que du débit de la parole. Selon Hirst (2000), les paramètres de Ramus sont intéressants mais d'autres paramètres (tels que le cv) pourraient améliorer la discrimination. Dès lors, l'application de ce nouveau paramètre est testée par Fernandez-Cruz (2000) qui utilise le coefficient de variation pour les intervalles consonantiques afin de distinguer le portugais européen (PE), le portugais brésilien (PB) et le portugais afro-brésilien¹⁰(PAB). Les paramètres de Ramus ($\Delta C_ \% V$) permettent de distinguer le PAB du PB ; en revanche, la combinaison cvC-%V proposée par Hirst donne un meilleur classement, et ce au niveau des trois dialectes. La combinaison cvC-%V est la représentation paramétrique rythmique la plus robuste dans cette étude.

Herry (2001) et Herry & Hirst (2002) appliquent cette même méthode pour distinguer les productions de francophones de celles des natifs anglais. Parmi une quinzaine de paramètres, une analyse statistique en arbre de régression, effectuée avec le logiciel CRUISE, révèle que le coefficient de variation des intervalles consonantiques, corrélé avec d'autres paramètres intonatifs, est le paramètre le plus discriminant.

Dellwo (2006) réitère la proposition avancée par Hirst et préconise également d'utiliser le coefficient de variation et présente « VarcoC », qui se base sur le calcul du coefficient de variation des consonnes, c'est-à-dire l'écart type des durées des intervalles consonantiques divisé par la moyenne des durées des consonnes, multiplié par 100. Les variations de tempo sont normalisées. Dellwo (2006) teste VarcoC et ΔC sur des productions de différents tempos afin de discriminer l'anglais et l'allemand du français. Il observe que VarcoC discrimine clairement les langues en question et ce pour les différents types de tempo analysés. VarcoC varie en fonction des tempos analysés mais la corrélation n'est pas systématique selon la langue. Cette nouvelle mesure d'intervalles normalisée est reprise par White & Mattys (2007) et appliquée également aux voyelles. Les auteurs testent

¹⁰ Selon l'étude de Fernandez-Cruz (2000), le portugais afro-brésilien aurait les caractéristiques du portugais avec des bases prosodiques du Bantoue.

les différentes métriques proposées, les indices de variabilité (rPVI, nPVI), les intervalles de mesures (%V, ΔC , ΔV) et les intervalles de mesures normalisées (VarcoV, VarcoC) afin de distinguer quatre langues : l'anglais et le néerlandais (isoaccentuelles), le français et l'espagnol (isosyllabiques). Les résultats montrent que la combinaison VarcoV, nPVI_V et %V donne la meilleure discrimination.

Ces résultats donnent des perspectives positives d'application des métriques permettant de distinguer plusieurs langues.

Application des métriques: discrimination L1/L2

La première utilisation de ces métriques est destinée à distinguer des structures rythmiques différentes entre les langues. De nombreuses études¹¹ ont démontré des résultats intéressants dans la discrimination des langues. Low & al. (2000) examinent les différences de classification entre l'anglais britannique (isoaccentuel) et l'anglais de Singapour (isosyllabique) en utilisant l'index de variabilité. Ils montrent que les durées des voyelles sont à peu près égales, contrairement à celles de l'anglais britannique, et que les réductions vocaliques semblent moins fréquentes en anglais de Singapour (inspection F1/F2). Lin & Wang (2007) ont vérifié, par l'utilisation des corrélats acoustiques proposés par Ramus et par Grabe & Low, l'impression auditive que le chinois mandarin était une langue isosyllabique.

Parallèlement, peu d'études ont été menées sur l'utilité des métriques destinés à une application de discrimination L1/L2 ; et pour la plupart de ces études, la langue cible porte en général sur l'anglais (L2) mais aucune ne s'est réellement penchée sur les productions des francophones en langue anglaise.

Une première utilisation consiste à mettre en avant, ou non, une discrimination dans la production de deux langues chez les bilingues en utilisant la métrique proposée par Grabe & Low (PVI). Whitworth (2002) étudie le rythme des enfants bilingues anglais/allemand en utilisant le rPVI et le nPVI afin d'observer l'influence de la langue

¹¹ Nombreux exemples pourraient être cités mais le but n'est pas d'établir une liste exhaustive des études menées, nous citerons donc deux exemples représentatifs des travaux récents menés dans ce domaine.

maternelle des parents sur la production des enfants dans les deux langues. Carter (2005) analyse les productions d'espagnols natifs (L1), d'anglais américains natifs (L1), et de locuteurs espagnols bilingues en anglais à partir d'enregistrements d'interviews réalisés dans les deux langues. La métrique du PVI a été utilisée montrant des différences rythmiques entre les productions. Le taux de PVI des bilingues ainsi que celui des espagnols est bien plus bas que celui des natifs américanophones.

Quelques autres études tentent de discriminer des groupes d'apprenants, de divers niveaux, par rapport aux natifs. Une des premières analyses L1/L2 est celle proposée par Markus *et al.* (2003) qui présentent une étude comparative des deux méthodes de calcul proposées par Ramus (1999) et Grabe & Low (2000) : le but est de comparer les résultats obtenus des mesures acoustico-phonétiques des natifs létoniens et des apprenants russes ayant choisi le létonien comme seconde langue, sachant que le russe est classé comme langue isoaccentuelle (aura donc tendance à un fort taux de variabilité entre les intervalles) et le létonien comme langue isosyllabique (aura une faible proportion de variabilité entre intervalles). Le corpus utilisé est composé de 10 natifs létoniens et de 10 apprenants russes de deux niveaux différents (faible et fort) allant de 15 à 60 ans. La tâche était de lire un court passage d'un texte en létonien. Les résultats obtenus, selon la méthode Ramus, montrent des différences par la variable (ΔC : variabilité intervalles consonantiques) : les russes de bas niveau en létonien montrent une variabilité consonantique deux fois plus grande que les russes d'un bon niveau et que les natifs. La variabilité des intervalles vocaliques (ΔV) est presque identique pour les russes de bon niveau et les natifs, le taux est plus faible pour les russes dits « faibles ». Le pourcentage d'intervalles vocaliques (%V) ne montre aucune différence entre les trois groupes. Pour la méthode Grabe & Low, la variabilité des intervalles vocaliques (nPVI) montre des valeurs plus élevées pour les russes de faible niveau, mais aucune différence significative entre les russes « forts » et les natifs. La variabilité des intervalles intervocaliques (rPVI) est similaire pour les natifs et les russes de bon niveau, par contre le taux de variabilité des russes de faible niveau est beaucoup plus élevé. Cette étude met en avant, par des propriétés rythmiques et en utilisant des mesures acoustiques proposées par Ramus et Grabe & Low, les distinctions entre les schémas rythmiques de natifs et non natifs, de la même façon que ces calculs sont opérés pour distinguer les catégories rythmiques des langues. Pour conclure leur recherche, les

auteurs se questionnent afin « de déterminer jusqu'où les propriétés rythmiques peuvent caractériser les types particuliers d'accent étranger » (2003 :10).

Stockmal & al. (2005) s'inspirent de l'étude de Markus *et al.* et expérimentent les métriques de Ramus et de Grabe & Low sur des productions d'apprenants russes (L1) en létonien (L2). Les résultats obtenus pour les groupes de natifs létoniens et des apprenants d'un bon niveau sont très proches tandis que les apprenants d'un niveau faible s'éloignent des deux premiers groupes. Utilisant les mêmes métriques ainsi que le Rhythm Ratio (RR) de Gut & al. (2001), Nava & al. (2008) montrent une discrimination entre les productions des espagnols en anglais. Les apprenants espagnols de bon niveau obtiennent des résultats proches de ceux des natifs anglais, tandis que les apprenants de niveau faible ont des résultats proches des productions des natifs espagnols.

Lin & Wang (2005) choisissent d'utiliser uniquement la métrique proposée par Ramus afin d'étudier l'influence du chinois (L1) sur l'anglais canadien (L2). La discrimination des productions de chinois en anglais est établie par la combinaison ΔC , %V.

Les dernières études, présentées ci-dessous, analysent les productions d'apprenants afin d'observer l'influence du rythme de la L1 sur la L2. Le procédé utilisé est celui d'une analyse comparative des précédentes métriques. Sont également incluses les dernières propositions détaillées précédemment, à savoir le coefficient de variation des intervalles consonantiques (cvC) et le coefficient de variation des intervalles vocaliques (cvV). White & Mattys (2007) analysent l'interaction L1/L2 entre le castillan (L1) et l'anglais britannique, et entre le néerlandais (L1) et l'anglais britannique. Pour cette distinction L1/L2, les résultats montrent que VarcoV et %V offrent la meilleure distinction entre les deux niveaux natifs/non natifs. Ces résultats donnent des perspectives positives d'application des métriques permettant de distinguer des productions natives/non natives et d'observer ainsi l'influence du rythme de la L1 sur la L2.

Une seconde étude menée par Grenon & White (2008) analyse l'interaction entre l'anglais canadien et le japonais par des apprenants japonais. L'étude indique que les corrélats utilisés (%V, varco V, rPVI_C) révèlent des distinctions rythmiques entre les différentes productions des natifs et des apprenants. Les auteurs concluent que ces

métriques peuvent « indubitablement » être appliquées dans le cadre d'une distinction L1/L2.

De manière générale, ces études ont montré que les *scores* obtenus pour les productions des locuteurs L2 se situent à l'intermédiaire, c'est-à-dire entre celles des natifs de la L1 et celles des natifs de la langue L2. Il est à noter qu'en fonction de l'étude menée, différents corrélats étaient utilisés et tous n'ont pas donné le même résultat. Il semble donc que certaines combinaisons de paramètres soient plus efficaces que d'autres. Pour l'ensemble des travaux présentés ci-dessus, il est possible de conclure que les métriques appliquées permettent de donner une tendance rythmique des productions et de distinguer ainsi les différences « natifs/non natifs ». Tout particulièrement, les travaux de White & Mattys (2007) montrent une claire distinction entre les productions d'apprenants castillans et néerlandais avec des natifs anglais et mettent en avant l'utilité des métriques, plus particulièrement la combinaison VarcoV-%V dans la discrimination du rythme de locuteurs natifs/non natifs.

Synthèse

La présentation de la notion de rythme a montré que ce dernier jouait un rôle important dans le quotidien de l'être humain. Une définition s'est dégagée de cette notion complexe que nous proposons de décrire selon les termes de Di Cristo & Hirst (1977) comme « l'organisation temporelle des proéminences ».

Par la suite, les principales différences prosodiques énoncées dans ce chapitre ont permis d'établir certaines interférences rythmiques qui pourraient caractériser les productions du locuteur francophone en anglais et permettre de rendre compte des difficultés auxquelles l'apprenant est confronté.

Par ailleurs, il a été noté que la perception du rythme repose sur la répétition de syllabes ou de syllabes accentuées. Traditionnellement, le rythme est associé à l'idée d'unités isochrones, soit des syllabes, soit des pieds (Lloyd James, 1940 ; Pike, 1945). Néanmoins, l'incapacité à contrôler par des données acoustiques ce qui est perçu auditivement conduit à une remise en question des impressions auditives annoncées qui sont à la base de la classification des langues (isosyllabiques vs. isoaccentuelles). Plusieurs analyses instrumentales (Bolinger, 1965 ; Hoequist, 1983) ont été menées, mais les résultats obtenus ne se sont pas révélés satisfaisants.

Dauer (1983) s'appuie sur le postulat traditionnel de la catégorisation des langues et propose une liste de propriétés structurelles qui varient en fonction des langues. Une nouvelle catégorie de langues, appelées langues à rythme mixte, est créée et se situe entre les deux « opposés ». Le concept de classification stricte des langues s'assouplit, on parle alors plutôt d'un continuum.

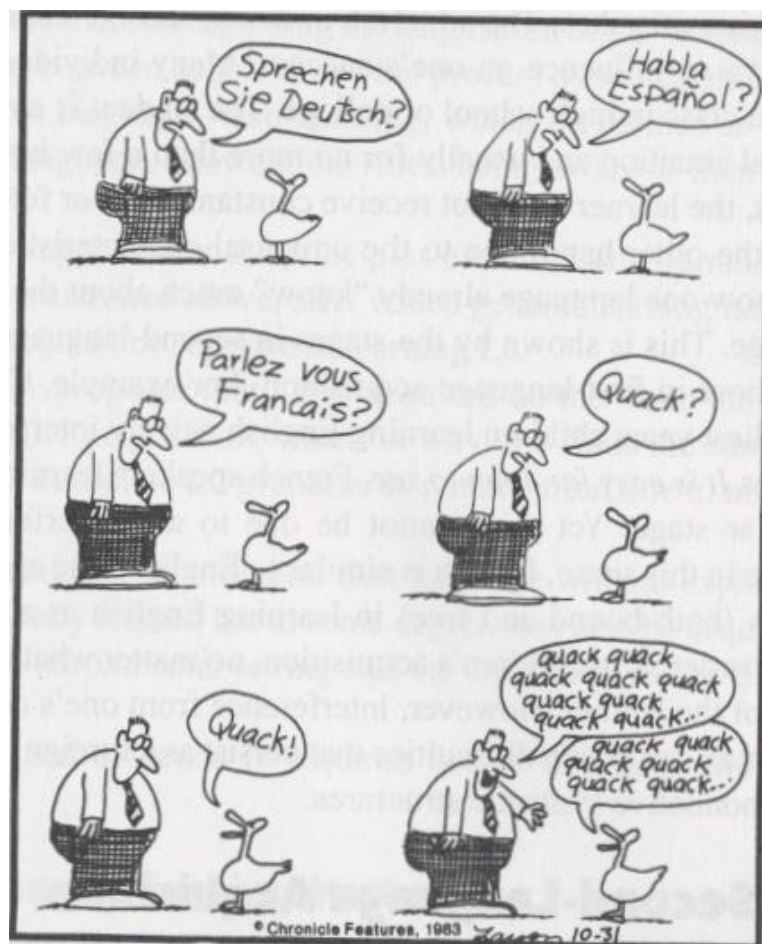
Récemment, la proposition de différentes métriques se basant sur les propositions de Dauer introduit de façon objective des données acoustiques, prenant en compte la durée des intervalles consonantiques et vocaliques (Ramus, 1999 ; Ramus & al., 1999 ; Grabe & Low, 2002 ; Dellwo & Wagner, 2003 ; White & Mattys, 2007). Certaines faiblesses telles que la non normalisation du tempo de certaines métriques, ou encore l'utilisation de petits corpus due à la lourde tâche de segmentation en unités consonantiques ou vocaliques, ont été mises en évidence. Malgré tout, l'application de ces métriques a donné des résultats

intéressants dans la discrimination des langues, ainsi qu'au niveau d'analyses inter-dialectales.

Par ailleurs, peu d'études (Nolan, 2000 ; White & Mattys, 2007) ont été menées sur l'impact du rythme de la L1 sur les productions de L2 et l'utilité que ces métriques pourraient jouer dans ce rôle de discrimination L1/L2. A notre connaissance, aucune recherche n'a été entreprise sur les productions de francophones en anglais.

Enfin, les recherches conduites par White & Mattys (2007), proposant l'utilisation de nouvelles métriques, pourraient être intéressantes afin d'évaluer la production rythmique des apprenants francophones en comparaison de celle des natifs et de tester si le rythme produit est plus proche d'un rythme français ou bien s'il se rapproche d'un rythme anglais. C'est sur cette étude que nous nous fonderons pour mener les analyses expérimentales de notre recherche, en procédant à une analyse comparative des métriques les plus utilisées dans la littérature, appliquées aux productions en anglais, de locuteurs francophones.

CHAPITRE 2. ACQUISITION D'UNE LANGUE



The far side. by Gary Larson. 1983, 1984 Chronicle Features, San Francisco,

dans Fromkin & Rodman (1998:347) *An introduction to Language*, Harcourt

Dans ce chapitre sur l'acquisition d'une langue, nous exposons les différentes phases d'acquisition de la prosodie et plus particulièrement du rythme de la langue maternelle (L1) puis de la langue cible (L2). Nous présentons différentes approches pédagogiques et nous analysons ce qui constitue un frein pour l'apprenant dans son apprentissage. Ensuite, nous discutons le contenu d'approches technologiques qui tentent de répondre aux difficultés des apprenants et pourraient être bénéfiques pour l'enseignement de l'oral en langue seconde.

2.1 Introduction générale

2.1.1 Langage : définitions

En premier lieu, nous nous devons de distinguer langage, parole et langue. Le langage est le moyen par lequel nous communiquons, la parole en est le vecteur principal et la langue est le système de langage propre à une communauté.

Le langage est donc un moyen d'expression, de contact, de communication. D'un point de vue linguistique et pour céder à la tradition, nous expliquons la notion de langage selon les termes de De Saussure (1968) qui définit le langage comme étant : *« la faculté qu'ont les humains de communiquer entre eux par des signes articulés »*.

D'un point de vue psychologique, puisque nous nous apprêtons à aborder le thème de l'acquisition du langage chez le nourrisson, le langage va permettre à l'enfant de créer des relations avec autrui ; il est intimement lié au développement et à la structuration du fonctionnement psychique ; le langage est une fonction structurante de l'organisation du Moi de l'enfant (Rondal, 1983).

La grammaire générative présuppose qu'une grande partie du langage est innée. L'art de communiquer nos idées dépend moins des organes qui servent à cette communication que de la faculté propre de l'homme d'avoir un langage fondé sur une combinatoire de signes arbitraires (il suffit de nous référer à la langue des signes utilisé par les sourds et muets). Lorsque l'on voyage dans le temps, nous nous apercevons que la parole était perçue au XVIIIème siècle comme un don, comme « le miracle de la parole ». Dans son discours physique de la parole, Géraud de Courdemoy (1668) apparente la parole à une innéité de la raison :

« Je désire seulement qu'on observe une vérité très importante que nous découvre cet exemple des enfants, qui est que dès la naissance, ils ont la raison toute entière, car enfin cette manière d'apprendre à parler est l'effet d'un si grand discernement et d'une raison si parfaite qu'il n'est pas possible d'en concevoir de plus merveilleux [...]. Il est évident que la raison toute entière est dès le commencement puisqu'ils apprennent parfaitement la langue du pays où ils naissent et même en moins de temps qu'il ne le faudrait à des hommes déjà faits ».

Cette remarque nous pose le problème de l'innéité du langage et fait ressurgir de nombreuses questions. Entre autres, dans le cadre de l'acquisition du langage et plus

particulièrement du système rythmique, on peut notamment se demander si la structure rythmique chez le nourrisson est acquise et/ou innée. Nous exposons, dans la partie suivante, une brève représentation des différents concepts sur l'origine du rythme, dans le cadre de la description des systèmes linguistiques qui sont directement liés au débat du caractère inné du langage.

2.1.2 Origine du rythme dans le langage : horloge interne

Cette question de l'innéisme du langage nous ramène à un vieux duel qui oppose deux concepts linguistiques, à savoir celui de l'innéisme prôné par Chomsky (1968), qui parle de tendance instinctive au langage et celui de constructivisme (ou comportementalisme) suivi par Piaget (1950), qui cherche à rendre compte des mécanismes de formation des connaissances. Dans le cadre de notre travail, à partir de ce débat oppositionnel, nous cherchons à savoir quelle serait l'origine du rythme dans le langage.

Certains chercheurs mettent en avant la prédisposition d'une horloge interne plus communément appelée « horloge biologique » qui serait naturellement ancrée en nous. Cette horloge nous permettrait de reporter le rythme naturel biologique sur d'autres activités, telles que la musique par exemple.

Pour Fraisse (1956) les rythmes sont acquis pendant le stade de la petite enfance et une fois ancrés en nous, ils nous amènent à la création d'autres rythmes sur les événements qui nous entourent. D'autres auteurs sont allés plus loin dans cette réflexion.

S'appuyant sur les travaux de Grosjean (1983) montrant l'importance de la syllabe dans la structuration temporelle du rythme, Fant (1991) a cherché à prouver par une analyse statistique, la régulation de la rythmicité syllabique par une horloge interne. Son analyse tend à montrer l'émergence « d'attracteurs syllabiques » situés à des valeurs multiples d'une horloge sur une base d'environ 140 ms, ce qui correspond à environ 7 Hz, c'est-à-dire, une valeur très proche de la fréquence propre à l'oscillation mandibulaire (cf. paragraphe suivant). Lors d'une étude comparée de l'anglais, du français et du suédois,

Fant (1991 : 248) a également remarqué que ces attracteurs sont aussi présents pour les phénomènes de pauses :

« the average inter-stress interval within a short time memory span of about 4 seconds preceding pause...synchronizes an internal beat generating clock which sets a preferred pause duration ».

D'autres auteurs, tels que Davis & Macneilage (1995) ont montré l'importance du rôle de l'oscillation mandibulaire sur la structuration du développement du langage. L'être humain aurait donc physiologiquement une rythmicité naturelle.

C'est sur cette même lignée de rythmicité naturelle de l'humain que Konopczynski (1986) donne au rythme une origine neurologique. En effet, les données récoltées par cet auteur ont montré un caractère isochronique initial et universel du rythme jusqu'au stade du babil infantile, pour finalement incorporer la structure rythmique de la langue maternelle. Ce point de vue résout la dualité des deux concepts traditionalistes, opposant Chomsky et Piaget, puisqu'il réunit à la fois le caractère inné initial d'une rythmicité due à une « horloge neurale », dans le stade du « tout premier Proto Langage », pour basculer ensuite vers une acquisition de la structure rythmique du système linguistique imposée par la langue maternelle.

Par ces diverses études, nous constatons la proposition d'une origine biologique au rythme par le biais d'une horloge interne, qu'elle soit simplement sous forme d'influence reconduite sur d'autres événements, ou qu'elle soit composante physiologique de l'humain (mandibules et neurones). A la lumière de ces travaux, il semblerait que le rythme soit présent dès les premiers instants de vie. C'est ce à quoi nous nous intéressons maintenant.

2.2 Acquisition de l'anglais langue maternelle (L1)

2.2.1 Introduction générale : le statut de la prosodie dans l'acquisition du langage

Cela pourrait surprendre, surtout lorsque l'on s'adresse à des non spécialistes de notre domaine, mais les éléments prosodiques sont en fait les premiers éléments acquis par l'enfant et ce, bien avant l'acte parolier. La recherche sur l'acquisition du langage a donné lieu à de nombreux travaux et a connu un développement particulièrement important dans

les années 60 lorsque l'intérêt des chercheurs a basculé des inventaires lexicaux vers des recherches menées en phonologie avec, par exemple, l'ouvrage de Jakobson (1941). Tout un éventail d'études a été mené allant des relations entre formes phonétiques et leur pertinence linguistique (Stampe, 1969 ; Smith, 1973 dans Konopczynski, 1979), jusqu'à la grammaire ou le développement syntaxique du langage infantin. Pourtant un élément important reste oublié : la prosodie. En effet, même si quelques pionniers tels que Lewis (1936) ou encore Kaczmarek (1953, cité par Konopczynski, 1986) ont tenté d'en tenir compte, ce n'est qu'en 1962, avec Faure, que la prosodie commence à prendre son importance et que la littérature sur l'acquisition du langage prend en compte cet aspect prosodique. Pourtant, la prosodie, avec tous les éléments suprasegmentaux qu'elle englobe, joue un rôle déterminant dans l'appropriation de la langue maternelle chez l'enfant. Nous savons que petit à petit, le nourrisson s'éveille à la prosodie avant même d'avoir acquis les éléments du niveau segmental. L'imprégnation naturelle d'une langue se fait d'abord par les éléments prosodiques, avant même les caractéristiques segmentales. Par la suite, nous verrons que dans les premières semaines de vie de l'enfant, la prosodie (outre la gestualité) est le seul moyen de communication.

2.2.2 Acquisition du rythme de la langue maternelle

Le premier constat est que les études focalisées sur l'acquisition du rythme en L1 sont peu nombreuses. Certains chercheurs l'expliquent par le fait que la nature de la structure rythmique, chez l'adulte, reste évasive ou n'est pas complètement élucidée (Kent & al., 1991 :135 ; Whitworth, 2002 :175) et que l'on n'a pas encore trouvé, de manière absolue, le moyen de la quantifier (Whitworth, 2002 :175). Malgré cela, nous avons tenté de faire état des travaux existants, en essayant de retracer l'évolution de la structure rythmique dans l'acquisition du langage. Notre propos ici, n'est pas tant de donner une description minutieuse des étapes de l'acquisition du langage, mais plutôt de nous intéresser à l'acquisition des éléments prosodiques et plus particulièrement du rythme¹².

¹² Même si notre travail cherche à se focaliser sur le rythme, il reste délicat de négliger les autres éléments supra-segmentaux, ces derniers ainsi que l'élément rythmique formant un tout tissé et complémentaire qu'est la prosodie ; il paraît donc difficile d'omettre les autres éléments prosodiques.

Perception rythmique du fœtus

Dès sa vie prénatale, le nourrisson semble soumis aux sons, entre autre par le biais de la voix de ses parents. Mais comment les sons parviennent-ils jusqu'au fœtus et comment ce dernier les perçoit-il ? A quoi ressemblent ces sons perçus de l'intérieur ? Nous nous sommes donc hasardée dans l'univers du monde utérin, essayant de comprendre quels éléments sonores étaient perçus et parmi eux, si le rythme de la parole en faisait partie. Nous tentons, premièrement, de trouver une justification anatomique à cette question, pour nous pencher ensuite sur des expériences qui démontreraient une perception possible des sons et des séquences rythmiques.

▪ *Éléments neurophysiologiques*

Nous exploitons brièvement les éléments neurophysiologiques du fœtus afin de comprendre l'existence d'une réalité anatomique qui soutiendrait l'hypothèse de certains linguistes (Querleu & Renard, 1981) ayant analysé une réelle audition fœtale.

La création de tous les neurones corticaux a lieu entre la sixième et la dix-septième semaine de gestation. L'organe de Corti est alors morphologiquement mature vers 4 mois. Il semblerait que vers 6 mois, le fœtus acquiert des possibilités de perception auditive. On sait donc que les systèmes sensoriels sont matures anatomiquement in utero. La perte de neurones et d'axones commence dès la fin de gestation. Après la naissance, le développement cognitif consiste en une « épuration » et un « aménagement » : la perte de neurones s'accompagne d'une création exubérante de jonctions entre les neurones, les synapses. Un neurone forme plus de mille connexions et en reçoit encore plus. Il peut recevoir plus de dix mille messages en même temps. Le cerveau de l'homme contenant plus de 100 milliards (10^{10}) de neurones (Williams & Herrup, 2001), les spécialistes ont estimé à 10^{15} le nombre de jonctions. Cette densité synaptique explose à la naissance. Le cerveau se sculpte ainsi sous l'influence de l'expérience interne et externe. Le fœtus baigne donc déjà dans une atmosphère rythmique ; grâce à un équipement biologique et cognitif sophistiqué, le nourrisson peut recevoir les sons et ce, même pendant la période intra utérine. C'est ce que montre l'expérience de Querleu & Renard (1981) que nous exposons dans la partie suivante.

▪ Approche expérimentale

Le domaine de la perception intra-utérine est largement étayé par les recherches menées ces vingt dernières années. D'une simple constatation que l'enfant « entendait » avant de naître, on tend à rentrer aujourd'hui dans le détail de ses performances auditives.

Dans une interaction constante avec la vie de sa maman, le fœtus grandit et se construit, baignant dans une atmosphère rythmique : le rythme respiratoire et le rythme cardiaque, réguliers et binaires de sa mère, le rythme de son propre cœur. A l'exception de l'odorat et de la vue, les sens du fœtus s'élaborent avant la naissance et c'est d'ailleurs l'ouïe qui est le premier sens actif. Notons, par ailleurs, que l'oreille possède trois fois plus de connexions au système cérébral que l'œil.

Pendant la période de gestation, le fœtus n'est pas à l'abri des bruits extérieurs. D'après Lecanuet (1995 : 9), « *l'utérus est un endroit relativement tranquille, les pressions moyennes étant comparables à celles généralement rencontrées à l'extérieur* ». Les hautes fréquences sont atténuées, et les principales propriétés acoustiques du signal sont préservées, la prosodie également. Les travaux de Querleu & Renard (1981) ont montré, en plaçant un hydrophone miniaturisé dans le vagin, puis dans l'utérus gestant, que l'enfant était dans une ambiance assourdie. Tous les sons lui parviennent filtrés en passe-bas, de sorte que, physiquement, il peut capter tous les bruits énumérés ci-dessus, y compris la voix maternelle (et selon l'auteur, cette dernière est tout particulièrement repérable dans les enregistrements) ; mais c'est surtout la partie grave des sons qui est acheminée, la zone aiguë, au-delà de 3000 Hz, étant atténuée.

C'est par le liquide amniotique, jouant le rôle de filtre passe-bas, que le fœtus est plongé dans un univers acoustique grave, qu'il entend et baigne dans une atmosphère rythmique, soit biologiquement soit par les bruits perçus. C'est ainsi que Konopczynski (1990 :172) émet l'hypothèse que le bébé, dès la période prénatale, nage dans une ambiance rythmée :

« Si, en outre, on retient l'hypothèse, non confirmée actuellement, qu'il existe chez le fœtus un apprentissage, une adaptation à des stimulations répétées [...] et sachant que le rythme biologique et le rythme de la parole ne sont pas synchrones, il s'ensuit que le rythme de la langue maternelle est l'un des premiers éléments linguistiques auquel est soumis le bébé, et ce dès sa vie prénatale ; a fortiori plus tard, lorsque, se trouvant dans un environnement sonore bruyant, lui parviendront, non point les éléments segmentaux, mais l'essentiel des mélodies et surtout des séquences rythmiques ».

Si l'auteur insiste sur le caractère non confirmé de l'hypothèse et le manque d'analyses, de plus récentes études renforcent l'idée que le rythme est déjà perçu et reconnu par le nourrisson et ce, dès la naissance. Reste à savoir si une exposition à la langue maternelle peut favoriser, dès avant la naissance, un réglage perceptif sur les paramètres phonétiques et/ou prosodiques qui caractérisent cette langue et la différencient des autres. La question est trop vaste pour être traitée en quelques lignes. Nous citerons l'exemple d'une expérience de lecture faite dans le dernier mois de grossesse, pendant lequel la mère lisait toujours le même passage à haute voix (Fromkin & Rodman, 1998). Il semblerait que grâce aux expériences auditives reposant sur l'habituation (apparition de réponses motrices ou cardiaques et disparition de ces réactions due à la répétition), les fœtus ont pu être entraînés à la reconnaissance de certains stimuli (histoire, extrait musical,...). Les résultats de l'expérience, obtenus en enregistrant le taux de succion, montrent qu'à la naissance les nouveau-nés rythment leur succion en fonction de la réception du passage de prose, lu cette fois par une autre voix féminine (passage habituellement lu par la mère en période intra-utérine). Les auteurs concluent que le fœtus est perceptif à des propriétés acoustiques du signal et que tout échantillon de langage, avec une intonation et un rythme normaux, alerte le fœtus et le pousse à régler son écoute sur cet échantillon, en augmentant alors son taux de succion. Le fœtus semble donc percevoir le rythme, l'intonation, la mélodie de la voix humaine.

Afin de mieux se rendre compte de l'ambiance utérine dans laquelle baigne le fœtus, il est possible d'écouter un extrait de vie familiale d'un couple, tel que cela serait perçu par le fœtus (la famille se retrouve autour du petit déjeuner, écoutant de la musique classique, et la maman fredonnant l'air écouté) (cf. annexes cd-rom, pour des exemples de fichiers son).

Emergences de structures sonores

▪ *Rythme dans les cris infantiles*

Le cri est la première expérience que le nourrisson expérimente dans sa vie post-natale et ce sera son seul moyen d'expression durant ses premiers mois de vie. En effet, de la naissance jusqu'à l'âge de 4 mois environ, sa communication se fera principalement par

des pleurs exprimant ainsi des sentiments (Fitzpatrick, 2002). Plusieurs études ont été menées analysant les cris de l'enfant et il semblerait que la structure rythmique du cri de l'enfant ait des conséquences sur les schémas communicatifs formés lors de la première année de vie. Lester (1985) analyse trois phases rythmiques présentes dans les cris des nourrissons qu'il classe en trois périodes :

- les périodes de longues durées, indiquant le moment et la fréquence des pleurs, qui auraient une implication dans le développement social de l'enfant. Des mesures acoustiques faites à 3 mois, ont permis de montrer à Lester une corrélation entre les pleurs du nourrisson et les structures rythmiques proposées lors d'une interaction enfant-mère. En effet, c'est souvent par le biais de stimulations rythmiques tels que le chant, les tapotements, (etc.) que les pleurs du nourrisson sont apaisés, ce qui permet à Lester de conclure que cela aide le nourrisson à établir des schémas temporels, en fonction des interactions sociales.

- les périodes plus courtes mettant en évidence une coordination entre cycle respiratoire (phase d'expiration) et pleurs ce qui crée un schéma rythmique.

- les périodes courtes, telles que les changements de fréquence fondamentale du cri. L'étude de Wolff (1969) établit un cadre descriptif d'un schéma rythmique de base d'une séquence de cris « spontanés » qui serait d'ailleurs observable 30 minutes après la naissance du nouveau-né, et resterait constante jusqu'au deuxième mois. Wolff affirme l'existence de séquences de temps régulières intrinsèques aux pleurs ; il observe qu'une séquence de pleurs s'articule comme suit : des pleurs, une pause de 0,2 seconde, un court sifflement inspiratoire, une pause, des cris.

■ *Perception du nouveau né*

Si dès les premiers instants de vie, un schéma rythmique, quelle que soit sa forme, est mis en place, plusieurs études montrent que le nourrisson est capable de percevoir le rythme de la parole, voire de distinguer deux classes rythmiques différentes (si tant est que l'on puisse parler de classes rythmiques).

Les travaux de Millot (1985) montrent que le nourrisson de moins de dix jours est capable de distinguer différents rythmes non langagiers. Dans son étude sur la décélération et l'accélération du rythme cardiaque infantin, Millot explique que l'enfant pose déjà une signification sur les divers rythmes et qu'il adopte un comportement de défense pour réagir à un rythme inhabituel. D'après ces expérimentations, le nouveau-né réagit différemment à un bruit de rythme cardiaque de 72 à 120 battements par minute (ce qui est l'équivalent de la mère et de celui de l'enfant éveillé), qu'à un même rythme mais produit par un métronome. L'auteur conclut que la divergence de réactions face aux différents rythmes est attribuée à l'expérience du rythme acquise pendant la période fœtale.

D'autres expériences ont été effectuées utilisant de la parole naturelle, à l'envers ou encore filtrée passe-bas. L'expérience de Mehler & al. (1988) montre que les nouveau-nés français de quatre jours arrivent à différencier le russe du français, par le biais d'une lecture d'un passage de texte lu par une bilingue franco-russe ; les préférences des nourrissons étaient indiquées par l'utilisation d'une tétine non nutritive. Les auteurs indiquent que les bébés parviennent à différencier leur langue maternelle d'une autre langue ; cependant, il semble que les langues à distinguer doivent être de classes rythmiques différentes pour être discriminées.

L'étude de Bosch & Sebastian-Gallès (1997) montre que des nourrissons de 4 mois peuvent distinguer deux langues de même classe rythmique (ici, l'espagnol et le catalan), si l'une d'entre elle leur est familière. Leur hypothèse se base sur le fait que le catalan possède des réductions vocaliques qui n'existent pas en Espagnol ; cette différence rythmique devrait être détectée par les nouveau-nés. Pour leurs analyses, les auteurs manient des stimuli de phrases normales et filtrées passe-bas et utilisent une procédure d'orientation visuelle avec des mesures de temps de réaction. Elles concluent que les nourrissons seraient capables de discriminer des langues par des indices rythmiques.

Une autre expérience (Nazzi & al., 2000) confirme les précédentes études et montre que la discrimination de deux langues d'une même classe rythmique est possible mais seulement, si l'une des deux est la langue maternelle du nourrisson.

L'expérience de Ramus & Mehler (1999) montre que le nourrisson possède dès la naissance une capacité à analyser le rythme de la parole. Afin de mener à bien leur étude, Ramus & Mehler ont procédé à des mesures acoustico-phonétiques sur huit langues ;

l'expérience consiste à démontrer que les sujets discriminent les langues selon les indices rythmiques. Les auteurs formulent l'hypothèse que les nourrissons perçoivent la parole avant tout, comme une succession de voyelles, séparées par des consonnes. Une segmentation en consonnes/voyelles d'un corpus de phrases parlées en huit langues a donc été faite. La durée des intervalles vocaliques et consonantiques a été mesurée. Les calculs ont été faits selon 3 variables : proportion des voyelles, écart type des intervalles vocaliques et consonantiques. Les résultats montrent que les valeurs reflètent la classe rythmique attribuée à chaque langue (cf. chapitre 1). Selon cette étude, le rythme d'une langue pourrait donc être signalé par des indices acoustiques et phonétiques, mesurables dans le signal de la parole.

Cette expérience se distingue par le fait que les auteurs utilisent une technique de resynthèse de parole qui permet, après une dégradation de stimuli, de ne garder que le rythme. Une autre expérience consistait à resynthétiser les phrases, en remplaçant les plosives par -t-, les liquides par -l-, les nasales par -n-, les semi-voyelles par -y-, les fricatives par -s- et les voyelles par -a-, éliminant ainsi les différences phonétiques entre les deux langues mais préservant leurs propriétés rythmiques. L'expérience a montré que la discrimination était effective, même avec la neutralisation de l'intonation. En revanche, sans les propriétés rythmiques, c'est-à-dire avec l'intonation seule, la discrimination ne se fait pas. Le test a montré que les nouveau-nés étaient capables de distinguer le japonais du néerlandais, et ce, par les propriétés rythmiques seulement. Les auteurs concluent que le nourrisson possède, dès la naissance, une capacité à analyser le rythme de la parole. Cependant, Ramus (2002) montre tout de même que rythme et intonation restent la combinaison la plus favorable à une meilleure discrimination. L'auteur formule l'hypothèse que cette capacité peut être utile aux nourrissons dans l'acquisition de la structure des syllabes de la langue maternelle. Le même type de remarque avait déjà été formulé par Kent & al. (1991 :143), considérant le rôle des schémas temporels dans l'acquisition du langage.

Vocalisation et babillage

▪ *Renforcement de la structure rythmique*

La prochaine étape de communication pour le nourrisson est celle que l'on connaît très communément sous le nom de « babillage », qui s'installe aux environs des 4 à 12 mois¹³. Avant de prononcer les mots, l'enfant pratique les sons, l'intonation et le rythme de la langue (Fitzpatrick, 2002). Il utilise des formes prosodiques exagérées, des pauses ainsi qu'une accentuation amplifiée. L'expérience de Pettito & Marantette (1991) montre d'ailleurs, que pendant la période de babillage, les contours intonatifs produits par les nourrissons sont très proches de ceux produits par les adultes. Face à ce développement langagier, l'entourage de l'enfant aura tendance à utiliser le « baby talk » (Dodane, 1999).

Di Cristo (2002) fait remarquer que des phénomènes articulatoires sont déjà présents et ancrés, et ce avant même que la parole n'apparaisse, donnant lieu à une rythmicité : de fait, on observe pendant la période du babillage, un mouvement répétitif d'ouverture et de fermeture des mandibules sur lequel le langage va se former et se développer. C'est le début d'une alternance ouverture/fermeture du chenal phonatoire mais avec des durées irrégulières. Ainsi Di Cristo explique qu'un cadre rythmique est déjà défini, avant même que la parole ne survienne.

C'est pendant cette période que la structure rythmique se renforce : l'identification du système rythmique va permettre à l'enfant de développer un outil de segmentation qui va conduire à la construction de structures syntaxiques et d'y ajouter par la suite les phonèmes (Di Cristo, 2002).

Dès deux mois, le nourrisson accorde un traitement particulier à la syllabe, qu'il perçoit comme un tout, et non comme une combinaison d'éléments distincts. La prosodie naturelle force l'écoute des bébés et par ce fait, elle aiderait leur attention à se porter sur les

¹³ La durée diffère selon les auteurs, nous avons donc englobé toutes les dates données en définissant cette période.

variations phonétiques. La prosodie serait une « glue perceptive » pour les séquences de parole.

Face à ce « comportement rythmique » observé dans la phase de babillage, plusieurs études tentent de rendre compte du développement langagier de cette phase en utilisant comme unité d'analyse, la syllabe, allant de la simple phonation à une combinaison de type CV.

Oller (1978) distingue deux périodes à l'intérieur de la phase « babillage ». La première, est une séquence répétée de syllabes ouvertes, type CV qu'il nomme le babillage canonique (ba-ba-ba-ba, par exemple). Suite à cela, (d'environ 6 à 10 mois) intervient la période de babillage multi-syllabiques qui consiste en un enchaînement de syllabes variées, mais toujours de type CV (ba-bi-da-de, par exemple). Oller montre dans son étude que l'émergence de syllabes s'accompagne de variations prosodiques telles que changement de la qualité de voix, de la mélodie et de l'intensité ainsi que des variations du taux de pauses.

Elbers (1982) observe son propre fils et distingue trois types de séquences développementales dans le babillage de l'enfant ; elle constate : syllabe simple, syllabe répétée, syllabe concaténée. Un bémol pour cette étude qui repose sur une observation et non sur des analyses acoustiques concrètes.

Kent & al. (1991) expliquent que le babillage du bébé partage une caractéristique avec la parole qui est une série de structures de mouvements organisés en syllabes. Une attention particulière est portée à la syllabe en CV à cause de sa primarité dans le babillage. Selon l'auteur, les vocalisations permettent à l'enfant de produire des syllabes CV dans une structure prosodique large, qui peut être modulée de façon à produire des changements de durée, d'intensité ou de fréquence fondamentale.

D'un point de vue perceptif, l'expérience de Burns & al. (2007) montre, lors d'une analyse comparée de nourrissons monolingues (anglais) et bilingues (anglais/français), que de 6 à 8 mois aucune distinction n'est faite entre les deux catégories ; les nourrissons répondent de la même façon, indépendamment de leur environnement linguistique, pour distinguer le français et l'anglais. En revanche, de 10 à 12 mois, les monolingues montrent plus de difficultés à percevoir les deux langues ; l'analyse montre qu'ils se réalignent sur les caractéristiques de leur langue (l'anglais), qui vient compromettre le développement

des représentations phonétiques de langues additionnelles (alors que les bilingues distinguent parfaitement les deux langues).

- *Prosodie : structure d'accueil*

L'écoute des premiers babillages montre que le ton caractéristique de la langue a été déjà bien saisi par nos jeunes apprentis ; ce n'est pas par un espéranto universel que le nourrisson commence à parler ! Les schémas rythmiques et intonatifs d'une langue constituent la véritable charpente à partir de laquelle le système phonologique pourra s'organiser. Selon Konopczynski (1999 :68) la prosodie constitue en cela « la structure d'accueil de tous les autres éléments de la parole, qui sans elle, restent vides de signification ».

Acquisition de la structure rythmique

Suite à la période de babillage, une phase plus longue, allant du premier mot à des combinaisons de mots, jusqu'à la phrase plus ou moins élaborée, caractérise la période du langage articulé. Nous tentons de nouveau de nous attacher seulement à la partie rythmique de l'acquisition, bien qu'il semble y avoir une totale interaction dans le développement des niveaux segmental et suprasegmental.

A partir de deux ans, l'enfant valide les caractéristiques rythmiques de base de sa langue maternelle.

Certaines études montrent que l'acquisition des durées segmentales et l'acquisition du rythme vont implicitement de pair.

Kent (1976) explique que les durées segmentales auront un effet sur l'apparence rythmique du langage. Ses analyses révèlent que la durée des segments vocaliques est généralement acquise vers l'âge de trois ans dans les syllabes accentuées, et à l'âge de 5 ans pour les syllabes inaccentuées. La production des durées consonantiques devient égale à celle des adultes seulement à l'âge de 8 ans pour les syllabes accentuées, et 10 ans pour

les syllabes inaccentuées. Par son étude, il indique donc que le schéma rythmique ne peut s'acquérir complètement, tant que les durées segmentales ne sont pas stables.

Deux autres études mettent en avant la tâche difficile, pour un enfant, d'arriver à produire des syllabes réduites lorsque la langue possède un taux de variabilité vocalique important. Konopczynski (1995) et Grabe & al. (1999) dirigent une étude comparée dans plusieurs langues, s'appuyant sur le principe de dichotomie des langues.

Konopczynski compare le français (isosyllabique) et l'anglais (isoaccentuel). Elle postule qu'il serait plus facile d'acquérir la structure rythmique du français et conclut que l'acquisition du rythme est liée à la complexité de la structure rythmique, comprenant la prédiction de l'accentuation lexicale ainsi que l'allongement final.

Grabe & al. comparent trois langues (le français, l'anglais et l'allemand) sur un corpus d'enfants de 4 ans. Les mesures de variabilité des intervalles vocaliques impliquent un classement possible des langues, selon l'ordre d'acquisition du rythme : les plus rapides sont les Français, puis les Allemands et enfin les Anglais. Les auteurs concluent que moins il y a de variabilité des intervalles vocaliques comme en français, plus la structure rythmique sera acquise facilement.

Cruttenden (2001 :254) rapporte dans son ouvrage, sans faire part d'études analytiques à ce sujet et précisant qu'il y a très peu de preuves à l'appui, que l'acquisition du rythme par les natifs anglais se ferait d'abord par la production des formes pleines des mots grammaticaux, et que la longueur des syllabes tendrait à être égale ; cela sous-entend qu'un anglais commencerait à communiquer avec un rythme isosyllabique.

■ *Acquisition de l'accent*

D'un point de vue imitatif, il semble que les accents soient très bien perçus au départ et ce, quel que soit l'endroit où ils se trouvent (Leopold, 1948 ; Allen, 1972) ; par ailleurs, les enfants le reproduisent correctement. Ce serait vers 5 ans que les enfants ne reproduiraient plus que les modèles correspondant à leur langue maternelle, et commenceraient à transformer ceux qui leur paraissent différents. Allen explique que l'enfant perdrait petit à petit les capacités auditives parce qu'il ne s'en sert pas.

L'étude de Todd & Reynolds (1973, cité par Konopczynski (1986)) montre que les mots accentués, porteurs de sens, sont mieux retenus que les mots inaccentués.

▪ *Indices prosodiques à la structuration des énoncés*

Une fois l'acquisition du langage plus avancée (maîtrise de quelques mots, début de syntaxe etc.), comment l'enfant va-t-il parvenir à extraire les phrases, les syntagmes et les mots, des unités qui font sens ? Nous mettons en avant la capacité des nourrissons à détecter les caractéristiques physiques des sons de la parole. Dans toutes les langues, le rythme et l'intonation font partie de la structure formelle de la parole. Dans certaines d'entre elles, les traits mélodiques convoient des différences de sens et de structure. Ainsi, en chinois mandarin, la forme « ma » peut avoir quatre sens différents, correspondant aux quatre tons utilisés dans cette langue.

L'utilisation des indices prosodiques dans les différentes langues a montré leur rôle dans la structuration de ces langues. C'est la prosodie qui va fournir aux nourrissons des possibilités de segmentation de la parole continue en unités de sens. Les différents types d'unités de parole, (propositions, syntagmes, mots) en tant que formes organisées, sont repérés progressivement : les premières segmentations portent sur les unités larges, bien finies prosodiquement, qui ensuite faciliteront l'identification de structures plus fines, qui à leur tour, permettront l'identification des mots. Le rôle des indices prosodiques apparaît clairement.

C'est aussi la simplification et l'exagération des structures et de l'intonation adoptée par l'adulte, qui facilitent la tâche de segmentation de l'enfant. La prosodie va alors jouer un rôle majeur dans la structuration. Ce rôle n'est d'ailleurs pas seulement joué pendant la période infantile.

De manière générale, la prosodie a une fonction structurante et organisatrice. Hirst (1977) illustre cette fonction par l'exemple d'un énoncé, dont seule la prosodie permettra d'en distinguer deux sens : « he hit the woman with a stick », peut signifier, que l'homme a utilisé un bâton pour frapper la dame, le découpage en unités intonatives correspondrait alors à : /he hit the woman/ /with a stick/ ; ou bien que l'homme a frappé la dame qui est munie d'un bâton ; dans ce cas, le découpage serait /he hit/ /the woman with a stick/.

Au travers des différentes études menées sur l'acquisition du langage au cours de ces dernières années, il est aisé d'observer et de constater que le rythme tient une place

prépondérante dans l'acquisition de la langue maternelle. Les éléments segmentaux arrivent au second plan se développant sur la « structure d'accueil » qu'est la prosodie. Mais qu'en est-il pour la L2 ? Dans la section suivante, nous discutons de la place de la prosodie dans l'apprentissage d'une L2 et des méthodes d'enseignement pratiquées, puis nous nous interrogeons sur ce qui peut constituer « une difficulté » pour des apprenants, et enfin nous exposons les solutions envisagées pour pallier les difficultés de l'apprenant.

2.3 Acquisition de l'anglais langue seconde (L2) chez l'adulte

2.3.1 Introduction générale : le statut de la prosodie dans l'apprentissage d'une langue seconde

Ces trente dernières années ont donné lieu à une véritable ascension de la prosodie dans les études linguistiques ainsi que dans d'autres domaines (psycholinguistique, (traitement automatique des langues (TAL), etc.). Autrefois en marge de la linguistique, la prosodie semble aujourd'hui ubiquiste, et ce, à de nombreux niveaux d'analyse. Elle bénéficie aujourd'hui d'un statut fort différent de celui « subi » plusieurs décennies auparavant. Selon les propos de Di Cristo (2004 :68), on assiste à un changement soudain du statut de la prosodie, jadis véritable « *Cendrillon de la communication* » pour en arriver à une « épidémie de *prosodimania* ».

Si l'essor de la prosodie est significatif, notamment à partir des années 80, cette affluence masque malgré tout, une absence de réflexion didactique sur l'enseignement de l'oral. En effet, l'enseignement de la phonétique semble avoir toujours posé un problème important pour les enseignants : que faut-il enseigner en priorité, et comment ?

L'enseignement et les recherches sur l'apprentissage des éléments prosodiques n'ont été entrepris, que de manière sporadique, cela malgré l'importance de la prosodie soulignée par plusieurs phonéticiens/didacticiens et ce, à de multiples reprises.

James (1976) souligne l'importance de l'enseignement des traits prosodiques d'une langue seconde, qui, selon ses propos, a été fortement négligé. Propos confirmés par Major

(2001 :14) : l'auteur souligne que l'enseignement des facteurs prosodiques est longtemps resté en retrait, au détriment de l'apprentissage de la prosodie : « *Segmentals have been the most thoroughly studied area in L2 phonology* ».

Munro & Derwing (1995) ; Hardison, (2004) rapportent que les éléments prosodiques jouent un rôle très important, voire plus important, que l'aspect segmental dans le discours en L2.

James (1976:227) évoque les conséquences négatives d'une situation de communication, dans laquelle l'intonation serait incorrecte :

« [...] une intonation incorrecte ne risque pas seulement de créer des malentendus. Elle a parfois comme conséquence une rupture complète de la communication ».

Pritchard (1985) fait part du « contentement » général, face à l'enseignement d'une articulation des segments en français L2, et constate que l'intonation, en tant que phénomène suprasegmental, n'a pas souvent fait l'objet d'études systématiques.

Vaissière (1989:3), citée par Freland-Ricard (1995) fait remarquer que lorsqu'un énoncé est jugé par un auditeur, la prosodie est tout aussi essentielle que la grammaire et la syntaxe : « *Une phrase prosodiquement anormale est aussi peu naturelle qu'une phrase agrammaticale* ».

Ginestet (2002 :87) évoque la portée de l'intonation dans les échanges discursifs :

« Il s'agit d'un phénomène très important vu que dans la communication, en dehors des gestes, nous utilisons, pour nous faire comprendre, l'intonation à 38% et les mots pour seulement 7%. Cependant au niveau de la didactique du FLE¹⁴, le problème de l'intonation est encore laissé de côté ».

Crystal (1969) souligne que le vocable « suprasegmental » était ce qui venait « en plus » d'autres éléments, et seul le segmental était prioritaire ; par cette approche, on considère alors que les éléments prosodiques sont redondants et que leur étude n'est pas nécessaire, contrairement à celle des éléments suprasegmentaux. Cela peut, entre autres,

¹⁴ Français Langue Etrangère

expliquer pourquoi, pendant de nombreuses années, seul l'enseignement des phonèmes paraissait crucial.

Si l'on retrace brièvement l'évolution de la prosodie dans le domaine de l'apprentissage des langues, si l'on ne peut pas parler ici de « *prosodimania* », on peut constater une ascension de la prise en compte des facteurs prosodiques, même si ces derniers ont longtemps été absents de l'enseignement d'une langue.

Les années 60-70 représentent un souffle novateur pour l'enseignement de la prosodie, notamment avec des premiers essais de méthodes d'apprentissage de la prosodie, lancés par Guberina (1965), suite à l'élaboration d'une approche verbo-tonale.

Par la suite, un peu plus de recherches ont été entreprises, mettant en avant le rôle fondamental de la prosodie dans l'apprentissage d'une seconde langue.

Alliaume (1989:98) fait primer l'apprentissage de la prosodie d'une langue seconde sur la grammaire. L'auteur explique que l'apprentissage de la grammaire ne serait possible que par le biais des éléments prosodiques :

« Un étudiant qui perçoit mal les éléments prosodiques de la parole étrangère est privé du seul véritable moyen d'intégration dynamique de la grammaire de la langue ».

L'auteur explique que l'intégration du système prosodique de la langue mère est une contrainte pour l'apprentissage de la prosodie d'une seconde langue. Alliaume (1989:98) préconise donc une méthode, afin de « déconditionner l'apprenant de la prosodie de sa langue maternelle, en le reconditionnant à la prosodie de la langue étrangère qu'il apprend ; ce serait lui faciliter l'intégration de la grammaire de cette langue, des contenus sémantiques et affectifs du discours [...] ». C'est ce que l'auteur a cherché à exploiter par la création d'I-lang (la méthode de l'auteur est développée en fin de ce chapitre).

Les années 90 semblent « voir » le retour des manuels de phonétique de correction articulatoire, qui reprennent, sur les principes classiques de l'apprentissage des phonèmes (en exemple, Ginésy (1995), quelques exercices sur l'intonation et le rythme sont proposés.

Actuellement, même si encore parcimonieuses, les études sur « l'accent étranger » commencent à prendre en compte l'aspect prosodique de ce phénomène. Lepetit (1992) faisait remarquer qu'à la fin des années 80, les travaux attachés aux éléments

suprasegmentaux liés à l'accent étranger, ont fait l'objet de peu de recherches et pour la plupart, ces études ne portent que sur l'enseignement de l'intonation de l'anglais. A l'heure actuelle, les recherches sur l'influence des facteurs prosodiques sur l'accent étranger commencent à émerger (Magen, 1998 ; Major, 2001 ; Jilka, 2000, Munro & Derwing, 2001 ; Poquelin, 2009).

Ces études récentes montrent l'importance du rôle de la prosodie, à la fois dans la production des apprenants mais également dans la perception d'un accent étranger. Jusqu'ici, l'accent étranger a toujours été considéré comme causé par des stratégies articulatoires mal imprégnées ou mal reproduites. Si les facteurs prosodiques n'ont guère été mis en cause, il est indispensable de considérer une production, à la fois au niveau segmental mais également suprasegmental ; il semble difficile d'envisager un énoncé oral sans support prosodique. Il faut donc considérer à la fois la structure des sons et la prosodie, qui sont essentielles dans l'apprentissage. Même si nous constatons une avancée dans la considération portée à la prosodie par la communauté scientifique et enseignante, il reste pourtant difficile à croire, au vu des approches pédagogiques, que la prosodie ait réellement trouvé sa place dans l'enseignement des langues. Pourtant, le nombre de travaux portant sur l'acquisition de la prosodie de la L2, de plus en plus grandissants, se concentrent essentiellement sur des notions d'erreurs de prononciation ou bien d'accent étranger, et examinent la contribution des erreurs segmentales et suprasegmentales dans la perception de l'accent étranger. Ainsi plusieurs études, dont celles de Jilka (2000), montrent que les erreurs segmentales ont un effet moindre que les erreurs prosodiques sur l'intelligibilité et la perception auditive d'un accent étranger dans les productions d'apprenants.

Si la progression de la prise en compte de la prosodie dans l'apprentissage des langues a été lente, on peut observer qu'aujourd'hui, dans la littérature, son rôle fondamental (i) dans l'acquisition L1/ L2, et (ii) dans la perception de l'accent étranger, est reconnu. Les recherches fondamentales sur le rythme et l'intonation se sont multipliées et ont largement progressé. Malgré tout, cette avancée reste insuffisante, et la prosodie continue d'être mise à l'écart au profit de pratiques articulatoires. Pourtant la prosodie devrait être à la base de l'enseignement, comme elle l'est à la base de l'acquisition de la langue maternelle. De nombreuses futures études restent incontournables, notamment sur son fonctionnement, mais aussi son dysfonctionnement en L2. Des travaux

interdisciplinaires réunissant phonéticiens, didacticiens, enseignants, informaticiens, techniciens seraient les bienvenus afin de poursuivre l'étude de concepts théoriques mais aussi de trouver une application dans l'apprentissage d'une langue seconde.

A partir de ces constats, nous nous intéressons aux différentes méthodes ou approches actuelles destinées à l'apprentissage d'une seconde langue par la prosodie et notamment par le rythme ; nous tentons également de comprendre en quoi la prosodie constitue une barrière dans l'apprentissage L2.

2.4 Acquisition du rythme de la langue seconde

Depuis les années 80, la floraison de travaux dans le domaine de l'acquisition des langues secondes a permis d'importantes avancées. Ces dernières ont très certainement été rendues possibles par les résultats d'études de grande ampleur, à la fois longitudinales, c'est-à-dire suivant des locuteurs, au long des étapes successives de construction d'une nouvelle langue, et aussi translinguistiques, c'est-à-dire comparant les évolutions chez des groupes d'apprenants de différents couples de langues. Mais là encore, on peut constater le manque de travaux dans l'acquisition de la prosodie des L2 ; si les études sur l'acquisition du rythme en L1 sont peu nombreuses, celles sur l'acquisition du rythme en L2 sont d'autant plus rares. En effet, il existe une réelle paucité des travaux menés dans le domaine de l'acquisition du rythme de la parole en langue seconde.

Analyse et constat

L'étude d'Adams (1979) se place dans le cadre de l'acquisition du rythme de l'anglais par des apprenants étrangers. Suite à une série d'analyses auditives et instrumentales, l'auteur montre que le rythme des apprenants diffère de celui du rythme des natifs anglais et ce, selon plusieurs paramètres : (i) le manque d'écart de durées entre les syllabes accentuées et inaccentuées, ce qui sous-entend que les syllabes inaccentuées ne sont pas assez réduites, (ii) le manque de phénomènes de liaisons, (iii) des pauses inappropriées et (iv) des accents lexicaux mal placés. Les remarques d'Adams vont pouvoir être mises en évidence par les métriques, proposées à partir des années 90 ; en effet, les éléments d'un rythme inapproprié seront saisis par la mesure des intervalles vocaliques qui mettent en évidence, par exemple, les problèmes de différence de durées entre syllabes accentuées et inaccentuées, ou encore les intervalles consonantiques qui

reflètent le manque de liaison, qui impliquerait l'insertion ou la perte de consonnes dans le flux de la parole.

Cette première étude met donc en avant un certain nombre de problèmes rencontrés par les locuteurs de L2 et indique à quel niveau se situent les différentes variations dans le rythme des apprenants. La section suivante expose diverses propositions pédagogiques tentant de remédier à ces variations rythmiques.

Approches pédagogiques dans l'enseignement L2

- *Approche académique*

Nombreux sont les manuels sur l'enseignement de la prosodie présentant une approche tout à fait traditionnelle. En effet, il s'agit de donner les connaissances théoriques, en exposant les diverses règles d'accentuation lexicale pour fournir ensuite une application avec des exercices cherchant à repérer la syllabe accentuée. Ainsi, Ginésy (1995) propose un « mémento » répertoriant toutes les règles d'accentuation lexicale (terminaisons fortes, les suffixes forts/neutres/mixtes, les préfixés, etc.). Il expose ainsi toute une série de règles permettant à l'apprenant de « calculer » la position de la syllabe accentuée dans le mot. Sont également présentées les règles d'accentuation de phrase et cinq schémas intonatifs de base de l'anglais (glide up/glide down/high jump/dive/take off). Une forme similaire se retrouve dans l'ouvrage de Carr (1999), dans lequel il consacre plusieurs chapitres au rythme de l'anglais avec des exercices ciblés sur, par exemple, l'accentuation lexicale et les réductions vocaliques. Le but de l'auteur est de donner aux étudiants quelques concepts clefs dans le but de les encourager, par la suite, à approfondir leurs connaissances (cf. « Preface for teachers »).

- *Approche quasi –traditionnelle*

Une première approche, ici, consiste à familiariser l'apprenant avec, uniquement, le rythme et l'intonation de la langue cible par le biais de suites déphonématisées dans lesquelles ne sont gardées que les éléments prosodiques.

Montredon & al. (1975) proposent une approche pédagogique d'exercices variés de prosodie pure, visant l'étude de certaines modalités expressives de base, telles que l'interrogation, la double question, l'énumération, l'exclamation, le doute, la surprise, etc. et ce, dès les tous débuts de l'apprentissage linguistique. L'apprenant doit écouter des enregistrements de suites syllabiques dénuées de signification ou encore des intonations à échelons régressifs (de la fin vers le début). La reproduction des séquences de type « dadada » lui permet ainsi de se sensibiliser au rythme et à l'intonation (par la suite est introduit un travail sur des micro-dialogues). Basé sur le même concept, Faure & Di Cristo (1977) proposent 67 petits dialogues représentant des situations de communication pour le français. Ces exercices visent plus particulièrement le travail de l'intonation ; le rythme n'est pas vraiment pris en compte et les exercices s'adressent à un niveau avancé d'apprenants.

▪ *Travail de la prononciation par le rythme*

Un autre argument en faveur de l'importance du rythme, prôné par certains professeurs et chercheurs, est celui de l'apprentissage de la prononciation de l'anglais. Dans nos recherches, nous avons constaté un certain nombre d'articles portant l'utilisation du rythme comme aide et entraînement à la prononciation. C'est sur cette idée que Dickerson (1989) propose un manuel appelé « Stress in the Speech Stream (SSS) qui est constitué de textes dédiés à l'exercice de prononciation. Le but de ce manuel est d'être plus particulièrement destiné à des apprenants, de niveau intermédiaire ou avancé, afin de « devenir des locuteurs compréhensibles ». La méthode se focalise sur le rythme en proposant des exercices sur les accents de mots, de phrases, les contractions, les phénomènes de liaison, et les réductions de voyelles. Le livre propose aussi des entraînements sur la prédiction du nucleus dans les énoncés.

Norris *et al.* (1995), reprochant aux méthodes d'être inefficaces en donnant de longues listes de vocabulaire, préconise la reconnaissance des schémas accentuels des mots, ainsi que de la structure rythmique de l'anglais, des fonctions du rythme et de l'intonation comme structure informationnelle de la phrase ; il recommande essentiellement l'enseignement des formes réduites afin que les étudiants prennent conscience de l'importance de l'accent. Norris cherche à donner aux étudiants un outil de

compréhension du fonctionnement de la langue, qui les amène à mieux écouter pour percevoir la langue.

C'est également sur cet enseignement des formes réduites et des réductions vocaliques que Cruttenden (2001), dans son ouvrage, porte son attention dans un encart spécial intitulé « Advice to Foreign Learners » (2000 :255) : il préconise aux apprenants de se concentrer sur l'apprentissage des formes faibles des mots grammaticaux, et de les considérer comme des prononciations régulières de la langue et non comme exceptions. Il conseille l'utilisation des formes fortes uniquement en cas d'emphasis ou de contraste. Par ailleurs, il précise que les apprenants, parlant une langue caractérisée par une isochronie syllabique telle que le français, devraient porter une attention particulière aux raccourcissements rythmiques des voyelles pleines positionnées dans les syllabes suivies de /@,I,i/ ; à titre d'exemple, l'auteur propose de pratiquer ce raccourcissement par le travail de paires de mots, telles que *short* vs. *shorter* ; *lead* vs. *leading*.

▪ *Approches non conventionnelles*

Une autre approche, qui au fil du temps semble devenir quasi conventionnelle car de plus en plus utilisée, est la confluence des éléments verbaux et para-verbaux. Il s'agit de la mise en place d'une gestualité qui accompagne la modalité verbale enseignée. Ainsi, la méthode consiste à apposer un geste « universel », compréhensible par des locuteurs de divers horizons, à une mélodie ciblée (mais là, différente selon les langues). Llorca (1998) commente un exemple à partir de l'expression de la déception, et compare plusieurs langues : en français, « dommage » serait prononcé avec une proéminence finale de mot accompagnée d'un mouvement de haussement des épaules ou d'un mouvement de la main ; l'allemand « wie schade » place une accentuation sur l'avant dernière syllabe, l'espagnol « que lastima » sur l'antépénultième, etc. Un geste caractéristique est donc associé à une prosodie particulière et facilite de cette manière l'apprentissage de la prosodie ; la gestualité est donc utilisée pour mieux percevoir et produire la prosodie de la langue cible. Llorca développe une technique pédagogique visant à installer le rythme de la parole par l'intermédiaire d'un rythme corporel en utilisant des rythmimots. Elle explique que le mouvement permanent caractérisant l'activité parolière est marqué dans l'organisation du discours et est donné par le rythme et l'intonation. La proposition faite ici se base sur une pédagogie gestuelle de la prosodie.

Au niveau prosodique, « prosodier » correctement une langue, autre que sa langue maternelle, implique, entre autres, une bonne perception de ses structures rythmiques et intonatives. Cependant, comme nous allons le voir, la perception phonétique d'une langue étrangère pose un certain nombre de difficultés. La section suivante présente les différentes difficultés auxquelles un apprenant est consciemment ou inconsciemment confronté lors de l'apprentissage d'une langue seconde.

2.5 Problèmes particuliers liés à l'apprenant

Nous avons vu dans une partie précédente concernant l'acquisition de la langue maternelle, que le nourrisson était capable de différencier divers contrastes non natifs, mais qu'au fil des mois passés, la discrimination devenait de plus en plus ardue, en raison de son exposition à la langue maternelle au détriment des autres langues. On peut donc postuler que la tâche d'apprentissage d'un adulte face à une langue seconde sera bien plus difficile dans la mesure où celui-ci sera moins sensible aux phonèmes mais aussi aux éléments prosodiques de cette nouvelle langue.

Il semble donc qu'avec l'âge, l'influence de la langue maternelle provoque différents types de difficultés que nous exposons ci-dessous.

Contraintes extra-linguistiques

Des facteurs extralinguistiques peuvent influencer la production des apprenants L2. Tarone & al. (1978) pensent que ce sont les facteurs sociaux, personnels (la motivation par exemple), affectifs ou situationnels qui doivent être incriminés, car ils forment un « crible psychologique ». Ce crible entraîne alors très vite une fossilisation des erreurs qu'il devient difficile de « combattre » une fois fixée. Anderson (1983) parle de « pidginisation », c'est-à-dire que cela correspond à une acquisition linguistique où l'exposition à la langue est limitée, et où des facteurs socio-affectifs et des contraintes de temps viennent affaiblir la saisie des données linguistiques.

Les facteurs proposés par ces auteurs rappellent l'impression selon laquelle certains apprenants, indépendamment de leurs niveaux respectifs de travail et d'intelligence, ont plus de facilités que d'autres à parler une langue seconde. Par déduction fataliste, on pourra

entendre que certains « sont doués » pour les langues contrairement à d'autres. La considération des facteurs psychologiques ainsi que l'ensemble des facteurs exposés ci-dessous, nous permettent de mieux considérer ces phénomènes et de mettre alors à profit les « causes inconscientes » des difficultés rencontrées par les apprenants pour un meilleur enseignement.

Contraintes physiologiques

Alliaume (1989) affirme que le rythme et l'intonation mettent en œuvre des processus musculaires et respiratoires particuliers à chaque langue. Il explique que ce sont des éléments constitutifs de la parole qui sont dynamiques et qui font appel à une activité respiratoire et musculaire particulière, propre à chaque langue, d'où la difficulté de modifier un système respiratoire et musculaire ancré dans la langue maternelle. Ainsi, l'aspect physiologique conditionnerait le bon apprentissage prosodique de la L2 (Alliaume, 2008: 14):

« Chacun est physiologiquement conditionné par sa langue maternelle, ce qui implique une capacité perceptive réduite aux sons, rythmes, intonations de toute autre langue, mais aussi une sorte de « sclérose » des organes impliqués dans la production de parole dès qu'il s'agit de « dire » dans une autre langue »

Un autre « outil », très important pour l'apprenant dans son apprentissage d'une langue seconde, est bien évidemment son oreille. Cependant, l'oreille a une capacité limitée de discrimination. Landercy & Renard (1977 :185) expliquent que chaque personne a un seuil de discrimination et d'intelligibilité qui lui sont propres, et qu'il existe un seuil d'audibilité, universel, qui ne permet pas d'entendre des sons au-dessous de 60 Hz chez l'adulte. L'audition est le fruit d'un mécanisme complexe assuré par les deux oreilles, pour permettre la perception. On distingue l'audition, qui est donc le mécanisme dépendant de chaque sujet, de la perception qui serait l'activité mentale ; à partir de cette distinction, nous pourrions postuler que les problèmes d'audition relèveront des causes physiologiques, tandis que les problèmes de perception relèveront des causes linguistiques. L'être humain a donc un mécanisme d'audition défini selon un seuil minimum et maximum, qui permet la perception d'une palette de sons (sons de la langue maternelle) ; lors de l'écoute d'une langue étrangère, la barrière linguistique empêche la perception de sons nouveaux, malgré les prédispositions auditives possibles.

L1 source d'influence de la L2

Les études menées en neurologie montrent, qu'à partir de 10 ans, le cerveau opère auditivement en fonction de la langue maternelle ; dès lors, l'influence de la L1 est gravée dans les circuits neuro-moteurs de l'apprenant. On a donc ici une incapacité d'ordre neurologique, représentée par la perte de la plasticité cérébrale à l'âge adulte. Avec l'âge, l'emprise de la L1 grandit et provoque des phénomènes d'interférences avec les autres langues.

▪ *Crible et Interférence*

Cette première contrainte est établie par la langue maternelle elle-même. Le système prosodique (de la L1) fonctionne comme un état réflexe : il est ancré dans le cerveau et il prend le pas sur le système de la seconde langue. C'est ce que Troubetzkoy (1949) appelle « le crible phonologique » :

« Le système phonologique d'une langue est semblable à un crible à travers lequel passe tout ce qui est dit [...]. Les sons de la langue étrangère reçoivent une interprétation phonologiquement inexacte, puisqu'on les fait passer par un crible phonologique de sa propre langue ».

L'apprenant aura donc tendance à calquer sur la langue cible le modèle rythmique et intonatif qu'il applique depuis sa naissance ; la langue maternelle constitue donc un obstacle dans l'apprentissage d'une seconde langue, et plus particulièrement avec les années passant. Ce sont les éléments linguistiques de la langue mère qui empêchent l'apprenant de s'imprégner et acquérir la langue cible.

Le deuxième phénomène, similaire à l'idée du crible phonologique est celui de l'interférence, terme emprunté à la physique pour désigner «une rencontre, conjonction de deux séries distinctes de phénomènes ; interaction» (Larousse, 2003). Dans la didactique actuelle, lorsqu'on observe dans une production d'un apprenant L2, un glissement des éléments de la langue maternelle vers la langue cible, on parle plutôt de «fautes» que «d'interférences». Ce terme a été introduit par Weinrich (1953:1). L'auteur définit l'interférence comme un exemple de déviation des normes d'une langue, qui apparaît dans l'expression des bilingues comme un résultat de contact:

"Those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language, i.e. as a result of language contact, will be referred to as interference phenomena."

A partir de ce concept, Weinrich établit une grille d'analyse des erreurs d'interférences au niveau phonologique, morphologique, syntaxique et lexical.

Dubois *et al.* (1994) définissent l'interférence comme un phénomène qui entraîne un « emprunt » ou un « calque » de la langue mère sur la langue cible :

« On dit qu'il y a une interférence quand un sujet bilingue dans une langue cible A a un trait phonétique, morphologique, lexical ou syntaxique caractéristique de la langue B. L'emprunt et le calque sont souvent dus, à l'origine, à des interférences. Mais l'interférence reste individuelle et involontaire, alors que l'emprunt et le calque sont en cours d'intégration ou sont intégrés dans la langue A ».

L'apprenant étant marqué par des habitudes linguistiques de sa propre langue, ces phénomènes s'imposent à lui, et constituent donc des contraintes incontournables qui ne facilitent pas la tâche d'apprentissage d'une seconde langue. Seuls les plus jeunes sont moins marqués par ces contraintes. Mackey (1965) distingue deux périodes: l'âge du «contenant» et l'âge du «contenu»: la période du contenant se réfère à l'enfance, pendant laquelle on acquiert et développe le système prosodique sans posséder un grand contenu ; inversement, dans la seconde période, qui correspond à l'âge adulte, l'apprenant possède un grand contenu mais est limité par le système prosodique insuffisamment acquis. Cependant, plus l'enfant prend de l'âge, plus l'acquisition se corse: au cours de l'apprentissage, le locuteur va écarter progressivement les sons qui ne font pas partie de la langue ambiante. C'est pourquoi plus l'enfant vieillit, plus l'apprentissage d'une deuxième langue devient difficile ; il ne perçoit plus les sons qu'il a éliminés et la langue seconde sera marquée par le système prosodique de la langue maternelle ; on dira alors communément que l'on parle avec un accent étranger.

▪ *Transfert et Interlangue*

Dans les modèles d'acquisition, un fait, qui semble indiscutable, est que l'on retrouve très aisément dans les productions de L2 les caractéristiques phonologiques de la

langue maternelle, ce qui sera très certainement perçu par les natifs comme un accent étranger.

Lado (1957) propose de comparer les différences et similarités L1 & L2 afin d'anticiper les difficultés des apprenants, expliquant qu'il est nécessaire de livrer une description scientifique de la langue cible, de manière à la confronter à celle de la langue maternelle de l'apprenant. Lado pense que l'apprenant tend à reproduire les habitudes de sa langue maternelle dans ses productions en langue cible, l'auteur (1957:2) parle alors de transfert positif ou négatif :

« [...] the students who come in contact with a foreign language will find some features of it quite easy and other extremely difficult. Those elements that are similar to his native language will be simple for him, and those elements that are different will be difficult ».

Cette théorie est à la base de l'analyse contrastive (AC) dans laquelle les structures de L2 identiques à celles de la L1 s'acquièrent rapidement et assez facilement, ce qui correspond au transfert positif ; en revanche, les structures particulières à la L1 ou absentes en L2 s'acquièrent difficilement, ce qui correspond au transfert négatif.

Cette approche de l'analyse contrastive reçoit diverses critiques, notamment par sa vision limitative basée sur le fait que toutes les erreurs seraient le résultat d'un transfert. Or, d'autres facteurs (notamment ceux exposés ci-dessus) sont à prendre en compte dans les difficultés d'apprentissage des apprenants. Richards (1971) considère que le phénomène de transfert n'est qu'un facteur secondaire. L'auteur distingue alors deux types d'erreurs: (i) les erreurs intra-systémiques (« intra-lingual errors ») qui représentent les caractéristiques générales de l'apprentissage de règles, que ce soit en L1 ou en L2, dues à une mauvaise acquisition des règles, (ii) les erreurs développementales (« developmental errors ») qui sont basées sur une mauvaise compréhension des fonctions des mots, dues à des méthodes d'apprentissage contrastives de constructions de la langue cible.

Klein (1989) conteste l'analyse contrastive et formule un postulat contraire, stipulant que lorsque les structures des langues L1 et L2 sont proches, l'apprenant ne perçoit pas cette similarité, et cela le pousse à négliger sa production ; il fournira alors moins d'effort à l'apprentissage de la langue cible, ce qui réduit son processus d'apprentissage. A l'inverse, quand l'apprenant est conscient des différences de structures et des difficultés que cela engendre, cela le pousse à mieux acquérir un énoncé différent de

sa L1. Il nous semble pourtant qu'une structure identique entre les deux langues L1 et L2 pourrait au contraire, être assimilée plus facilement par un processus de mémorisation demandant moins « d'effort », la ressemblance pouvant mener à une meilleure assimilation de langue cible.

Au vu des différentes critiques envers l'analyse contrastive et au travers de l'observation des erreurs dans les productions d'apprenants, une image de l'erreur aboutit au concept d'interlangue (IL). En tout premier lieu, émise par Corder (1967), puis reprise par Selinker (1972), la notion d'interlangue décrit une phase transitoire du langage des apprenants en évolution entre la langue maternelle et la langue cible. Corder (1971) distingue les fautes occasionnelles et non systématiques relevant de la performance, et les erreurs plus systématiques qui décèlent la « compétence transitoire ». Dès lors, le rôle joué par la langue maternelle dans la production des erreurs est toujours pris en compte, en revanche la L1 n'est plus le seul facteur en cause. Les caractéristiques de l'interlangue incluent toujours la notion de transferts, mais d'autres stratégies d'acquisition sont reconnues comme la simplification. Pour Selinker (1972), cette langue transitoire résulte de quatre processus psycholinguistiques. L'auteur élabore une grille dans laquelle il place en (i), les transferts linguistiques (ainsi que le transfert d'enseignement (« transfer of training ») qui correspondent à la façon dont l'apprenant va privilégier certaines formes, selon la manière dont elles ont été enseignées), en (ii) les stratégies d'apprentissage propres à l'apprenant (comme l'évitement, la simplification), et enfin, en (iii) les stratégies de communication et le phénomène d'hyperm généralisation des structures de la L2. A cela, Selinker ajoute ce qu'il appelle le stade de « fossilisation », qui correspond au stade d'apprentissage où le niveau de l'interlangue de l'apprenant stagne, dans la mesure où le système linguistique de ce dernier se rapproche fortement du système de la L2, mais où subsistent des formes linguistiques de la L1.

Nous pouvons le constater ici, plusieurs théories exposant les modèles d'acquisition en L2 coexistent. Nous avons cité ici, les principaux concepts étayés dans la littérature qui représentent des théories « générales » dont le point commun est bien l'acceptation des concepts de « transfert » ou « d'interférence ». Basés sur ces mêmes concepts, il existe bien d'autres modèles théoriques de l'acquisition, tels que le Native Language Magnet Theory (Kuhl & Iverson, 1995), ou encore le Perceptual Assimilation Model (Best, 1995). Nous citerons ici, plus en détail, la théorie du Speech Learning Model (SLM) de Flege

(1995). Selon le modèle SLM, l'accent étranger serait dû à la forte tendance de l'apprenant à classer un son L2, voisin d'un son L1, dans une catégorie phonique (perceptuelle ainsi que productive) préexistente. Un tel son serait représenté dans une catégorie que l'apprenant aurait développée au cours de l'acquisition de la langue maternelle (voir aussi Best, McRoberts & Goodell, 2001 ; Kuhl, 1998 ; Kuhl, 2000). L'auteur explique donc, que lorsque deux sons sont quasiment similaires, l'apprenant aura une plus grande difficulté d'acquisition du son de la langue cible. En revanche, les segments différents peuvent être acquis sans grande difficulté car la différence de perception entre les deux sons est plus évidente pour l'apprenant. Flege et Hillenbrand (1984) ont démontré que les apprenants tardifs anglophones parviennent à reproduire la voyelle française /y/ d'une manière authentique. Les auteurs concluent qu'il convient alors d'examiner la prononciation de segments qui sont très voisins dans les L1 et L2.

Cependant, comme on peut le constater à partir des explications données sur l'exemple du modèle de Flege, la plupart des recherches sur les modèles d'acquisition en L2 restent concentrées sur les phonèmes uniquement, et malheureusement très peu d'études portent sur l'acquisition des traits prosodiques en L2.

Malgré la prédominance des travaux portant sur l'aspect segmental de la L2, Lado a consacré une petite partie de sa recherche, sur les aspects d'accentuation lexicale et de rythme ; cependant, aucun modèle d'acquisition de la prosodie n'a réellement été proposé.

Pourtant, à la même époque, Abercrombie (1967:36) souligne bien que les premiers aspects acquis (donc la prosodie) sont les plus durs à modifier, une fois le locuteur adulte. Une description solide des phénomènes de transfert prosodique permettrait d'établir les caractéristiques de la L1 qui influencent la prosodie des productions en L2. Le concept de transfert prosodique a donc été introduit, mais peu de travaux ont été menés dans cette optique-là.

▪ *Interlangue et transfert prosodique*

Considérant la théorie de l'interlangue, Calbris & Montredon (1975) ont montré que le rythme et l'intonation de la langue maternelle sont les éléments les plus ancrés dans

l'interlangue de l'apprenant, dans la mesure où ils relèvent du domaine psychologique. L'apprenant subit alors à un blocage inconscient au niveau prosodique.

Wenk (1985) constate que l'interlangue est caractérisée par des unités rythmiques qui empruntent les caractéristiques à la fois de la langue maternelle et de la langue cible.

Plus récemment, Jun (2005:4) reconnaît l'existence d'un transfert prosodique, selon lequel les traits prosodiques de la langue maternelle affectent la production de la prosodie de la langue cible :

« [...] by comparing the intonational categories and their realizations in the target languages, they can pin down the sources of prosodic interference and transfer. »

L'étude récente de Trofimovitch & al. (2001) analyse les effets de l'âge d'exposition à la langue cible, ainsi que le degré d'expérience sur les caractéristiques prosodiques des productions des apprenants. Les auteurs montrent que le facteur de l'âge influe sur la vitesse d'élocution, la fréquence et la durée des pauses, alors que le degré d'exposition aurait un effet positif sur les traits rythmiques.

Mennen (2006:4) considère qu'un transfert prosodique se présente sous la forme d'une interférence, à la fois phonologique et phonétique. Elle explique qu'une influence phonologique relève des différences d'inventaire de tons phonologiques, de leur forme et de la signification qui leur est attribuée, par opposition à une influence phonétique qui s'opère par une différence de réalisation phonétique d'un ton phonologique identique.

Enfin, Rasier & Hiligsmann (2007) étudient le rôle de la L1 dans l'acquisition de la prosodie en L2, et tout particulièrement dans l'acquisition des « pitch accents » par des apprenants néerlandais et français. Les auteurs se basent sur le modèle intégré contrastif « Integrated Contrastive Model » (ICM) qui consiste en une analyse contrastive des productions comparatives des apprenants L2 et des natifs. S'appuyant sur divers travaux antérieurs, ils postulent que les apprenants L2 ont beaucoup de difficultés à distinguer entre l'information nouvelle et ancienne présentes dans la phrase. De ce fait, les apprenants ont tendance à accentuer voire emphatiser chaque mot de l'énoncé. De manière générale, selon eux, l'accent est le phénomène prosodique le plus difficile à acquérir en L2. Les résultats, obtenus à partir d'enregistrements de phrases nominales de type

-déterminant+nom+adjectif-, montrent des différences de placement d'accentuation assez importantes et que les contrastes entre le français et le néerlandais influencent le processus d'acquisition. Par ailleurs, les auteurs ont noté une correspondance entre les erreurs d'accentuation et les erreurs de pauses et suggèrent alors que les erreurs prosodiques sont cumulatives.

Ces récents travaux ouvrent la voix à de nouvelles perspectives et ont une importance capitale pour le statut de la prosodie dans l'acquisition des langues secondes. Malheureusement trop peu nombreuses, ces études ont besoin d'être approfondies et développées, ce qui nous permettrait de mieux cerner les conséquences de cet « accent étranger » sur la production de nos apprenants L2 et d'en tirer profit pour une meilleure pédagogie de l'oral. La partie suivante expose la question de « surdité » de l'apprenant, au travers de trois approches électroniques qui ont été créées dans le but de remédier aux problèmes de perception de l'apprenant en langue seconde.

Approches électroniques non conventionnelles : la perception au cœur du problème

Lorsque l'on apprend une seconde langue à l'âge adulte, la maîtrise totale, notamment des phénomènes prosodiques de la langue cible, est difficile. Cela, d'ailleurs, nous amène souvent à parler d'accent étranger, ce que nous avons vu précédemment. Une hypothèse possible à cela est que nous avons du mal à articuler des sons que nous n'avons pas pour habitude de prononcer. Une autre cause possible, exposée ci-dessus, est que l'apprenant adulte a de grandes difficultés à percevoir les sons de la L2 et donc à les articuler correctement (en effet, cette absence de perception ne permet pas à l'apprenant d'entendre ce qu'il produit réellement et qui n'est que le calque de la L1).

Pour répondre à cette surdité, diverses méthodes, donnant naissance à de petits appareils électroniques, replacent l'oral au centre de différentes approches pédagogiques. Le travail de la prosodie fait partie intégrante de la rééducation de l'oreille. Ainsi, des méthodes audio-visuelles se développent afin de pallier les difficultés inhérentes pour la perception et la production d'énoncés.

▪ *SUVAGs : Système Universel-Verbo-Auditif-Guberina*

Les premières générations, au début du XIXème siècle, se contentent d'ajouter une image à une méthodologie audio-orale. C'est à partir des années 60 que l'on peut considérer l'apparition d'une deuxième génération de méthodes qui doivent leur existence à la méthodologie structuro-globale.

En effet, en 1955 Guberina développe sa théorie verbo-tonale et en 1961 a lieu la commercialisation des premiers SUVAGs (Système Universel-Verbo-Auditif-Guberina), appareils électro-acoustiques (cf. figure 3). Cette méthode se démarque du structuralisme classique, dans la mesure où ce n'est pas la structure de la langue qui est considérée, mais plutôt la structure de la parole. Cette méthode est le fruit des travaux de deux équipes de recherches, celle de Guberina à l'université de Zagreb et celle de Rivenc au CREDIF (Centre de Recherche et d'Etudes pour la Diffusion du Français). Se basant sur la théorie de Saussure selon laquelle le rôle de la langue est d'être un instrument de communication, les équipes de recherche vont alors utiliser l'audio-visuel pour rétablir cette fonction et redonner priorité à la communication parlée par le biais de l'image.

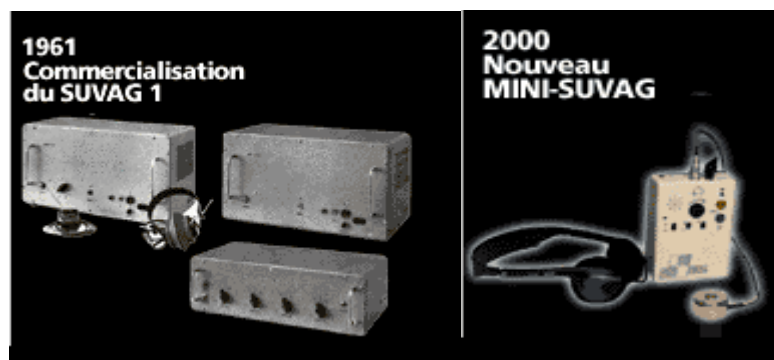


Figure 3. Illustration des appareils SUVAGs en 1961 puis 2000¹⁵

¹⁵ <http://www.suvag.com/fr/histoire/chronologie.html>

Les premiers travaux de thèse de Gubérina (1939) mettaient déjà en avant l'importance du rythme et de l'intonation dans l'acquisition des langues. Les principes du système verbo-tonal soulignent l'importance du rythme et de l'intonation comme facteurs optimaux de la structuration des langues, et donc de leur acquisition. L'objectif est de transmettre le son de telle manière que l'apprenant le perçoive dans son « champ optimal auditif ». Gubérina (1974) explique que dans la bande totale de l'audition humainement perceptible (16 à 16 000 Hertz), il s'agit des « plages » dans lesquelles existent des « restes auditifs ». Ce champ optimal, qui correspond au recensement précis de ces zones d'audition, permet de différencier plus facilement les sons de la parole. Gubérina applique ses recherches en pathologie ainsi qu'à l'enseignement des langues étrangères et crée les SUVAGs LINGUA. L'auteur base alors son approche sur le fait que l'apprenant est un « sourd phonologique » et qu'il est incapable de percevoir des sons qu'il ne possède pas dans son champ auditif restreint à celui de sa langue maternelle ; le procédé consiste alors à modifier le signal acoustique, pour mettre en valeur les zones de fréquences considérées comme optimales dans la langue cible. Les SUVAGs donnent la possibilité de travailler à l'élargissement du champ optimal afin que l'apprenant puisse différencier les sons de la parole dans un champ plus étendu que l'initial. Gubérina utilise donc ses SUVAGs LINGUA dans les premières étapes de l'enseignement pour mettre en relief l'intonation et le rythme ; plus précisément, c'est la deuxième étape de correction sur les 7 envisagées qui porte sur le rythme et l'intonation. C'est la visualisation, et non la traduction, de la situation par l'image en s'appuyant sur des aspects non verbaux tels que l'intonation, le rythme, l'intensité, les gestes, qui est mise en avant. La compréhension se fait alors par l'audio-visuel. L'apprentissage des structures de la langue orale s'effectue dans un contexte « global », qui insère la démarche dans l'ensemble des situations réelles où se trouvent engagés les sujets parlants. Cette approche n'a cessé d'être développée par son auteur au fil des années. Cependant, Konopczynski (2001) fait remarquer que la notion de fréquence optimale d'une langue suscite de sérieuses réserves parmi la communauté scientifique.

■ « *L'oreille électronique* »

Ce premier travail d'audiophonologie a inspiré par la suite la création d'autres appareils. Les recherches de Tomatis (1971) lient le handicap linguistique à l'évolution de la maturation de l'oreille.

Tomatis, oto-rhino-laryngologiste, observe que chaque langue utilise de façon préférentielle certaines plages de fréquences sonores appelées « bandes passantes ».

Selon Tomatis, les différences de fréquences sont liées à la différence de l'impédance de l'air en fonction du lieu, c'est-à-dire de l'altitude, de la végétation, de l'humidité etc. En fonction de ces caractéristiques, l'air propagerait mieux certaines fréquences et en atténuerait d'autres. La méthode Tomatis se fonde sur les relations existant entre l'oreille et la voix, par extension, l'écoute et la communication.

« Si l'on restitue à l'oreille la possibilité de percevoir correctement les harmoniques perdus ou compromis, ceux-ci sont instantanément restitués dans l'émission vocale ou instrumentale ». (Tomatis, 1978:45)

Pour aider l'intégration d'une langue étrangère, il convient, selon Tomatis, de donner à l'oreille la capacité d'entendre les fréquences propres à la langue en question et, aussi, de suivre le tempo des temps de latence caractéristique de cette langue ; le sujet deviendrait alors capable d'en reproduire les sons, le rythme et l'intonation avec une plus grande facilité. A partir de ses observations, Tomatis crée un appareil appelé « oreille électronique » qui faciliterait l'apprentissage des langues étrangères. La mise en application s'effectue de la manière suivante : étant donné que le système neuromusculaire de l'apprenant est déjà formé pour l'audition de sa langue maternelle, l'oreille électronique va permettre d'entraîner le sujet à entendre à la manière de la population linguistique de la langue cible, et à lui faire adopter la posture adéquate de la langue cible. Ce processus se réalise par un filtrage, en renforçant la bande passante et en introduisant les temps de précession et de latence ainsi que la latéralité auditive de la langue visée. Cela permet de stimuler l'écoute, d'enlever les blocages, d'intégrer plus rapidement la langue cible en réorganisant les mécanismes de communication. Progressivement, il se produit un assouplissement audio-oral. Si cette méthode est intéressante, à notre connaissance, aucune étude objective n'a évalué l'efficacité de cette approche.

▪ *I-LANG : approche neurolinguistique*

Suite à des recherches et des travaux en neurolinguistique sur le fonctionnement de l'activité cérébrale, Alliaume (1989) propose une nouvelle méthode utilisant l'activité du cerveau dans l'acquisition prosodique d'une seconde langue. Il crée un appareil appelé I-LANG afin de faciliter l'apprentissage du système prosodique de la langue cible. La

méthode consiste à envoyer des fréquences basses directement à l'hémisphère droit (dimension prosodique musicale et émotionnelle), alors que le gauche reçoit les énoncés, la dimension linguistique du message. L'auteur explique que c'est le cerveau qui commande simultanément les organes de la parole et qu'il est important de s'adresser directement à lui en le stimulant. Alliaume distingue deux dimensions : l'expression de la langue et l'expressivité du sujet parlant qui sont inséparables et complémentaires ; chacune, comme le montre le schéma ci-dessous (Alliaume & Hesling, 1998 cité dans Aliaume 2008), est gérée par l'hémisphère droit ou gauche :

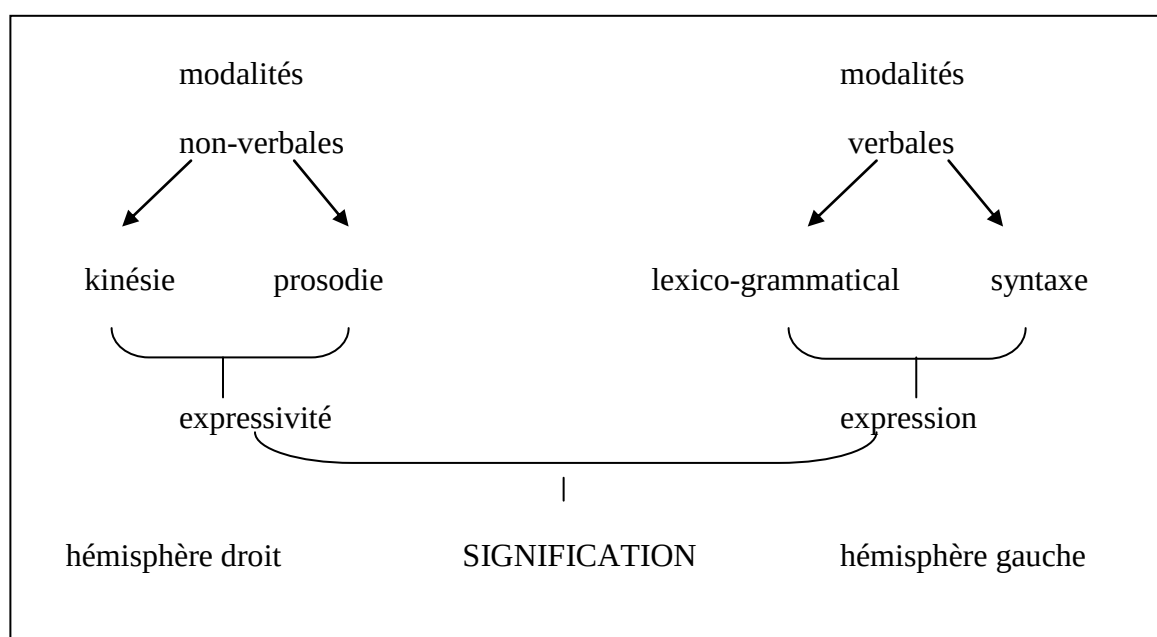


Figure 4. Illustration des modalités de la communication et fonctionnement cérébral.

Dans ses travaux, Alliaume explique que chacun des hémisphères joue donc un rôle bien précis, l'hémisphère droit traite des modalités non verbales, quand l'hémisphère gauche gère les modalités verbales, et l'un sans l'autre ne pourrait rendre compte de l'acte langagier complet. Alliaume le résume de cette façon (2008:15-16):

« -HG est plus impliqué dans les traitements des sons de parole et des intonations linguistiques (accents, accentuations spécifiques de la langue)

-HD est plus spécialisé dans le traitement des intonations affectives qui expriment par la parole l'état émotionnel du sujet. [...]

-la parole de l'hémisphère gauche isolé (HG), purement mécanique, perdrait son pouvoir de communication humaine qui est la fonction même de la parole.

-La parole de l'hémisphère droit isolé (HD) serait une modulation sans langage exprimant les émotions à la manière d'un animal familier. »

A partir de ces constatations, Alliaume propose ainsi un nouveau système, basé sur la prise en compte du fonctionnement cérébral, dans lequel il utilise par l'intermédiaire de I-LANG la spécificité de l'hémisphère droit, qui traite de façon privilégiée les fréquences basses du spectre acoustique, sans négliger les aspects verbaux. Il s'agit donc d'adresser à l'hémisphère droit la dimension prosodique (musicale) de la langue cible, pendant que l'hémisphère gauche reçoit simultanément la dimension linguistique. Par cette méthode, la perception des traits prosodiques (rythme et intonation) de la L2 est placée au cœur de l'apprentissage. Alliaume semble offrir la possibilité aux enseignants de travailler la composante la plus difficile à apprendre de la L2 qu'est la prosodie, par l'usage de l'appareil I-LANG, petit et simple d'utilisation.

Hesling (2000) expérimente I-LANG et précise qu'une hyperactivité de l'hémisphère cérébral, exercée artificiellement par l'appareil I-LANG, permettrait aux apprenants de progresser dans leur apprentissage.

Les recherches en neurosciences pourraient donc constituer un apport considérable à l'enseignement de la prosodie d'une L2, et à l'apprentissage des langues étrangères en général.

De ces trois méthodes se dégage un trait commun, se basant sur l'idée que l'apprenant d'une langue seconde ne pourra pas reproduire ou communiquer correctement ce qu'il ne perçoit pas correctement. Les trois techniques exposées, SUVAGLINGUA, l'oreille électronique et I-LANG, utilisent des appareils destinés à déconditionner l'apprenant de ses habitudes auditives et articulatoires acquises avec sa langue maternelle, et à le reconditionner au système de la langue visée. L'enseignement de l'anglais, s'il veut aider les apprenants à surmonter les difficultés inhérentes à la perception et à la production d'énoncés, doit considérer le niveau segmental (ce qu'il fait déjà) mais aussi le niveau prosodique et favoriser la place de l'écoute et le travail de perception.

Cependant, ces appareils ne parviennent pas à être intégrés aux contextes institutionnels. En effet, ces méthodes sont apparues il y a maintenant plus de 35 ans. Nous pourrions évoquer quelques raisons à cela. Tout d'abord, le manque d'intérêt pour la prosodie de la plupart des enseignants, qui privilégient l'apprentissage des éléments segmentaux, mais aussi le rôle linguistique, ignoré par manque de formation ou une formation incomplète. On peut évoquer la relative technicité à l'emploi de ces appareils ainsi que le coût de l'investissement technique. Mais il faut également mentionner le grand scepticisme que l'efficacité de ces techniques peut poser. En effet, ces méthodes n'ont pas fait l'objet d'évaluation, il est donc difficile de se baser sur de telles techniques, sans preuves objectives de leur efficacité. Il pourrait donc être intéressant de les tester sur plusieurs mois, et de les compléter, en incluant un système d'évaluation de la prosodie, de manière à « mesurer » la progression de l'apprenant et de lui fournir une évaluation objective de ses productions, ce qui n'existe pas à l'heure actuelle.

Synthèse

Dans ce chapitre, nous avons rendu compte de la place essentielle de la prosodie autant, dans l'acquisition de la langue maternelle, que dans l'apprentissage d'une langue seconde.

Nous avons présenté différents concepts qui donnaient au rythme « la casquette » (Astésano, 2001) d'horloge interne, lui attribuant une origine biologique dans le langage. Nous avons décrit les différentes phases d'acquisition de la prosodie de la langue maternelle, en mettant en évidence le rôle prépondérant du rythme dans la formation du langage.

Cela nous a permis d'observer, dans une troisième section, que la prosodie était un facteur essentiel dans la production des apprenants L2 ; cependant, elle n'était pas réellement placée au cœur des enseignements dispensés, car souvent délaissée pour un enseignement de correction articulatoire.

Il a par ailleurs été démontré que le système de la langue maternelle, mis en place dès le babillage, influence les apprentissages futurs de la L2, ce qui a permis de rendre compte d'un certain nombre de difficultés « inconscientes » chez les apprenants d'une L2. L'apprentissage de l'oral d'une langue seconde s'effectue par la perception auditive que le « crible phonologique » semble rendre sélectif. Dans cette optique, nous avons présenté quelques méthodes pédagogiques tentant d'améliorer la qualité de la prosodie des productions L2.

La littérature sur l'acquisition des langues L1 & L2 nous permet de constater le hiatus gigantesque entre les besoins des didacticiens et l'incapacité de la prosodologie à satisfaire leurs requêtes, malgré ses avancées immenses. Pourtant, la prise en compte de l'oralité en L2 passe nécessairement par celle de la prosodie.

Les approches de Guberina, Tomatis, ou encore Alliaume pourraient être envisagées, dans la mesure où des études objectives d'évaluation de ces méthodes prouvent leur efficacité.

CHAPITRE 3.

PROSODIE ET MULTIMEDIA :

PROBLEMES D'EVALUATION

SILICON AT WORK: COMPUTER PROCESSING OF HUMAN LANGUAGE



from Fromkin & Rodman (1998)

Dans ce chapitre, nous faisons état des différents logiciels et outils d'apprentissage les plus pertinents de la prosodie assistée par ordinateur ; les différents atouts et inconvénients qu'ils génèrent sont discutés. La notion d'évaluation est également abordée, ainsi que les récentes études tentant de proposer une évaluation prosodique objective.

3.1 Les logiciels d'apprentissage de la prosodie de l'anglais

3.1.1 Introduction

Dans le domaine musical, le rythme a une place privilégiée dans l'apprentissage de la musique. Il a été remarqué que le processus d'éveil chez les tous jeunes enfants les conduit, dès qu'ils sont en mesure de saisir des objets, à les entrechoquer de façon répétitive. Dans sa méthode d'apprentissage musical, Martenot¹⁶ (Laurendeau, 1990) affirme que « c'est par le rythme, sous sa forme la plus directe, la plus rudimentaire, la plus instinctive, que doit commencer l'éducation musicale ». Ne pourrait-on pas reproduire le même schéma d'apprentissage en langue seconde ? Pourquoi ne pas apprendre une langue seconde en passant d'abord par la prosodie et plus particulièrement par le rythme de celle-ci ? Nous avons vu que l'acquisition d'une langue maternelle passe par les éléments prosodiques, avant les éléments segmentaux. Ce sont donc l'intonation et le rythme qui priment et qui sont les premiers éléments acquis. Comme dit précédemment, avec l'âge, la langue maternelle entraîne des interférences auditives. Il devient donc de plus en plus difficile d'acquérir les éléments prosodiques de la langue cible. Par ailleurs, lorsque nous écoutons une langue qui nous est tout à fait étrangère, la première chose qui nous marque est sa musique ; parallèlement à cela, il nous est impossible de segmenter le flux des mots ou d'en comprendre le sens. L'enseignement d'une seconde langue par le biais d'un outil multimédia serait une solution intéressante. De fait, les technologies multimédia pourraient jouer un rôle de rééducation prosodique, ou être un complément à l'enseignement de la prosodie, en rendant observables et accessibles les déviations produites par l'apprenant.

¹⁶ Martenot est un ingénieur et musicien français. Il est le créateur de l'instrument de musique électronique à clavier appelé *ondes Martenot* (1928).

3.1.2 Constat

Nous avons cherché à savoir s’il existait des logiciels sur l’apprentissage du rythme uniquement. Dans cette recherche, nous constatons en premier lieu, que parmi les logiciels de langues, la tendance majeure est de se focaliser sur l’apprentissage de la grammaire, la conjugaison, la syntaxe ou encore du vocabulaire, alors que la prosodie est restée bien souvent « hors jeu ». Cela peut s’expliquer certainement par le fait que, dès le début de la création de logiciels, ce sont les « matières scientifiques » telles que les mathématiques, la physique ou la chimie qui étaient concernées, car en tant que sciences exactes, il était relativement pratique et facile de les transposer à cet outil. De ce point de vue là, il n’est donc finalement pas étonnant si, concernant le domaine des langues, les premières applications concernaient les composantes les plus aisées à systématiser et automatiser. Dès lors, il a fallu attendre la fin des années 90 pour voir la naissance de logiciels d’apprentissage de l’anglais par la prosodie incluant le rythme.

Le deuxième constat, face à cette recherche, concerne la place du rythme dans ces logiciels. Il n’existe pas de logiciels consacrés uniquement au rythme de la langue cible ; si rythme il doit y avoir, il va toujours de pair avec l’intonation (ce qui n’est pas étonnant, les deux éléments étant complémentaires d’un tout formant la prosodie). En revanche, il est aisé de trouver des logiciels destinés uniquement à l’apprentissage de l’intonation. La section suivante propose un bref historique de l’évolution des visualiseurs des courbes intonatives, qui ont donné naissance aux nombreux logiciels multimédia actuels.

Historique sommaire de l’évolution de la visualisation de l’intonation

- *Premières tentatives*

L’apprentissage des éléments prosodiques a été pendant longtemps négligé au profit des éléments segmentaux. C’est à partir des années 60 que certains travaux ont donné lieu à de nouvelles technologies basées sur le travail de l’intonation. La démarche réside en la représentation graphique par une visualisation de l’onde sonore sous la forme d’oscillogrammes voire parfois de spectrogrammes.

Ainsi, Vardanian (1964) et Lane & Buitén (1965) développent des méthodes qui consistent à visualiser la courbe intonative lors de répétition de phrases données. Des expériences ont été menées pour chacun des visualiseurs mais aucune n’a montré de

résultats pertinents dans les progrès réalisés par les apprenants. En effet, soit ces méthodes ne comprenaient pas de feedback visuel adapté (utilisation de l'oscillogramme et des courbes d'amplitude ininterprétables), soit elles n'étaient pas suffisamment exploitées pour être efficaces.

- *Visualisation de la courbe mélodique*

C'est entre les années 70 et 80 que les progrès techniques et les recherches en sciences du langage sur l'acoustique, en parole de synthèse, et en reconnaissance de parole ouvrent la voie vers des technologies permettant la visualisation de la courbe mélodique par des types de représentations beaucoup plus encourageants. Les recherches de Léon & Martin (1971) ont abouti à la mise au point d'un analyseur, montrant sur un écran d'ordinateur la phrase modèle de l'enseignant et la reproduction de l'apprenant. L'idée de logiciels permettant de visualiser les courbes intonatives est alors approfondie et renforcée par l'apparition d'un feedback visuel. L'apprenant peut alors comparer sa production avec celle du modèle et repérer les différences majeures de production, en particulier par la durée et la forme de la courbe mélodique. Le principe du feedback visuel prend alors toute son importance.

Cette avancée technologique va permettre également à James (1972, 1977) de mener des expériences confirmant qu'un feedback visuel d'apprenants anglophones représente un élément important dans l'acquisition des schémas intonatifs de la langue cible, ici le français. Plus particulièrement, il remarque qu'une visualisation comparée est plus efficace qu'une simple visualisation de la courbe modèle. De Bot (1983) montre dans une étude qu'une meilleure imitation du modèle était plus rapidement atteinte lorsqu'il y avait un feedback visuel et qu'un facteur de motivation entrait en jeu.

Malgré les progrès apportés par l'intégration d'un visualiseur de mélodie, le feedback semble être insuffisant et ne permet pas de différencier les variations prosodiques pertinentes des variantes individuelles possibles.

- *Nouvelles technologies : phonétique et multimédia*

Une décennie plus tard, les techniques d'analyse du signal et d'extraction de la fréquence fondamentale ainsi que de reconnaissance de la parole deviennent plus fiables et

plus accessibles. Divers logiciels sont développés. Leur utilisation première est plutôt destinée aux chercheurs ou enseignants-chercheurs, mais il nous semble qu'une utilisation détournée de ces produits est tout à fait envisageable dans l'apprentissage d'une L2.

Parmi eux, PRAAT, développé par Boersma & Weenick (2001) est un logiciel en libre accès¹⁷, en constante évolution, qui présente plusieurs fonctions telles que l'analyse (Fo, formants, courbe d'intensité), la manipulation de la parole, la synthétisation. Ce programme informatique, d'une apparente complexité, pourrait s'intégrer dans un cursus de formation pour des étudiants de niveau avancé (masters), pour qui le logiciel PRAAT constituerait un outil intéressant pour mener une auto-évaluation prosodique et/ou segmentale de leur propre production. En effet, PRAAT permet de s'enregistrer, ce qui permettrait aux étudiants de visualiser leurs énoncés et de les analyser au travers des différentes fonctions proposées par le logiciel (courbes, durées, etc.).

Un second logiciel d'analyse, nommé Speech Analyzer, proposé par le SIL (Summer Institute of Linguistics) se voit développé par l'ajout d'extensions telles que SaRP¹⁸ (Speech Analyzer Rapid Plot) créées par Nikolov & Dommergues (2007).

SaRP est une extension de Speech Analyzer, version 3. Il permet de manipuler la représentation graphique de données acoustiques, en créant un diagramme des voyelles¹⁹ (propre à chaque production). Plusieurs fonctions sont à disposition des utilisateurs, telles que :

- l'étude des propriétés acoustico-phonétiques des voyelles, par élaboration d'un diagramme avec représentation graphique des valeurs moyennes,
- la possibilité de mener des analyses comparatives des voyelles par superposition graphique,
- la possibilité d'utiliser le logiciel comme un outil d'aide à la compréhension orale.

¹⁷ www.praat.org

¹⁸ <http://web.uni-plovdiv.bg/rousni/sarp/>

¹⁹ SaRP permet également une visualisation de la f0 et du spectrogramme, comme représenté sur la figure 5.

De la même manière que PRAAT peut être utilisé par des « non experts » (suite à une formation), SaRP peut aussi constituer un outil intéressant pour les apprenants. Après une formation pour une prise en main du logiciel, l'apprenant pourrait tout à fait auto-évaluer les voyelles de ses productions, d'autant qu'à partir de la deuxième année universitaire, les apprenants maîtrisent cette charte des voyelles. SaRP offre une application pédagogique intéressante qui touche à la fois, phonéticiens, didacticiens et apprenants.

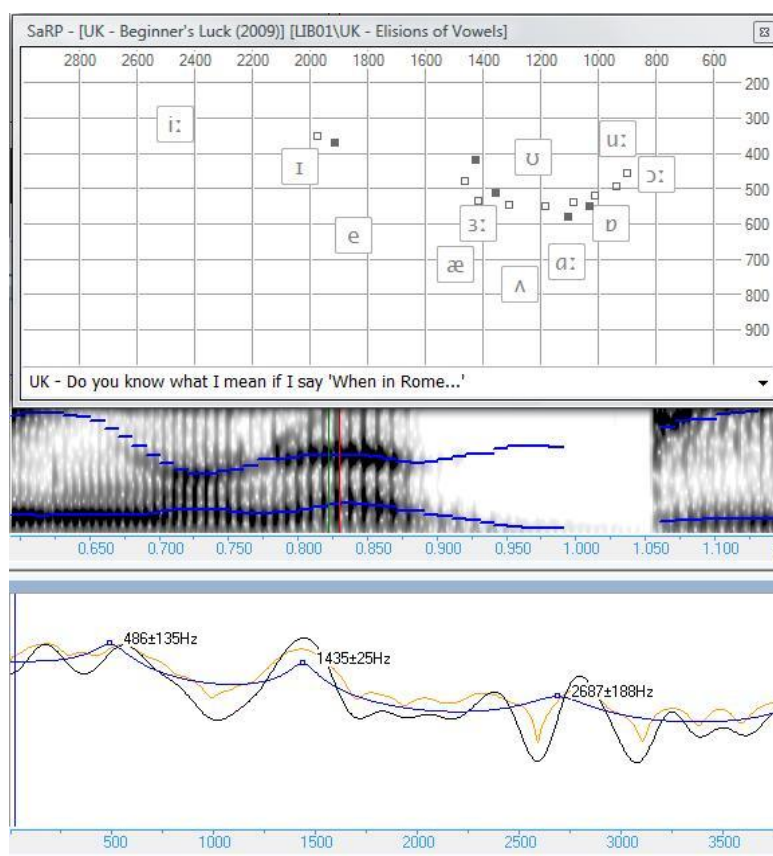


Figure 5. Exemple de représentation graphique proposée dans le logiciel SaRP

L'utilisation première des logiciels, présentés ci-dessus, n'est pas directement destinée à l'apprentissage de la prosodie d'une L2, mais il est tout à fait envisageable de détourner leurs fonctions premières dans une visée pédagogique et formatrice pour l'apprenant. Le seul bémol est bien entendu, comme dit précédemment, le besoin que cela soit effectué au sein d'une formation, permettant une prise en main de ces logiciels.

C'est ainsi qu'à partir des années 1990, on peut constater l'émergence de nombreux logiciels, qu'ils soient outils d'analyse (comme vu précédemment) ou bien encore, logiciels d'apprentissage multimédia (cf.ci-dessous). Cependant, beaucoup de ces outils ne traitent encore que des aspects grammaticaux et lexicaux de la langue, ou bien ciblent uniquement l'apprentissage des éléments segmentaux.

En illustration de ces logiciels, il était possible de trouver « **Escape from Arizona planet** » (1997) présenté par UC Berkeley²⁰, qui était un programme informatique proposé par "EF multimédia". Ce logiciel, destiné aux apprenants d'un niveau avancé, proposait un apprentissage interactif au travers de diverses mises en situation. L'utilisateur pouvait travailler la compréhension, étendre son vocabulaire, pratiquer la lecture, et tester sa prononciation en écoutant et répétant des phrases extraites des mises en situation. Pour résumer, l'apprenant pouvait expérimenter une situation de communication, comme il était mis en exemple dans la présentation du logiciel : « *You are an alien whose spaceship has crashed outside of a small town in the middle of Arizona's desert. Your mission is to find the seven objects necessary to repair your ship [...]* ».

WebGrader, proposé par Neumeyer & al. (1998), est un outil plus récent basé sur la reconnaissance de parole, conçu pour la pratique de la prononciation d'une seconde langue, au travers de différentes leçons regroupées par thème, telles que « repas au restaurant », ou encore « les transports en commun ». Pour chaque leçon, un ensemble de phrases est proposé ; l'utilisateur peut écouter les phrases prononcées par un natif et peut s'enregistrer, afin d'obtenir un score sur ses réalisations.

PLASER, présenté par Mak & al. (2003) est un logiciel destiné à l'enseignement de la prononciation de l'anglais pour des apprenants chinois. Utilisant la reconnaissance de parole, cet outil donne un feedback immédiat aux utilisateurs qui pratiquent deux types d'exercices, à savoir la prononciation, soit par le biais de paires minimales, soit par l'utilisation de mots.

²⁰ <http://www-writing.berkeley.edu/chorus/call/reviews/archives/escape.html>

Suivant l'idée d'aider l'apprenant à analyser la prononciation des mots, au travers de la transcription phonémique, des logiciels destinés à l'entraînement de la transcription phonémique voient le jour. **Phthong**²¹ (1999), par exemple, offre une initiation à la phonétique anglaise ; ce programme est un tutorial réservé à la transcription phonémique. Il est composé de deux types d'exercices : soit l'utilisateur transcrit un mot anglais, en phonèmes à l'aide de l'alphabet phonétique, soit il analyse un mot transcrit en phonèmes pour retrouver sa forme orthographique.

Si ces exemples de logiciels assignés à l'apprentissage des phonèmes et de la prononciation (bien loin d'être exhaustifs), démontrent l'engouement pour les aspects segmentaux d'une langue, ces dernières années ont vu émerger un bon nombre de logiciels destinés à l'enseignement de l'intonation. Cependant, le lancement d'un nouvel outil multimédia d'apprentissage des langues n'est pas forcément toujours le reflet d'une nouveauté méthodologique, mais plutôt l'aboutissement d'une opération commerciale chargée de rendre l'outil plus attractif. Ces outils ne semblent pas s'associer à une véritable pédagogie de l'oral et les possibilités offertes sont encore limitées et peu accessibles pour les apprenants.

Parmi les logiciels existants sur le marché, deux types de regroupements (listes non exhaustives) peuvent être établis.

Le premier groupe concerne les logiciels proposant un feedback, basé sur la visualisation d'un spectrogramme ou oscillogramme (illustré par la figure 6) :

-**SoundForge** (1991)-Sonicfoundry& RealNetworks: <http://www.Sonicfoundry.com>

-**Tell me more** (1998)-Auralog: <http://www.auralog.fr>

-**VOICEbook** (1998)-Englishear system (Paris): <http://www.voicebook.com>

-**LAVAC** (Laboratoire Vidéo Actif Comparatif) (1999)/ Tony Toma : <http://www.alizes.fr/cp3i>.

²¹ www.chass.utoronto.ca/~stairs/phthong/

-**Speaker** (version 4.0-1999)-Neuroconcept: <http://www.neuroconcept.com>

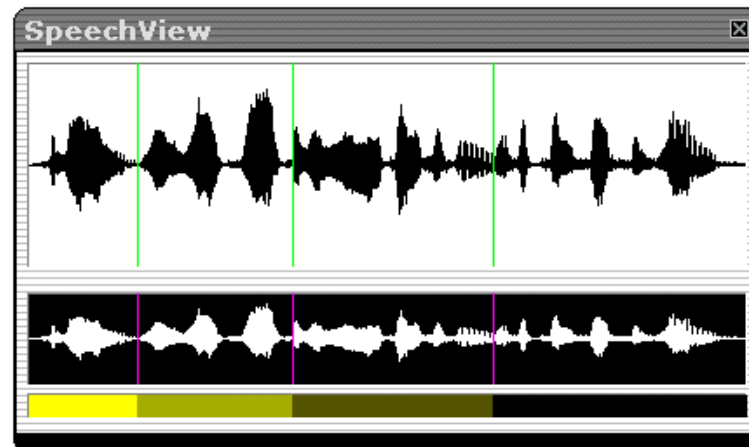


Figure 6. Proposition de feedback par affichage de l'oscillogramme (Voicebook, 1998)

Le deuxième groupe proposé, utilise la courbe d'amplitude ou encore la courbe de fréquence fondamentale non lissée afin de représenter les courbes intonatives des productions.

-**Speaker** (version 3.0- 1997)-Neuroconcept (cf.above)

-**Wincecil** (1997)-SIL: <http://www.jaars.org/itcs/software/cecil/wincecil/we22.zip>

-**Talk to me** (1998)-Auralog: <http://www.auralog.fr>

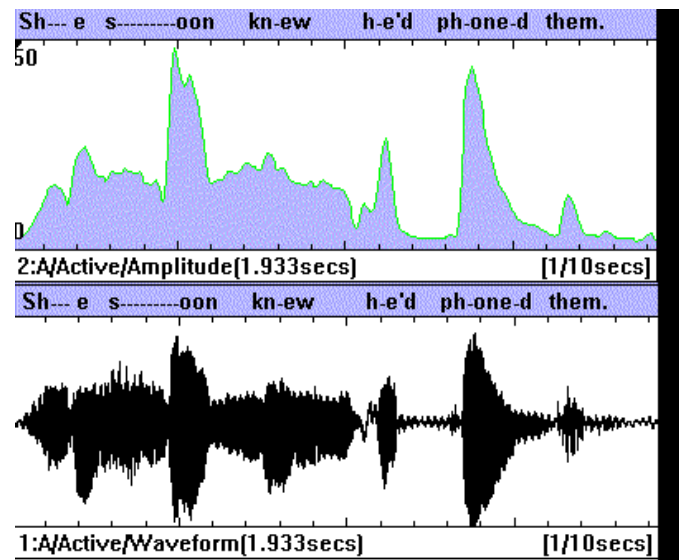


Figure 7. Proposition de feedback par affichage de la courbe d'amplitude (Wincecil, 1997)

La courbe d'amplitude proposée par Wincecil (figure 7) ne semble pas plus convaincante que l'oscillogramme de Voicebook (figure 6). Dans les deux cas, on peut bien imaginer que l'apprenant aura du mal à comprendre et décoder ce qu'il faut retravailler. Sans une aide experte, il paraît difficile pour l'utilisateur d'accéder à une interprétation de sa production. Il est à noter un point positif pour l'affichage de Wincecil qui propose d'ajouter les mots du texte au-dessus de la courbe d'amplitude. Cela permet à l'apprenant de mieux visualiser l'élément sonore qu'il souhaite étudier. Sans connaître les correspondances entre le texte et l'oscillogramme ou toute autre forme, ce qui peut être compris assez facilement par l'utilisateur, il paraît difficile pour l'apprenant de bénéficier du meilleur parti de ces affichages pour améliorer la qualité de ses productions.

Les diverses représentations proposées jusqu'ici, quelle que soit la forme utilisée (spectrogrammes, oscillogrammes, etc.), semblent difficilement accessibles pour des non spécialistes. L'apprenant semble être considéré comme ayant fatalement les connaissances phonétiques instrumentales nécessaires à l'interprétation des courbes visuelles. L'efficacité de ces logiciels peut être ici incriminée, et laisse un doute sur l'efficacité de la représentation de telles courbes, qui ne sont pas pédagogiquement parlantes pour l'apprenant.

Malgré ces derniers points négatifs, il est constatable que, depuis ces toutes dernières années, se développent quelques rares logiciels, voire didacticiels, cherchant à faire évoluer l'outil pédagogique que peut constituer un logiciel, mettant en exergue le travail de la prosodie et prenant en compte l'accessibilité des représentations données aux apprenants. Une évolution, qui vise à dépasser les faiblesses constatées auparavant, semble donc déjà se dessiner.

Le logiciel **SLIM** (Multimedia Interactive Linguistic Software) présenté par Delmonte (1998) se base sur deux objectifs focalisés sur l'amélioration (i) de la perception (ii) de la production de l'apprenant au niveau prosodique. Deux activités principales sont proposées afin de travailler rythme et intonation. La première se situe au niveau du mot et est liée à la durée des segments syllabiques, la seconde au niveau de la phrase, travaillant sur l'accent de phrase, les patrons intonatifs.

Ce logiciel semble être un outil intéressant, dépassant l'idée traditionnelle de simple visualisation de la courbe intonative ; SLIM tend à faire de nouvelles propositions d'apprentissage. Malheureusement, aucune démonstration n'est proposée ; nous ne

pouvons commenter plus en détail ce logiciel. Mais il est à noter que Delmonte est un des rares concepteurs à mettre l'accent sur le facteur « rythme », sans favoriser outre mesure, la représentation graphique de la mélodie. Martin (2005) l'explique par le fait que des recherches en synthèse de la parole ont démontré que le rythme est parfois « plus important que la mélodie pour la compréhension de l'organisation hiérarchique (syntaxique) de la phrase ».

Selon la même conception, **BetterAccentTutor** (que nous appelons désormais BAT) développé par Kommissarchik & Kommissarchik (2000), est un logiciel d'apprentissage du système prosodique de l'anglais américain. Ce programme propose un feedback visuel immédiat sur l'intonation, l'accentuation et le rythme des productions ; les erreurs segmentales ne sont pas prises en compte, l'attention est mise sur la prosodie des productions.

Le système BAT propose deux modes d'entraînement par visualisation, (i) sur l'intonation et (ii) sur l'intensité et le rythme. L'apprenant peut visualiser sa production en la comparant avec celle du natif. Sur l'ensemble des thématiques de travail proposé, l'utilisateur a la possibilité de pratiquer des exercices sur l'accent de mots, les affirmatives, les Wh-questions, les questions de répétition, les tags, les ordres, les exclamations, « tongue twisters », les longues phrases.

Concernant le rythme, des exercices ciblés sont proposés comme par exemple, des exercices d'accentuation lexicale où la fonction distinctive de l'accent (en anglais) est mise en avant ; voici un exemple avec l'exercice « present vs. present » représenté par les 3 figures proposées ci-dessous. L'apprenant doit prononcer un des deux mots ou les deux ('present ou pre'sent) en répondant à la question « he said what ?! ».

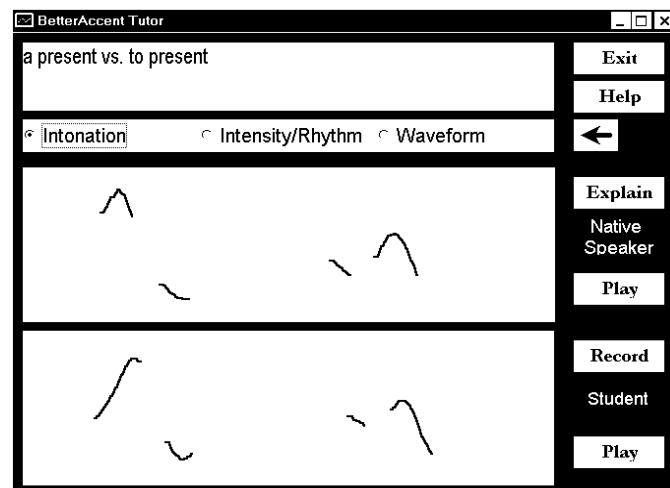


Figure 8. Feedback visuel des mots « 'present vs pre'sent » avec comparaison des productions natif/apprenant selon les contours intonatifs.

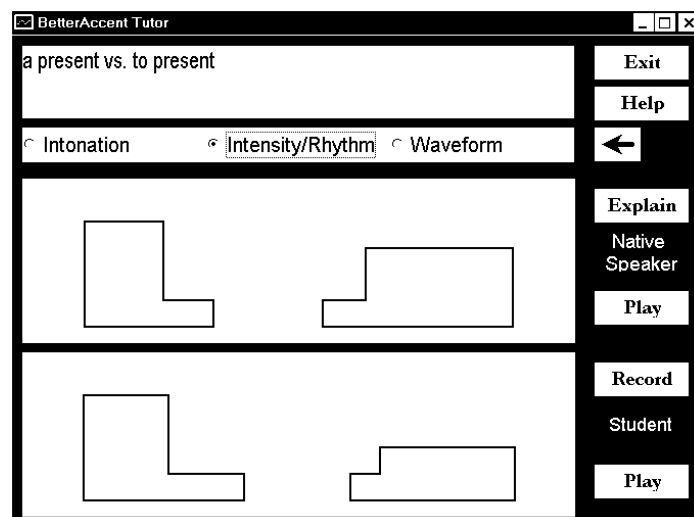


Figure 9. Feedback visuel des mots « 'present vs pre'sent » avec comparaison des productions natif/apprenant (rythme/intensité)

Les syllabes sont représentées par des marches ; la longueur de la marche représente la durée de la syllabe et sa hauteur correspond à l'énergie du nucleus de la syllabe (la voyelle).

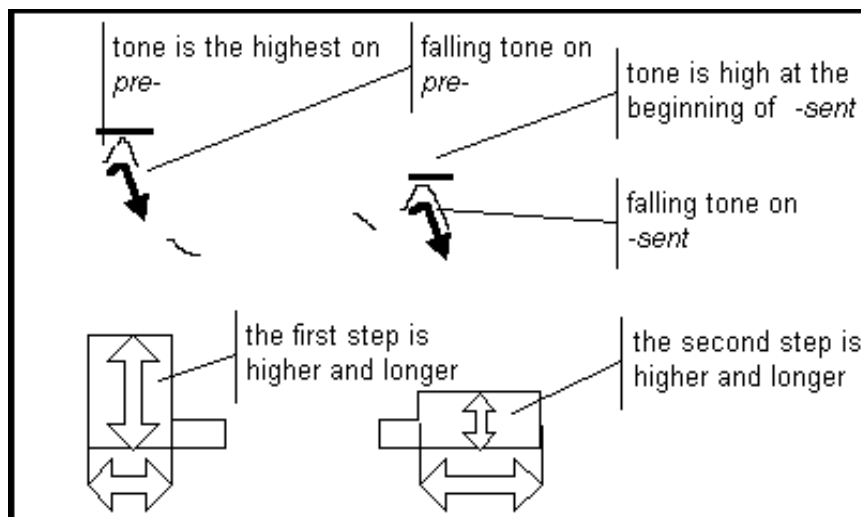


Figure 10. Illustration du feedback explicatif des points saillants à travailler par l'apprenant

La troisième étape est un module explicatif proposé à l'apprenant mettant en avant les points saillants de sa production, qui devraient se rapprocher plus de celle du natif.

Il est dommage que ce logiciel ne soit focalisé uniquement sur l'accent américain ; par ailleurs, il est à noter la recherche de nouvelles propositions concernant l'amélioration du feedback proposé, même si l'on peut s'interroger sur son efficacité face à la progression de l'utilisateur.

Le logiciel **Winpitch LTL** (Language Teaching Learning) a été récemment conçu au Canada par Martin et son équipe. Il est basé sur le fruit d'une collaboration entre un phonéticien-ingénieur et une enseignante phonéticienne spécialisée en didactique des langues (Germain & Martin, 2000). Winpitch LTL semble être un des logiciels les plus développés actuellement proposant diverses fonctions comme l'analyse en temps réel, la visualisation de la courbe f_0 , l'écoute ralentie. Son originalité réside dans l'annotation des commentaires et segments pertinents intégrés sur l'affichage. Des segments de courbes mélodiques peuvent être surlignés en couleur et du texte peut être facilement ajouté sur l'écran, pour mettre en relief des propriétés mélodiques ou rythmiques particulièrement pertinentes pour l'apprenant (cf. figure 11, ci-dessous), afin de localiser les différences entre la L1 et la L2. Cela requiert la présence d'un enseignant spécialisé, en mesure de commenter les visualisations proposées. Sans ces commentaires, Winpitch LTL retombe dans une configuration de feedback, améliorée certes, mais traditionnelle.

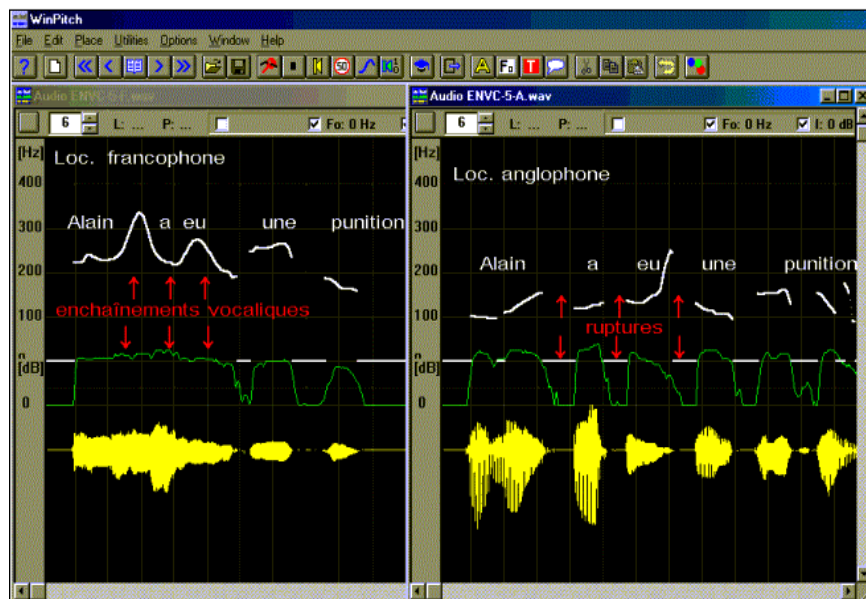


Figure 11. Exemple de feedback visuel du logiciel Winpitch LTL,

<http://www.winpitch.com>

Malgré l'attention portée sur la pédagogie proposée, la conception de Winpitch LTL, tout comme Winpitch Light et Winpitch Pro, reste celle d'un outil de recherche avant tout, et ne peut se placer en didacticiel. Notons pas ailleurs, que les logiciels Winpitch ne sont pas téléchargeables gratuitement et que Winpitch Pro est un logiciel payant.

D'autres logiciels tels que **Prosodia** (Herry, 2001) démontrent qu'il est pédagogiquement possible de créer un environnement efficace pour un travail d'apprentissage de la prosodie. La méthode utilisée s'appuie sur les propositions de schémas intonatifs présentés par Ginésy (1998). Prosodia traite de l'intonation et du rythme au travers de divers exercices de trois différents niveaux de difficulté, allant de débutants à experts. 500 phrases sont proposées, à pratiquer soit en mode répétition, soit en mode production libre, avec l'indication du modèle intonatif à suivre, par exemple :

E.g. Level A, Unit 2, Lesson 2:

Glide up (ton normal): Is it a 'doll? / is it Mary's 'doll?

Glide down (ton détaché): Is it a 'doll? / is it Mary's 'doll?

High jump (exprime la surprise): Is it a 'doll? / is it Mary's 'doll?

Concernant le rythme, ce sont des comptines qui sont proposées à l'utilisateur :

e.g. Level A, Unit 1, Lesson 1:

'One, 'two, 'three, 'four,

*'Mary at the 'cottage 'door
'Five 'six 'seven 'eight,
'Eating 'cherries off a plate.*

Le feedback proposé est commun aux autres logiciels, à savoir la visualisation de la courbe mélodique, qu'il est possible de comparer avec celle de l'apprenant ; en revanche, la courbe est bien plus modélisée, et ainsi plus accessible pour l'apprenant au niveau de l'interprétation (cf. figure 12, ci-dessous).

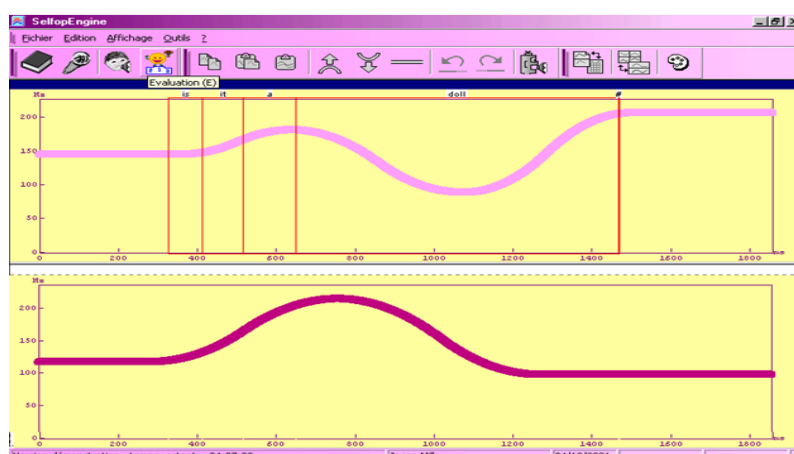


Figure 12. Exemple de feedback de la production d'un apprenant (en bas) comparée à celle d'un natif (en haut), <http://www.lpl.univ-aix.fr/~lpldev/prosodia>

Prosodia est le résultat d'un projet mené au laboratoire Parole et Langage. Il a pu être réalisé grâce à la collaboration de diverses personnes telles que Ginésy, Ghio, Nishinuma. Ce projet a été financé en partie par le ministère de l'éducation nationale, de la recherche et des technologies. Herry (2001) explique que l'objectif d'un tel logiciel est d'améliorer l'enseignement et l'apprentissage des éléments prosodiques de la L2. L'aspect segmental a aussi été pris en compte, en incluant dans les phrases un travail sur des paires minimales. Les avantages, qui différencient Prosodia des autres modèles, sont les propositions d'exercices élaborés, afin de cibler les points faibles et les difficultés particulières rencontrées par l'apprenant lors de l'apprentissage de la prosodie L2. Le logiciel est également accompagné d'un manuel explicatif « *sounds, rhythm and intonation of English* ». Nous regrettons que le contenu soit limité à 500 phrases et que le texte ne soit pas aligné avec la courbe modélisée.

Prosodia est toujours en cours de développement, faisant l'objet d'une collaboration entre le LPL et l'équipe de recherche de l'université de Paris 8.

Certains chercheurs se sont consacrés uniquement à l'apprentissage de l'accentuation. Le projet SWAMS (*Synchronised Web Authoring Notation System*), lancé en 2004, se tourne entièrement vers la réalisation du rythme, au niveau de l'accent lexical des énoncés. Le logiciel, toujours en cours de développement (Stenton & al., 2005), est le projet d'une collaboration menée au sein du LAIRDIL²² de l'université de Toulouse. SWAMS permet une annotation des textes en fonction des accents et de l'intention des auteurs. Par l'utilisation de couleurs et de la taille de la police, il est possible de mettre en relief les syllabes accentuées d'un texte sonore, tout en affichant et en synchronisant les images vidéo (cf. figure 13). En l'état actuel du système, cette fonctionnalité est exécutée de manière interactive, et permet à n'importe quel enseignant sans compétences particulières en informatique, de l'utiliser sans contrainte.



Figure 13. Illustration d'une fenêtre du logiciel SWAMS

Ces derniers logiciels montrent que des recherches sont effectivement en cours de réalisation afin de « promouvoir » un apprentissage prosodique d'une seconde langue,

²² Laboratoire Inter-Universitaire de Recherche en Didactique des Langues

tendant de pallier les différentes critiques de ces dernières années. Toivanen (2003) remarque que l'étude de l'intonation a été délaissée jusqu'à présent, malgré l'importance de son rôle dans l'apprentissage d'une seconde langue :

« The lack of interest shown in prosodic aspects of interlanguage does seem to do a disservice to the discipline of SLA²³ research as intonation is undeniably an important aspect of especially the pragmatic aspects of interlanguage speech ».

Ces propos ont été démontrés, puisque les premiers essais remontent seulement à une quarantaine d'années. Cependant, notons ici, que si la constatation est pertinente, le rythme encore une fois est négligé. En effet, jusque là, l'intonation était le seul élément central des recherches et le rythme en était largement absent. Les dernières technologies développées apportent donc un souffle nouveau sur les outils multimédia et permettent d'envisager une réelle application de ces outils dans l'enseignement de la prosodie en L2 par l'apport qu'ils pourraient constituer. Nous faisons état, dans la partie suivante, des différents avantages apportés par l'utilisation d'un outil multimédia, mais aussi ses limites.

3.2 Apports des logiciels de langues

Ces dernières années, les technologies multimédia ont subi un véritable essor par ce que l'on appelle l'Apprentissage des Langues Assisté par Ordinateur (A.L.A.O.), ou bien encore l'Enseignement Assisté par Ordinateur (E.A.O.). Si l'on peut s'interroger sur le nombre abondant de ces logiciels²⁴ masquant un éventuel intérêt commercial, il semble que ces technologies multimédia soient assez prometteuses et ce, pour différents motifs exposés ci-dessous.

Utilité dans l'apprentissage de la prosodie

Un matériel, sur support imprimé, ne semble pas adapté à un travail sur la perception et production d'une langue seconde. Les logiciels qui associent l'image au son ainsi qu'à l'animation, sont plus adaptés au travail de l'oral. Ainsi, les phénomènes

²³ Second language acquisition

²⁴ Il est même possible de trouver un logiciel d'apprentissage de la prosodie du Breton (Aubry, 2004)

inconscients tels que l'ancrage de la prosodie de la L1 rendant difficile l'apprentissage de la prosodie d'une L2, nécessitent de commencer par rendre visibles, concrets et observables les déviations prosodiques, afin que l'apprenant en prenne conscience. Dodane & Konopczynski (2001) font remarquer que c'est seulement en passant par cette étape que l'apprenant pourra corriger ses erreurs :

« En permettant la visualisation des contours d'intonation et de la structure accentuelle de la langue cible, elles donnent justement à l'apprenant la possibilité de les objectiver [...]. Cette prise de conscience constitue une étape indispensable vers la correction ».

L'utilisation de l'A.L.A.O. pourrait donc constituer un véritable apport pour lutter contre la « surdit  prosodique » de l'apprenant adulte.

L'apprenant au c ur de l'outil

L'usage des logiciels place l'utilisateur au c ur du processus : en effet, celui-ci occupe la place centrale, contrairement   un apprentissage en salle de classe ou l'enseignant garde le r le principal. Il n'est plus sujet passif, comme il pourrait l' tre dans une salle de cours surcharg e par un effectif trop important o    peine une minute de cours lui serait consacr e pour son travail de l'oral. Il est donc actif face   cette technologie, va chercher lui-m me les informations. Par ailleurs, il perd son statut d' «  l ve » et prend le r le d' « apprenant » ; dans ce cadre, l'enseignant perd son statut de « ma tre » et endosse le r le de « conseiller », ce qui change le rapport d'autorit   l ve/ma tre et redistribue un r le   chacun o  l'apprenant devient le centre d'int r t de la formation.

En pla ant l'apprenant au c ur de l'apprentissage, cela ne signifie en aucun cas le d nigrement du r le de l'enseignant. Bien au contraire, s'il y a bien une red finition des r les, c'est bien la relation apprenant/enseignant qui  volue et non l'importance du r le de l'enseignant qui est remise en cause. L'enseignant est lib r  de son r le de r p titeur et devient le guide et le cr ateur du concept p dagogique sous-jacent au logiciel ; la machine ne peut   elle seule satisfaire tous les besoins et demandes de l'utilisateur, elle devient un plus pour l'enseignant, poss dant de surcro t une patience illimit e en tant que tuteur.

Loin de remplacer l'homme par la machine, comme cela peut  tre pens , la relation enseignant/ l ve, souvent inexistante dans le contexte de cours classique, est ainsi fortement renforc e.

Par ailleurs, l'apprenant se voit attribuer une évaluation individuelle, suivant une progression individualisée adaptée au rythme de chacun. L'apprenant est donc actif, mais aussi autonome, dans la mesure où il travaille à son propre rythme et en fonction de ses besoins. Cela palie, en outre, le manque de temps qui empêche l'enseignant de donner un suivi pédagogique individualisé en raison de groupes surchargés d'étudiants. C'est ce qu'explique Marty (1983) pour qui l'enseignement assisté par ordinateur permet à l'apprenant de suivre l'état de son évolution et lui offre un programme sur mesure, adapté à son rythme d'apprentissage :

« Je n'essaierai pas de prouver que les ordinateurs peuvent faciliter l'apprentissage d'une langue. Il est évident que des étudiants désireux d'apprendre, apprendront plus vite et obtiendront de meilleurs résultats s'ils ont accès à un ordinateur qui leur fournit une réponse immédiate et une analyse détaillée de leurs erreurs, qui stocke des informations sur leurs résultats et qui, selon les résultats obtenus, leur fournit de nouveaux exercices adaptés à leur 'profil'. Je crois que le profil tiré d'exercices de ce genre vaut la peine qu'on investisse dans l'E.A.O. ».

Par l'utilisation d'un logiciel, l'apprenant se voit évalué par une machine et n'a pas ainsi l'impression qu'il reçoit un jugement porté par un enseignant. L'apprenant peut donc repérer ses erreurs et observer ses progrès, sans pour autant recevoir une note « sanction ». Cette forme d'apprentissage est donc « psychologiquement » encourageante et gratifiante pour l'utilisateur.

Murray (1999) montre, par ailleurs, dans son étude, que ce type d'apprentissage permet également à l'utilisateur de travailler dans un environnement privé, moins stressant dans la mesure où l'apprenant semble moins embarrassé de pratiquer une tâche orale et de réaliser de mauvaises productions.

Feedback pédagogique

« La visualisation de l'intonation nous a permis de transposer un geste auditif spécifique propre à un monde linguistique fermé, en un geste visuel décodable par une communauté sémiotique universelle » (Léon et Martin, 1970 : 9).

L'importance de la visualisation de la courbe mélodique a rapidement été mise en avant par plusieurs études (Léon & Martin, 1970 ; De Bot, 1983). Un certain nombre de travaux ont prouvé l'efficacité de ces visualiseurs dans l'apprentissage du système prosodique de la langue cible. Un des premiers facteurs positifs avancé par De Bot (1983) est celui de la « motivation » ; l'auteur explique que les apprenants sont beaucoup plus

motivés par ce type de feedback que par un simple feedback auditif proposé en laboratoire de langues.

Hardison (2004) reporte ses travaux sur l'apprentissage de la prosodie assisté par ordinateur, et montre au travers de deux expériences l'effet positif que ces programmes constituent dans l'application des technologies de la parole. Ces expériences montrent qu'un entraînement audio-visuel peut aider les apprenants du français à améliorer, non seulement leur prosodie, mais aussi les aspects segmentaux de leur production.

Delmonte (1999), Taniguchi & Abberton (1999) mènent également des études qui attestent de l'effet positif d'un feedback visuel. Taniguchi & Abberton (1999) montrent notamment que des apprenants japonais suite à un travail avec feedback visuel de la fréquence fondamentale de la voix améliorent leur intonation en anglais.

Herry (2001) teste l'efficacité de son logiciel Prosodia sur deux groupes d'étudiants (groupe témoin et groupe test). Le groupe témoin a suivi un entraînement traditionnel en laboratoire, l'autre groupe s'est entraîné par l'utilisation du logiciel. Deux notations ont été effectuées à 4 mois d'intervalles. Herry conclut que les résultats ne sont pas « statistiquement significatifs », mais la moyenne du groupe test reste plus élevée que celle du groupe témoin, ce qui permet de conclure à une certaine efficacité de la méthode.

Hincks (2003) étudie l'impact des programmes d'apprentissage de langue sur la prononciation d'un groupe d'immigrants, d'un groupe d'apprenants étudiant l'anglais en Suède, comparé à un groupe contrôle d'étudiants. Les groupes tests devaient travailler sur la thématique de l'anglais technique, en pratiquant 200 heures de plus que le groupe contrôle, et étaient encouragés à travailler également chez eux. Bien que les résultats soient moins probants pour les étudiants d'un niveau préalable assez bon, les résultats montrent un effet très bénéfique pour les étudiants qui ont débuté le programme avec un fort accent étranger.

Si les atouts proposés par les technologies multimédia s'annoncent encourageants, et si les points positifs développés ci-dessus montrent les implications méthodologiques et pédagogiques, l'objet multimédia présente aussi des limites, et la plateforme technique concernant le système d'évaluation ne semble pas réellement satisfaisante.

Limites de l'outil multimédia

Les produits multimédia se voient attribués divers avantages, tels ceux développés dans la section précédente, ou encore ceux annoncés par Eskenazi (1999), (i) l'apprenant entend une grande quantité de parole, (ii) il produit lui-même une grande quantité d'énoncés. Cependant, l'outil présente aussi quelques limites. Le facteur interaction est à prendre en compte dans la mesure où il permet à l'apprenant de se focaliser, non plus sur ses propres productions, mais vers celles de son interlocuteur. Cela constitue un changement de stratégie communicationnelle important. Un entraînement isolé face à un ordinateur apparaît insuffisant pour acquérir la véritable oralité d'une langue en situation authentique.

Par ailleurs, si la visualisation de la courbe mélodique semble être communément attestée dans le sens où cela faciliterait les progrès de l'apprenant et lui serait d'un apport important, il n'en reste pas moins que le feedback visuel proposé, comme nous l'avons vu précédemment, reste assez pauvre et difficilement interprétable par l'utilisateur ; Cazade (2000 :1-2) va jusqu'à remettre en cause l'utilisation des courbes sonores, dans la mesure où elles ne sont d'aucune utilité pour l'apprenant, puisqu' aucun feedback ne vient en complément :

« Ces courbes ont-elles un usage quelconque ? Ne sont-elles là que pour "faire joli", ou plutôt "faire sérieux, grâce à l'aspect scientifique, apparemment inattaquable et définitif, qu'un joli graphe semble pouvoir donner, pour un utilisateur non averti, à toute présentation de résultats dans un magazine scientifique ou non, à une présentation multimédia de courtier d'assurance ou même parfois à un exposé d'étudiant à court de démonstration ? »

Ces courbes (souvent présentées par un oscillogramme) sont trop complexes graphiquement parlant et sont fréquemment laissées à l'interprétation de l'apprenant. L'usage de l'imitation ainsi que d'un feedback visuel et sonore est loin d'être pédagogiquement satisfaisant. Une correction prosodique ainsi que certains commentaires explicatifs pourraient venir compléter la visualisation des courbes. Sous-jacent à cette constatation, est souvent évoqué le manque de coordination entre ingénieurs, phonéticiens et pédagogues (Germain & Martin, 2000 ; Dodane & Konopczynski, 2001):

« Il semble que, dans l'élaboration de tout nouveau matériel pédagogique, les phonéticiens ou les méthodologues de la prononciation se soient volontairement effacés ou bien qu'ils n'aient pas su adapter leurs recherches aux nouvelles orientations pédagogiques... » (Galazzi-Matasci et Pedoya (1989: 39), cité par Germain & Martin, 2000)

Martin (2005:104) explique que la fragilité de l'ensemble de ces logiciels destinés à l'apprentissage de la prosodie relève de considérations didactiques et pédagogiques :

« Mais le défi principal est extérieur à la visualisation : il s'agit de déterminer de manière simple et pédagogiquement convaincante quels sont les facteurs linguistiquement importants dans l'apprentissage de la prosodie d'une langue étrangère, et ensuite d'en établir les caractéristiques pertinentes par rapport à celles de la langue maternelle de l'apprenant, et de configurer les modèles et les leçons proposés en se concentrant sur ces différences. »

Certains logiciels très récents, cités précédemment (SLIM, Prosodia, Winpitch, SWAMS) ont effectivement dépassé le problème de la visualisation des courbes ; cependant, l'évaluation proposée reste à compléter. En effet, ce n'est pas l'utilisation des courbes qu'il faut remettre en cause, mais plutôt l'interprétation et l'évaluation qui l'accompagnent. Si le feedback audio-visuel a fait ses preuves et semble avoir été surmonté par certains logiciels, il manque, néanmoins, un véritable système d'évaluation reposant sur des critères objectifs permettant d'évaluer le niveau de l'apprenant. Malgré quelques premières propositions (Herry & Hirst, 2002 ; Beck *et al.* 2005), aucune évaluation basée sur des critères objectifs n'a réellement été avancée dans les recherches sur l'évaluation de la prosodie d'apprenants français. Nonobstant, lorsque cela concerne d'autres langues L1, comme le japonais par exemple, ciblant l'apprentissage de l'anglais, des études récentes attestent de l'intérêt que portent les autres nations sur le besoin d'élaboration de critères évaluatifs.

Avant de faire état des différentes études portant sur l'évaluation, nous proposons en premier lieu, dans la section suivante, de définir cette notion et d'en exposer les différents concepts et travaux qui en découlent.

3.3 L'évaluation prosodique en L2

Définitions et concepts

Après une présentation générale de la notion d'évaluation, l'attention est portée sur les évaluations prosodiques des productions orales en langue seconde et les propositions de critères évaluatifs proposés dans les récentes recherches.

Le terme « évaluer » fait référence au fait de fournir des informations pour éclairer une prise de décision (Larousse, 2003). L'évaluation est, étymologiquement, « une réflexion sur les rapports aux valeurs » (Bonniol & Vial, 1997 :16). Spécifier le type de rapports, à savoir, attribuer une valeur ou construire un jugement de valeur, consiste à lui attribuer un modèle.

La méthode d'évaluation la plus ancienne est probablement l'évaluation perceptive, que l'on peut également qualifier d'évaluation subjective. Deux types d'évaluation sont généralement distingués : (i) l'évaluation perceptive qui correspond à une production évaluée par un auditeur, et, (ii) une évaluation objective, qui correspond à des mesures instrumentales, acoustiques, etc.

Le protocole d'analyse se situe alors selon deux choix : (i) le choix qualitatif dans lequel les paramètres mis à l'étude sont définis, et, (ii) le choix quantitatif correspondant aux échelles de mesure. Selon ce concept, l'évaluation est alors une mesure, et cette notion renvoie directement à la question de mesurabilité à laquelle chercheurs et enseignants sont confrontés. La difficulté de la mesurabilité est en lien direct avec la problématique de l'évaluation des productions en L2, à savoir, qu'est-ce qu'on évalue (qu'est-ce qu'on mesure) et comment on l'évalue (comment le-mesure-t-on). La première étape consiste donc à définir l'objet, comme le précise Rost (2002 : 174) :

« In order for assessment to become valid and useful, it is essential to define the 'construct', that is, the set of traits in the learner that we are seeking to observe, evaluate and provide feedback on »

La deuxième étape consiste à attribuer une donnée chiffrée à cet objet (Bonniol, 1976) :

« Cela implique que l'objet, où l'événement, puisse être saisi sous une seule dimension, isolable, susceptible d'être pourvu d'une échelle numérique. »

Partant de ce principe, l'évaluation des productions orales devrait reposer sur une grille définie selon certains critères évaluatifs représentant les compétences à atteindre pour l'apprenant, critères applicables par l'ensemble de l'équipe pédagogique. La réalité actuelle est bien souvent loin de ce concept. Par ailleurs, l'évaluation appliquée dans le cadre universitaire repose sur une évaluation perceptive uniquement, ce qui implique un certain nombre de contraintes exposées dans la littérature. Santerre (1989) met en avant le manque de fiabilité d'un jugement auditif, expliquant que l'oreille n'est pas un outil de

mesure fidèle: « *si elle est un bon décodeur sémantique, l'oreille est un mauvais instrument de mesure* »

Ce type d'évaluation repose sur une appréciation perceptive, propre au jugement de chaque individu. Cela peut entraîner des écarts de notation assez importants. Herry (1997) lance une évaluation subjective par six experts phonéticiens sur des productions de l'intonation d'apprenants francophones et démontre un écart important entre le jugement auditif et la transcription de schémas intonatifs. De plus, l'étude menée par Caverni & al. (1975) met en évidence le caractère subjectif de l'évaluation par ce que les auteurs appellent « les biais d'évaluation ». Les notations obtenues présentaient le cas de divergences importantes, que Caverni *et al.* expliquent par le fait que l'évaluateur serait inconsciemment influencé par des « biais d'évaluation » tels que le sexe, l'âge, l'origine sociale (etc.).

Santerre (1989) fait remarquer qu'il serait intéressant d'utiliser des logiciels spécifiques dont le système possède un module d'évaluation défini. De fait, l'évaluation de productions orales peut être complexe dans la mesure où divers paramètres entrent en jeu. Il serait donc pertinent d'appuyer une évaluation perceptive sur une évaluation quantitative reposant sur des critères acoustiques objectifs, automatisés dans un logiciel. Cela rejoindrait alors la théorie des quatre /k/ de Hansen (2006) indiquant qu'une évaluation complète doit être : « corrective, compréhensible, qualitative et quantitative ». Cependant, rares sont les études qui se sont réellement penchées sur la question, et un manque évident se fait ressentir dans le domaine de l'évaluation des productions orales en L2.

... propositions de critères ?

L'élaboration d'une évaluation prosodique en langue seconde reste rare. Sans prendre en considération les modèles de logiciels de langue, mais en nous basant sur l'évaluation de l'oral en L2 de manière générale, le constat repose sur le manque de définition de critères évaluatifs.

Si l'on se reporte au système d'évaluation DIALANG²⁵, la prosodie en est absente. DIALANG est un système d'évaluation de compétences en langues, destiné à des apprenants qui désirent obtenir un « diagnostique » sur leur compétence dans la langue cible. Les échelles d'évaluation se basent sur le « cadre européen commun de référence » dans lequel figure une des échelles d'évaluation « l'expression orale » ; on aurait pu penser que des critères prosodiques serviraient à une évaluation de niveaux, mais ce n'est pas le cas. Le critère principal de jugement du niveau se base sur l'étendue du vocabulaire utilisé dans la conversation. Ce type d'évaluation s'accorde plutôt avec les tests de diagnostics qui visent à repérer les connaissances déjà acquises par les apprenants, mais les critères requis sont toujours basés sur la grammaire et le vocabulaire. Un diagnostic de niveau de langue DIALANG a été élaboré et automatisé ; il est aisé d'observer qu'aucune place n'a été laissée pour l'expression orale. Ainsi, il est possible d'être évalué sur la compréhension orale, la grammaire et le vocabulaire. Par ailleurs, une analyse est en cours d'expérimentation (Tortel, en cours), mettant en avant le manque de fiabilité de ce programme, sachant que sur l'épreuve de compréhension orale en anglais, testée par 10 natives anglaises, la moitié des sujets obtient le niveau répertorié comme « proche du natif ». Ce programme offre de grandes possibilités, mais de nombreuses perspectives de travail sont à envisager afin d'améliorer le système proposé (Lopez & al., à paraître).

Dans le cadre de recherche sur l'élaboration de critères prosodiques destinés à une évaluation de l'oral, les premiers travaux concernent surtout le domaine du FLE.

Ainsi, dans le cadre de la création d'un test de niveau, visant à évaluer la prononciation des anglophones (apprenant le français), Di Cristo (1975) propose une évaluation prenant en compte des critères prosodiques, dont le rythme. Le test se compose de 40 phrases (chaque phrase contient une difficulté précise) qui se répartissent selon trois catégories : phonémique, phonétique, prosodie. L'évaluation prosodique, comprenant accent et rythme, se base sur les critères suivants (1975 :17) :

- absence de neutralisation des voyelles inaccentuées,
- place de l'accent de groupe,

²⁵ <http://www.dialang.org>

- distribution des accents d'insistance,
- rythme : égalité des syllabes.

Ces critères sont les paramètres de divergences décrits dans le chapitre 1 sur le rythme. Un exemple de barèmes pour l'accentuation est cité ; il se base sur une échelle à trois catégories :

- 2=distribution correcte des accents
- 1=distribution parfois incorrecte des accents
- 0=distribution toujours incorrecte des accents

Ce test comporte plusieurs avantages: tout d'abord, il permet d'évaluer le niveau des apprenants ; ensuite, exécuté une seconde fois, le test va permettre de mesurer les progrès réalisés, et enfin il fournit les renseignements ciblant les points faibles des apprenants et sur lesquels ils pourront alors se concentrer.

Freland-Ricard (1995) tente d'élaborer des paramètres rythmiques en étudiant les déviations des productions d'apprenants étrangers en FLE. Les paramètres retenus sont les durées syllabiques et les pauses. Au niveau du premier paramètre, l'auteur conclut que les sujets étrangers ne maîtrisent pas les durées des syllabes de la même manière que les Français, et qu'ils calquent l'organisation temporelle de leur langue maternelle. Concernant l'étude des pauses, Freland-Ricard constate que la moyenne des groupements entre deux pauses se situe entre 3 et 5 syllabes, que les groupements internes se caractérisent par un allongement final, et la durée de ces groupements est nettement plus courte que celle des natifs français. Les conclusions de l'auteur permettent d'attester des théories sur l'influence de la L1 sur la L2 ainsi que de l'observation d'une interlangue mais les résultats annoncés n'ont pas été exploités pour l'élaboration d'une évaluation prosodique.

Lorsque l'on se penche sur les travaux de recherche dont le but est l'élaboration de critères évaluatifs à implémenter dans un logiciel, la littérature n'est pas plus fructueuse.

Une investigation récente a été menée dans le domaine du FLE par Dodane & Konopczynski (2001) examinant la possibilité d'appliquer cette recherche de critères rythmiques dans le logiciel Winpitch LTL. Les auteurs étudient les déviations rythmiques de 5 locuteurs étrangers (2 malaisiens, 2 japonais, et un taiwanais) dans une tâche de lecture et d'oral spontané en français. Les mesures relevées sont celles des durées

syllabiques (accentuées et non accentuées) et des pics de hauteur. Les résultats montrent que, dans les productions des francophones (loc. réf.), la durée moyenne des syllabes portant l'accent primaire, est plus de deux fois plus longue que celle des syllabes non accentuées. Pour l'ensemble des locuteurs, la durée des syllabes non finales reste trop élevée ; les auteurs concluent que cela a pour effet de réduire l'écart perceptuel avec les syllabes finales portant l'accent primaire. La déviation rythmique commune à tous les locuteurs est d'associer à la proéminence mélodique qui marque l'accent secondaire, un allongement de la durée. Les auteurs remarquent qu'ils brisent alors la régularité syllabique du français, l'un des locuteurs produit une rythmicité de type accentuel. Par la suite, Dodane & Konopczynski utilisent le logiciel Winpitch afin de mettre en évidence les déviations repérées par un feedback audio-visuel. Un énoncé est pris en exemple, sur lequel elles identifient l'endroit où devraient être réalisés l'accent primaire et l'accent secondaire en français, en faisant figurer la durée des syllabes concernées. Finalement, un diagnostic des éléments rythmiques à travailler est établi (par exemple, travailler sur l'égalité des syllabes inaccentuées en leur faisant prononcer toutes les syllabes avec approximativement la même force).

Cette étude montre des résultats intéressants et exploitables mais sa visée n'a pas pour objectif une automatisation de l'évaluation prosodique.

Teixeira *et al.* (2000) tentent de prédire le degré de déviation des phrases des apprenants par rapport aux natifs. Leur travail n'est pas focalisé sur l'aspect segmental des déviations, mais sur les aspects prosodiques.

Cucchiaroni *et al.* (2000) montrent que les taux de fluence verbale (« fluency ») de natifs, sur un corpus de parole lue, peuvent être prédits par le biais de mesures temporelles calculées automatiquement sur de la parole de qualité. Les facteurs les plus influents étant le taux avec lequel les locuteurs articulent les sons et le nombre de pauses réalisées.

Imoto *et al.* (2002) proposent une méthode qui évalue l'accent lexical en anglais parlé par des apprenants japonais. L'exactitude des prédictions se situent à 95,1% pour les natifs et 84,1% pour les apprenants japonais.

Quelques travaux commencent donc à émerger, mais ils ne concernent pas les productions de francophones en anglais. A notre connaissance, la seule étude qui a été faite est celle de Herry (2001) et Herry & Hirst (2002) pour le logiciel Prosodia.

Cette évaluation présente deux objectifs : (i) une évaluation subjective d'apprenants français basée sur la prosodie (anglaise), afin de vérifier l'efficacité de la méthode du logiciel, et (ii) une évaluation objective, établissant une corrélation entre le niveau des apprenants français et des paramètres prosodiques. Cette évaluation prend en compte des paramètres rythmiques, afin d'établir une comparaison entre les productions des natifs (britanniques anglais) et celles des apprenants français.

Les paramètres rythmiques ont été choisis selon la typologie traditionnelle des langues isoaccentuelles comme l'anglais, des langues isosyllabiques comme le français et sont ceux définis par l'étude de Ramus & Mehler (1999). A cela, ont été ajoutés: (i) le pourcentage de durée vocalique, (ii) la moyenne des durées consonantiques, (iii) la variation des durées consonantiques, (iv) le coefficient de variation des durées consonantiques, (v) la moyenne des durées vocaliques, (vi) la variation des durées vocaliques, (vii) le pourcentage du nombre de voyelles, (viii) la différence de durées moyennes de la phrase des apprenants avec celles de deux locuteurs natifs, (ix) la différence (en pourcentage) entre les variations d'intensité des étudiants avec celle des deux natifs, et enfin (x) la différence (en pourcentage) entre le coefficient de variation d'intensité des étudiants et la moyenne des coefficients de variation des deux natifs (Herry, 2001 :294).

Les variables présentées ci-dessus sont corrélées avec les résultats d'une évaluation subjective qui a permis d'obtenir 6 notes sur la production des apprenants et selon des critères de notation différents. Les calculs statistiques ont été réalisés avec le logiciel CRUISE. Les analyses ont permis de déterminer des paramètres prosodiques, leur degré d'importance, et leurs intervalles pour les 6 notes. Prenons l'exemple de la note 3 (décrire les résultats obtenus pour toutes les notes serait trop fastidieux). La note 3 représente la qualité de production. L'analyse prédit des classes allant de 1 à 6 dans lesquelles sont classés les sujets en fonction du paramètre pris en compte par l'analyse. Les résultats sont donnés sous forme d'arbre. Ainsi, avec l'aide du logiciel Cruise une prédiction de deux niveaux est établie, classes 2 et 3 : la classe 2 représente les étudiants ayant obtenu entre 5-8/20 et la classe 3 les étudiants ayant entre 9-12/20. Ces classes ont été établies en fonction d'un seul paramètre : la différence de durée par rapport à la moyenne des natifs. Dans la classe 2, sur 303 valeurs, 194 sont prédites, ce qui représente 64%. Pour la classe 3, 78 valeurs sont prédites sur 187, soit 42%. Si la différence de durée est inférieure ou égale à 8%, les étudiants sont classés en niveau 3, en revanche une différence de durée supérieure

à 8% classe les étudiants en niveau 2. La conséquence de ces résultats montre, qu'en phase de production (un apprenant produisant un énoncé donné), plus l'apprenant allonge la durée de l'énoncé, plus la note diminue et inversement, plus la durée se rapproche de celle du modèle, plus la note augmente.

L'auteur conclut que les résultats sont intéressants mais pas encore satisfaisants. Elle suggère de poursuivre une évaluation objective dans de futurs travaux testant de nouveaux paramètres en plus de ceux déjà établis. Cette proposition d'évaluation n'est encore qu'au stade de test et n'est pas encore assez significative pour être incluse dans le logiciel. Le logiciel propose donc, à ce stade, un feedback par écoute et visualisation des courbes intonatives.

La synthèse de parole

Une autre technique tout à fait intégrable dans les logiciels de langue est celle de la synthèse de parole. Plusieurs études (Van Santen & al., 2009 ; Wang & al., 2009 ; Felps & al ; 2009) s'appuient sur la proposition de reformulation (« *rephrasing* ») des productions incorrectes des apprenants. Selon Lyster (2001), cette méthode a été efficace lors d'une correction explicite des erreurs phonologiques. Tous ces auteurs mettent en avant ce procédé de synthèse et expliquent que cela constituerait un feedback pertinent. L'idée consiste pour l'apprenant à écouter sa propre production de la manière dont elle aurait dû être produite. L'apprenant entend donc sa propre voix mais corrigée par la synthèse avec un accent natif.

3.4 Conclusion

A la lumière des différents travaux exposés dans ce chapitre, il est légitime de constater le manque d'études dans le domaine de l'évaluation, et de ce fait, le vide laissé par cette absence d'évaluation pertinente. C'est sur cette constatation que notre intérêt se porte. L'évaluation, essentielle pour un apprenant en langue seconde, se constitue sur un certain nombre de critères et s'ouvre sur le champ de recherche que représente « l'analyse d'erreurs », qui selon Corder (1980 : 13) est une notion stratégique pour le chercheur :

« [...] elles [les erreurs] fournissent au chercheur des indications sur la façon dont une langue s'apprend ou s'acquiert, sur les stratégies et les processus utilisés par l'apprenant dans sa découverte progressive de la langue ».

Un apprentissage prosodique (ou tout autre d'ailleurs) n'est efficace que s'il est accompagné d'un « feedback », d'une évaluation concrète qui serait basée sur des critères pertinents. Les premières études, remontant à trois décennies, ont démontré la pertinence d'un feedback, mais sans pouvoir arriver à un résultat d'application suffisant. Aujourd'hui, les nouvelles technologies ont réussi à implémenter un outil pédagogique d'apprentissage prosodique accessible et à la portée de tout apprenant. Il apparaît donc pertinent de se pencher de plus près sur cet « handicap » que présentent ces logiciels, éclopés d'un système d'évaluation automatique concret.

C'est dans cette optique, que nous n'accorderons pas d'importance à l'interrogation défaitiste de Corder (1980 : 41) se demandant :

« A quoi bon investir du temps et de l'argent dans des recherches pénibles et techniquement difficiles sur les erreurs de groupes d'apprenants si nous ne pouvons mettre à profit leurs résultats pour améliorer la capacité de communication de ces apprenants? »

... et que nous poursuivons notre étude par la recherche de critères d'évaluation intégrables à un logiciel de langues, exposée dans la section expérimentale de ce travail.

Synthèse

Pour pallier les difficultés (énoncées dans les chapitres précédents) auxquelles les apprenants sont confrontés face à un changement de rythme, les logiciels semblent être une solution intéressante.

Suite à un bref historique de l'évolution de la technologie, il est à noter que les tentatives d'intégration d'outils technologiques, notamment par la visualisation de la courbe mélodique, n'est pas chose nouvelle dans l'enseignement de la prosodie L2. Néanmoins, les aspects segmentaux de la langue ont longtemps été favorisés au détriment de la prosodie, et la complexité de la représentation de la courbe mélodique semble avoir freiné l'emploi des nouvelles technologies dans l'apprentissage des langues. De fait, l'évaluation est la composante la plus critiquée pour ces logiciels de langue.

Il a été montré que les logiciels présentaient des avantages intéressants notamment en resituant l'apprenant au cœur de l'apprentissage.

La phonétique n'est pas suffisamment mise en avant dans l'enseignement des langues secondes, particulièrement dans le cadre collectif auquel les enseignants sont affectés, mais cela n'est pas étonnant, vu les diverses contraintes qui lui sont inhérentes.

Dès lors, il est assez difficile de consacrer le temps nécessaire à la pratique de la phonétique, sinon quelques corrections occasionnelles des rares productions des étudiants. Champagne-Muzar & Bourdages (1998) expliquent la marginalisation de la composante phonétique, et particulièrement de la prosodie, dans l'enseignement des langues par le « *contraste entre la rapide évolution des ressources et pratiques pédagogiques pour le cours de L2 et la monotonie des méthodes de phonétique corrective* ». Les auteurs proposent une solution basée sur l'utilisation de logiciels adaptés : « *Le besoin urgent de fournir à l'enseignant et à l'apprenant un véritable outil capable d'intégrer dans ses fonctions les nouvelles orientations pédagogiques* ».

Ainsi, un enseignement « mixte » alliant approches didactiques traditionnelles, pratiques et technologiques ne peut constituer qu'un apport bénéfique pour l'apprenant.

Cependant, même si les nouvelles technologies ont démontré qu'elles surpassaient le manque de feedback visuel accessible et interprétable par tout un chacun, le module d'évaluation de ces logiciels restent très incomplet. En effet, jusqu'à présent, l'évaluation proposée repose sur un feedback visuel mais qui ne prend pas en compte les multi-paramètres prosodiques sous-jacents. Ce constat, ainsi que le manque de travaux dans ce domaine, atteste du besoin de nombreuses perspectives de recherches.

La voie à de nombreux travaux, concernant l'évaluation prosodique automatique pour les logiciels d'apprentissage de la prosodie des langues, est donc pleinement ouverte.

Section 2.

Analyses expérimentales

CHAPITRE 4.

CONSTITUTION ET EXPLOITATION

D'UNE BASE DE DONNEES : ENGLISH



"Gina is by lingal . . . that means she can say the same thing twice, but you can only understand it once."

DENNIS THE MENACE® used by permission of Hank Ketcham and by North America Syndicate.

Ce chapitre décrit le déroulement des différentes phases qui ont mené à la constitution de la base de données ANGLISH, nouveau corpus oral d'anglais L1 & L2 (fusion des mots 'anglais' et 'English'). Sont également décrites les étapes d'annotation et de segmentation manuelle du corpus en unités intonatives, en mots, en syllabes, en unités rythmiques (modèle de Jassem, 1952) et en phonèmes (selon un codage CVC).

4.1 Remarques préliminaires

Nous avons soulevé dans la partie introductive de cette thèse, l'absence de disponibilité de corpus d'anglais L2 produit par des francophones. Or, pour mener à bien l'élaboration de critères prosodiques, il est nécessaire d'utiliser un corpus oral d'anglais britannique suffisamment important, qui regroupe des enregistrements de productions en anglais L1 et L2. Dès lors, un des premiers objectifs de cette recherche consiste en la création d'une base de données conséquente afin de mener à bien les analyses qui vont suivre (cf. chapitre 5), c'est-à-dire chercher à déterminer où se situent les écarts prosodiques entre des productions de natifs anglais et d'apprenants francophones dans le but d'établir des paramètres destinés à une évaluation prosodique.

Malgré tout, le manque d'un tel corpus semble tout à fait surprenant, tout particulièrement lorsque l'on considère le statut de l'anglais dans le monde et la place que cette langue occupe en France. En effet, l'anglais est la langue officielle de plus de 50 pays et elle est parlée en tant que langue maternelle par plus de 330 millions de terriens (Weber, 1997). Estimée comme l'actuelle « langue internationale », elle est assurément la seconde langue la plus apprise et étudiée au monde (Crystal, 2003) avec un nombre toujours accru d'apprenants (estimé à 150 millions en 2008 par Weber d'après une interview sur son ouvrage en 2008 : « The number of speakers of *all* the top ten languages have gone up in the last quarter century but relative to each other, the situation among the top ten remains unchanged »). Ces chiffres nous semblent un peu faibles, et effectivement lorsqu'on se réfère à l'étude de Graddol (2000 :10), ce dernier annonce 375 millions de locuteurs anglais L1 et 750 millions d'apprenants. L'anglais est en effet, la première langue étrangère étudiée à partir de l'école primaire dans de nombreux pays (Allemagne, Espagne etc.), y compris la France, pour laquelle le site www.nationmaster.com annonce environ 16 millions d'apprenants sur la population totale.

Conjointement, la littérature sur l'acquisition des langues est de plus en plus abondante, et les travaux sur l'acquisition de l'anglais L2 se font grandissants. Lorsque l'on se penche sur le domaine de l'acquisition des langues secondes et notamment l'acquisition de l'anglais L2 par des francophones, la constitution et le travail sur des corpus oraux sont plus rares. Pourtant, les travaux menés dans le cadre de l'apprentissage de l'anglais L2 ne cessent d'augmenter tant au niveau segmental qu'au niveau

suprasegmental. Actuellement de nombreuses études sont menées, par exemple, sur « l'accent étranger » d'un français parlant anglais (Horgues, 2005) ou de locuteurs d'autres nationalités parlant anglais (Cheong, 2007 ; Jilka, 2000).

La recherche en linguistique, quel que soit son domaine, semble se tourner vers la constitution et l'utilisation de production de parole naturelle. De ce point de vue là, le travail sur corpus représente la base fondamentale des travaux entrepris. Les études portant sur l'oral et l'acquisition des langues L2 ont alors grand intérêt à se fonder sur de grands corpus de parole, dont le style répond aux exigences et objectifs visés. L'utilisation de corpus dans l'apprentissage des langues n'est d'ailleurs pas récente, et c'est dans les années 1990 que le travail sur de larges corpus est apparu notamment avec la 'naissance' du TaLC (Teaching and Language Corpora). Cependant, comme le remarque Mauranen (2004) si la production et les échanges oraux en L2 font l'objet de nombreux travaux et occupent une place largement notable, le travail sur corpus oraux reste particulièrement mis de côté, les corpus utilisés étant majoritairement des corpus écrits. Le corpus ICLE²⁶ (International Corpus of Learner English) en est une parfaite illustration. En effet, l'ICLE est un important projet qui est le résultat d'une collaboration de plus de 10 ans d'activités ; sa conception est très riche et originale, mais ce corpus ne comporte qu'une base écrite reposant sur deux types d'exercices : un commentaire composé argumentatif et une dissertation littéraire (environ 500 à 1000 mots pour chacun des exercices). Nous ne trouvons aucune équivalence du même type pour des corpus oraux. Boulton (2008) fait état des différents corpus existants en DDL ; sur 39 recensés, nous constatons que seulement 11 portent sur l'oral des apprenants et semblent de surcroît être difficilement accessibles. Malgré tout, actuellement, un seul projet en cours se distingue des remarques formulées ci-dessus. Il s'agit du projet LONGDALE (Granger & al., 2009), qui a pour objectif de constituer un corpus d'apprenants avec une partie consacrée aux apprenants francophones de l'anglais L2.

Parallèlement à cela, nombreux sont les corpus d'anglais L1, et dont les utilisations sont spécifiques à certains types de travaux ou répondent à des exigences en fonction des

²⁶ <http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/iclc.htm>

diverses approches choisies. En effet, de nombreux travaux sur l'anglais, quel que soit leur cadre théorique, semblent s'appuyer sur la volonté d'utiliser des stimuli naturels. Ainsi plusieurs grandes bases de données, orientées selon le type de travaux, sont exploitables et disponibles (études inter-dialectales avec IVIE²⁷ (Kochanski & al., 2004), étude d'un phénomène particulier sur de l'anglais spontané avec la base de données Aix-Marsec (Auran & al, 2004), ou encore une base de données comparatives sur les langues européennes telle que EUROM1 (Chan & al. ,1995), etc.). On retrouve également plus d'une vingtaine de corpus d'anglais L1 sur le site de l'UCREL²⁸ (University Centre for Computer Corpus Research on Language) tels que le BNC (90% de données écrites), le Lancaster/IBM Spoken English Corpus devenu SEC puis Aix-Marsec ou encore ITU (International Telecommunication Union) pour exemples de corpus oraux. Egalement, des centres de ressources linguistiques tels que le LDC²⁹ (Linguistic Data Consortium) ou encore l' ELRA³⁰ (European Language Resources Association) mettent à la disposition de la communauté un grand nombre de corpus anglais L1, tel que AURORA (*ELRA*), base de données développée dans le but principal d'étudier l'influence sur la performance de la reconnaissance automatique de la parole pour des entrées de parole en mode mains-libres dans des environnements bruités.

De ce constat général s'est imposée la nécessité de la constitution d'une ressource telle que ANGLISH.

²⁷ http://www.phon.ox.ac.uk/old_IViE

²⁸ <http://ucrel.lancs.ac.uk/corpora.html>

²⁹ <http://www.ldc.upenn.edu/>

³⁰ <http://catalog.elra.info/index.php>

4.2 Origine du corpus ANGLISH

4.2.1 Corpus de base : Eurom 1

Les extraits/passages qui ont servi de base et de support à nos enregistrements (destinés à la création du corpus ANGLISH) sont issus de la base de données multilingue du corpus EUROM 1. Ce corpus a été créé dans le cadre de travail du projet européen SAM_A, lui-même extension d'un projet préliminaire appelé SAM (Speech Assessment Methods) dont le but était de produire une base de données de parole commune en langues européennes. La continuité du projet s'est appuyé sur le développement progressif de la base de données désormais appelé EUROM afin de mettre en avant l'existence d'un corpus de langues européennes disponible sur Cdroms.

Ainsi le corpus EUROM 1 contient des enregistrements de 60 locuteurs en 11 langues européennes : danois, néerlandais, anglais britannique, français, allemand, norvégien, suédois, néerlandais, grec, portugais, espagnol. Il a donc été explicitement conçu pour faciliter les comparaisons phonétiques entre langues, en s'appuyant sur des matériaux similaires et des protocoles d'enregistrement identiques dans toutes les langues.

4.2.2 Description du corpus Eurom 1

Pour chacune des 11 langues proposées, 4 types d'exercices de lecture ont été enregistrés suivant l'ordre annoncé ci-dessous :

- 121 logatomes isolés puis en contexte,
- 100 nombres allant de 0 à 9999,
- 40 passages de 5 phrases chacun relatant une histoire du quotidien,
- 50 phrases isolées provenant des passages proposés.

Le corpus a été divisé en 3 « sous-corpus » :

- **Many Talker Corpus** (MANY) : 100 nombres, 3 passages, 5 phrases, lus par 30 hommes et 30 femmes,

- **Few Talker Corpus** (FEW) : 5 x 121 logatomes, 5 x 100 nombres, 15 passages et 25 phrases, lus par 5 hommes et 5 femmes sélectionnés à partir du corpus MANY,

- **Very Few Talker Corpus** (VERY FEW) : 121 logatomes en contexte lus par 1 homme et 1 femme sélectionnés du corpus FEW.

Les enregistrements ont été faits en chambre sourde à une fréquence de 20KHz et une résolution de 16 bits, les données recueillies directement sur disque dur.

Eurom 1 : corpus anglais

Le corpus en langue anglaise est composé des enregistrements énumérés ci-dessous :

- **MANY** corpus : 60 locuteurs (30 hommes et 30 femmes) ont lu 3 passages, 5 phrases et 100 nombres,
- **FEW** corpus : 10 locuteurs (5 hommes et 5 femmes) ont enregistré 15 passages, 25 phrases, 5 x 100 nombres et les logatomes isolés,
- **VERY FEW** corpus : 2 locuteurs (1 homme et 1 femme) ont lu les logatomes en contexte.

4.2.3 Enrichissement du corpus Eurom 1 (corpus anglais)

Etat des lieux

Parmi les exercices proposés dans le corpus Eurom1 anglais, on trouve la lecture de 40 passages (comme indiqué ci-dessus dans la description du corpus Eurom1). La répartition de ces 40 passages se fait en 4 groupes de 10 passages nommés respectivement, O, P, Q, R [+ le numéro du passage], chacun sous-divisés en groupe de 5 passages. Ci-dessous la répartition des enregistrements effectués et classés par locuteur (tableau issu et adapté du corpus Eurom 1).

	Groups of passages							
speakers	o1-o5	o6-o0	p1-p5	p6-p0	q1-q5	q6-q0	r1-r5	r6-r0
fa	x	x	x					
fb					x	x	x	
fc				x	x	x		
fd	x	x						x
fe	x						x	x
ff		x	x	x				
fg			x	x	x			
fh						x	x	x
fi	x	x	x					
fj				x	x	x		

Tableau 4. Répartition des enregistrements des passages par locuteur (corpus EUROM 1)

Il nous a semblé intéressant d'exploiter l'ensemble des 40 passages proposés dans le corpus Eurom1 mais en application à un même locuteur. Dans la mesure où les enregistrements effectués dans ce corpus sont répartis de manière assez aléatoire, la seule 'directive' imposée était que chacun des locuteurs devait lire 15 passages parmi les 40. Une exploitation de la totalité des passages pour un même locuteur et ce, par plusieurs locuteurs, semblait être un apport important afin d'obtenir une base de données comparative inter-passages plus large.

Contributions personnelles : extension du corpus EUROM 1

Un premier corpus, nommé eurom1_EN_complete, a été enregistré afin d'élargir la base de données de EUROM1. Ainsi, cinq locuteurs ont été enregistrés dans une tâche de lecture des 40 passages en anglais, extraits du corpus Eurom1. Pour les besoins de cette thèse, certains passages de ces enregistrements ont été exploités dans la partie expérimentale, à savoir 4 passages lus des trois locuteurs britanniques, et font ainsi partie du corpus ENGLISH, présenté ci-dessous. Chacun des passages se compose de 5 phrases liées sémantiquement (soit au total 200 phrases) racontant un événement, une histoire de la vie quotidienne. La consigne donnée à la fois oralement et à l'écrit, en français et en anglais, (cf. tête de page des feuillets présentés aux sujets) était de :

« Lire les passages comme si l'histoire vous était arrivée et que vous la racontiez à un ami ». Les passages étaient donnés à l'avance de manière à ce que les locuteurs se familiarisent avec les textes et puissent éviter les bafouillages lors de l'enregistrement. Les passages ont été donnés 10 par 10, une pause ayant été prévue entre chaque groupe de textes, de façon à ne pas surcharger le locuteur.

Les enregistrements se sont déroulés en chambre sourde au laboratoire de recherche Parole et Langage à Aix-en-Provence. Un micro-casque a été utilisé pour les enregistrements. Les données ont été directement récupérées sur ordinateur sous format wav en qualité numérique mono 16 bits avec un échantillonnage de 44 Hz.

Parmi les 5 locuteurs, trois locuteurs, deux femmes et un homme d'une moyenne d'âge de 24 ans, sont britanniques anglais et de langue maternelle anglaise. Les deux locutrices viennent du sud de l'Angleterre (Hemel Hempstead et Chertsey, respectivement nord-est et sud-est de Londres) et le locuteur est originaire du nord-est du pays (Yorkshire). Les deux autres locuteurs, une femme et un homme d'une moyenne d'âge de 26 ans, sont américains originaires de l'état de Californie et de langue maternelle anglaise

(anglais américain). Tous les cinq sont lecteurs au département d'anglais de l'université de Provence. Ils ont tous été volontaires pour enregistrer les différents passages. Un consentement de participation et de droit de diffusion des enregistrements a été présenté et signé par chacun des participants.

A titre de récapitulation, a) pour l'accent anglais et b) pour l'accent anglais américain :

a). 120 passages soit 600 phrases prononcées par 3 locuteurs britanniques anglais sont enregistrées, ce qui représente un total d'environ 36 minutes de parole lue. La durée d'un passage varie de 10,6073 à 30,2146 secondes en fonction du locuteur.

Locuteur	durée (min. sec.)	durée (max. sec.)	moyenne (sec.)	nombre passages	durée totale (sec.) (min.)	
F01GB	10,6073	30,2146	18,5666	40	742,6651	12,37
F02GB	12,5439	27,4705	18,8120	40	752,4834	12,54
H01GB	12,0811	23,1789	16,7657	40	670,6307	11,17
Total				120	2165,7794	36,09

b). 80 passages soit 400 phrases prononcées par 2 locuteurs américains sont enregistrées, ce qui représente un total d'environ 25 minutes de parole lue. La durée d'un passage varie de 10,8442 à 27,8387 secondes.

Locuteur	durée (min. sec)	durée (max. sec)	moyenne (sec.)	nombre passages	durée totale (sec.) (min.)	
F1AM	10, 8442	12, 0021	19, 6934	40	787,7363	13,12
H1AM	27, 6658	27, 8387	17, 1488	40	685,9543	11,43
Total				80	1473,6906	24,56

4.4 ENGLISH : création d'un corpus anglais L1 & L2

4.4.1 Choix, réflexions et exigences

La conception retenue pour ce corpus s'imposait par différents souhaits et exigences que nous exposons dans les thèmes développés ci-dessous. Les quatre axes majeurs interdépendants qui ont guidé la conception du corpus reposaient sur :

- (i) la volonté de mener à bien des travaux sur l'évaluation de la prosodie d'anglophones et de francophones apprenants en anglais L2, permettant de créer une évaluation objective intégrable à des logiciels de langues, cadre de travail de cette thèse,
- (ii) créer au moins un exercice offrant une base de données comparable interlocuteurs,
- (iii) la volonté de combler le manque de larges corpus oraux en anglais L2 parlé par des francophones,
- (iv) la volonté de créer une ressource qui serait utilisable, outre ses propriétaires, par la communauté scientifique mais aussi enseignante,
- (v) une libre diffusion qui serait destinée à un public d'utilisateurs secondaires avec des buts différents, allant du simple apport de support sonore de natifs anglais (anglais L1) pour un contenu théorique sur un fait de langue particulier, à l'analyse d'erreurs segmentales ou prosodiques de productions de francophones en anglais L2 par exemple, ou tout autre étude possible.

ENGLISH : corpus, base de données ?

Avant de présenter une description détaillée de la constitution d'ENGLISH, nous souhaitons faire un bilan terminologique très bref.

« Corpus » est le premier terme qui émerge de manière dominante lorsque l'on nomme un matériel linguistique de ce type. Si nous prenons en compte la définition de Habert et al. (1997), nous pouvons alors considérer qu'ENGLISH a toutes les caractéristiques d'un corpus traditionnel :

« une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon de langage ».

Cela nous conduit à une définition classique du corpus, comme un rassemblement d'enregistrements audio et qui catégorise ANGLISH en tant que tel. Cependant, cette définition semble assez limitative. Durand (2003) nous propose une définition un peu plus complète issue de Gibbon et al. (1998), qui correspondrait à notre propre collection d'enregistrements:

« a corpus is any collection of speech recordings which is accessible in computer readable form and which comes with annotation and documentation sufficient to allow re-use of the data-in-house, or by people in other organisations ».

Ainsi, nous complétons notre première définition par l'idée que l'ensemble des enregistrements audio est alors accompagné par différents niveaux (ou niveau unique) de transcriptions. C'est d'ailleurs en ce sens là, que nous relevons la tâche considérable et la difficulté de travail qui nous sont apparues, tant au traitement des données audio (environ 5h30), qu'au niveau des enregistrements, et que pour les différents traitements orthographique, extractions des données, annotations (etc.).

Cependant, par contraste avec la représentation classique du corpus comme élément achevé ou abouti, un second concept consiste à examiner ANGLISH également en tant que base de données. Selon le jargon français³¹, une base de données se définit comme telle :

« une base de données doit être conçue pour permettre une consultation et une modification aisée de son contenu, si possible par plusieurs utilisateurs en même temps ».

ANGGLISH répond donc également aux exigences d'une base de données, comportant une structuration particulière des données (enregistrements structurés en catégorie, cf. figure 14, § alignement acoustique manuel) et une possibilité de requête (Auran, 2003), dans la mesure où les informations classées de type TextGrid dans le logiciel Praat (Boersma & Weenick, 2009) sont isolées sur des niveaux distincts nommés « tiers » (appelés désormais « rangées », selon le terme emprunté à Cho (2009)), et

³¹ <http://www.linux-france.org/prj/jargonf>

permettent commodément la formulation de requêtes diverses. Ainsi l'annotation des données peut se faire sur plusieurs années.

ANGGLISH, en tant que base de données, s'inscrit donc dans le cadre d'une conception dynamique qui attend de recevoir une constante évolution.

Ce point terminologique nous permet de conclure qu'ANGGLISH peut être considéré comme une base de données évolutive, mais pour des raisons de simplicité, nous nous permettrons d'utiliser à diverses reprises pour ce travail de thèse, le terme « corpus », sans bien entendu, que les différents points de vue développés ci-dessus, ne soient remis en cause.

Parole de laboratoire vs parole spontanée vers parole naturelle

Outre les corpus écrits, on distingue dans les travaux de phonétique et phonologie, deux types de corpus, à savoir les corpus de parole de laboratoire et les corpus de parole spontanée. La parole de laboratoire correspond à des enregistrements de parole effectués dans des conditions expérimentales, très souvent en chambre sourde de manière à obtenir une bonne qualité de son. Les données se composent de lecture de logatomes, de mots, de phrases isolées, de textes ou de phrases préparées en amont. Les exercices proposés étant plutôt de la parole contrainte, cela peut parfois aboutir à des résultats peu naturels voire artificiels. En revanche, cela permet à l'expérimentateur d'étudier un phénomène précis en utilisant des stimuli contrôlés. Nous retiendrons comme atouts pour le corpus ANGLISH que ce type de méthodologie permettra d'obtenir une qualité du signal sonore optimale (cf. qualité technique du corpus) mais aussi d'en retirer une base de données comparables pour l'ensemble des locuteurs, ce qui nous fera pencher pour une tâche de lecture de textes. Le deuxième type de parole, qui est celui de la parole spontanée, vient s'opposer à la parole de laboratoire dans le sens où aucune contrainte n'est imposée aux locuteurs, ces derniers étant enregistrés dans des conditions naturelles de production (échanges, débats, interviews, etc.) qui ne nécessitent pas à l'expérimentateur d'avoir préalablement organisé des stimuli. L'avantage de ce type de corpus est qu'il peut être recueilli en chambre sourde. De plus en plus d'études se tournent vers de la parole naturelle, de manière à obtenir des caractéristiques authentiques des productions des locuteurs. C'est dans ce sens là que nous avons souhaité allier à la fois qualité optimale du son (enregistrements effectués en

chambre sourde), données comparables par les tâches de lecture et répétitions (cf. variétés des genres) et enfin parole naturelle (il a été demandé aux locuteurs de parler spontanément d'un sujet de leur choix, racontant une anecdote sur, par exemple, leurs dernières vacances). Ainsi ANGLISH combine trois types de données orales tout en gardant une qualité acoustique idéale à des analyses phonétiques.

Focus sur l'oral

Comme nous l'avons expliqué précédemment, il existe une réelle paucité de corpus oraux d'anglais L2 produit par des francophones alors qu'il existe un nombre assez important de corpus écrits disponibles et exploitables pour travailler sur l'anglais et sur le français. Par ailleurs, parallèlement à cela, un travail considérable reste à faire dans le domaine de l'apprentissage de l'oral de l'anglais. A plusieurs reprises, nous avons pu constater le peu de place laissé à l'oralisation des langues proportionnellement à celle faite à l'écrit. Les épreuves de langues du baccalauréat, par exemple, sont d'abord menées sur un support écrit, l'oral n'arrive que de manière optionnelle ; il a d'ailleurs été constaté que plus de 70% des lycéens à l'échéance de leur scolarisation de second cycle, ayant étudié l'anglais pendant plus de 7 ans, atteignent un niveau de compétence médiocre selon les rapports européens³².

Par ailleurs, lorsque l'on se penche sur le premier système majeur d'évaluation fondé sur le cadre commun européen du Conseil de l'Europe, DIALANG, aucune place pour l'oral n'a été faite lors de la conception du 'test-diagnostique'. Ainsi, de nouveau, ce sont des compétences écrites de la langue qui sont évaluées. Il nous semblait donc important qu'une ressource de productions orales soit créée afin de (i) pallier un manque de corpus anglais L2 et (ii) obtenir une base de données qui pourrait donner lieu à de nombreux travaux sur l'évaluation de l'oral et qui permettrait d'inclure un système d'évaluation automatique de l'oral à ces tests déjà existants et incomplets.

³² European network of Policy Makers for the evaluation of Education Systems (2004) The Assessment of Pupils' Skills in English in Eight European Countries 2002. A European Project. <http://cisad.adc.education.fr/revu/pdf/lire16.pdf>

Variétés des genres

Pour cette base de données, l'intention première était de réaliser à la fois des tâches représentatives du contenu d'un logiciel de langue destiné à l'apprentissage de la prosodie de l'anglais, mais aussi des exercices que l'on retrouve en contexte d'apprentissage en parcours universitaire. Nous souhaitions également obtenir des données identiques en terme de contenu qui soient comparables pour l'ensemble des locuteurs ; notre choix s'est donc porté, tout d'abord, sur une tâche de lecture que nous avons complétée avec une tâche d'imitation, avec des répétitions de phrases et de la parole naturelle (cf. § contenu de la base de données pour une description détaillée).

Variétés des niveaux d'apprentissage des locuteurs

Afin de constituer nos différents groupes de locuteurs, la question de la représentativité a été de nouveau abordée. La constitution d'un précédent corpus (Tortel, 2004) a permis de mettre en avant la problématique des groupes de niveaux. Quatre groupes avaient été constitués : français non spécialistes de la langue, étudiants de 1^{ère} année, étudiants de 4/5^{ème} année, et natifs anglais. Il était apparu, au vu des résultats obtenus, que les groupes de non spécialistes et étudiants de 1^{ère} année avaient des résultats quasi équivalents, et même chose pour les étudiants de 4/5^{ème} année avec le groupe des natifs. Cette difficulté se retrouve également dans les travaux menés en pathologie, comme l'illustre les travaux de Revis (2004). Il était donc souhaitable de constituer des groupes de niveau plus distincts, par rapport à leur niveau de langue (niveau académique).

Qualité technique du corpus

Pour rendre possible l'analyse expérimentale, la constitution d'une base de données sonores impose une précaution particulière et des caractéristiques techniques précises quant à l'enregistrement effectué ; en effet, le choix du lieu d'enregistrement et du matériel utilisé (entre autres) sont des questions essentielles et primordiales pour mener à bien une analyse phonétique du signal (conditions identiques pour chaque enregistrement, absence

de bruits de fond, matériel de qualité, bon rapport signal/bruit, etc.). Ainsi, pour une qualité sonore optimale, tous les enregistrements ont eu lieu en chambre anéchoïque (dans les locaux du LPL) à une fréquence de 44KHz et une résolution de 16 bits et ont été effectués à l'aide d'un micro-casque. Les données ont été recueillies directement sur disque dur sous format wav., PCM, en mono.

4.4.2 ENGLISH : Description du corpus

Le choix du corpus a été un des points essentiels de cette étude. En effet, il représente la base de ce travail, et les résultats et interprétations possibles de cette recherche dépendent en grande partie du choix du type de corpus. La constitution d'un corpus ne signifie pas simplement de regrouper des données (écrites ou orales) ; en effet, cela entraîne un travail qui s'effectue en trois phases : (i). un travail préparatoire qui est réalisé avant la phase d'enregistrement des données, (ii) l'enregistrement des données, (iii) la structuration et mise en forme des données afin qu'elles soient utilisables.

Nous souhaitons, pour ce travail de thèse, un corpus qui regroupe les exigences énoncées ci-dessus. Nous constatons que de nombreux enregistrements sont constitués sur des productions de français en anglais répondant aux exigences d'une étude précise, mais peu d'entre eux se développent et acquièrent la casquette de « corpus », peu regroupent toutes les caractéristiques souhaitées pour notre recherche et aucun ne semble être accessible librement.

Nous présentons ici, une description détaillée du contenu de la base de données ENGLISH.

De Eurom 1 à ENGLISH : choix des textes et préparation des stimuli

Une préparation préalable aux exercices, comprenant le choix des textes et le montage de la bande son pour la phase de répétition, était nécessaire.

Le choix de base s'est porté sur les passages du corpus Eurom 1, car les textes proposés nous semblaient adaptés à une tâche de lecture, dans laquelle le sujet pouvait tout

à fait s'approprier le texte, étant une histoire du quotidien qui pourrait arriver à n'importe lequel d'entre nous.

La difficulté majeure a été la sélection des quatre textes. Eurom 1 en comporte 40 (comme expliqué précédemment), ce qui laisse un choix assez important. Mais il est à noter que parmi nos groupes, se trouve celui des faux débutants pour lesquels nous souhaitions absolument éviter une difficulté segmentale trop importante (puisque ce n'est pas ce qui nous intéressait en premier lieu), mais aussi limiter les bafouillages. Il est vrai qu'une difficulté de lecture trop importante aurait empêché la compréhension et la libre interprétation du texte, et ainsi le non respect « involontaire » de la consigne par le locuteur. Pour cela, nous avons évité de sélectionner les passages comportant des nombres, des dates, des numéros de téléphone, des noms propres, des noms de lieux tels que « Ireland », « Edgebaston », « Tewkesbury », des adresses, des noms de maladie (« hepatitis A », « typhoid fever ». Une fois ces critères pris en compte, le choix se faisait parmi une quinzaine de textes. Nous nous sommes alors penchée sur le nombre de syllabes dans les phrases et avons répertorié le nombre total de syllabes de chacune des phrases de chacun des passages (cf. annexes, section 1). Nous souhaitions éviter que les phrases soient trop courtes et comptent moins de 7 syllabes, mais aussi maîtriser le débit en contrôlant les longueurs de phrases par les syllabes.

Un autre critère était l'occurrence des mots polysyllabiques, envisageant un déplacement d'accent lexical vers la gauche de la part des locuteurs francophones. Ainsi nous avons tenté d'avoir au moins un mot quadrisyllabique dans chaque passage, puis plusieurs bi -et trisyllabiques. Les mots de deux syllabes semblaient suffisants pour observer le phénomène de déplacement d'accent et d'allongement de syllabes inaccentuées au niveau des locuteurs de niveau faible. En revanche, il était important de disposer de mots de plus de trois syllabes pour le groupe d'étudiants à qui ces mots posent plus de difficultés. Après ce premier tri, nous avons sélectionné six textes que nous avons testés sur une dizaine de sujets. Les sujets étaient des membres de la famille ou des proches qui n'ont pas participé par la suite aux enregistrements destinés à la base de données ANGLISH et avaient donc été choisis simplement pour une phase test. Nous leur avons demandé de lire les 6 textes à voix haute (dans les mêmes conditions que pour les « réels » enregistrements, c'est-à-dire, avec une prise de connaissance des textes en amont, et une prise de son dans la chambre sourde) ; dans un second temps, leur tâche était de sélectionner les 4 textes sur lesquels ils se sont sentis à l'aise, à la fois dans la tâche de

lecture et de compréhension. Sur une douzaine d'interrogés, la majorité avait choisi les 4 textes qui ont été présentés lors de la phase finale réelle d'enregistrements.

Concernant la phase de répétition (phase 2), nous avons créé une bande son en utilisant le logiciel CoolEdit, que nous avons réalisée à partir des énoncés des quatre textes proposés en lecture, ainsi qu'un texte supplémentaire. Nous avons donc extrait et isolé 25 phrases (cf. annexes, section 1) prononcées par les natifs anglais enregistrés pour le corpus Eurom 1 qui servaient de modèles de référence aux sujets. Nous avons récupéré un son de type « bip » qui servait d'alerte à chaque fois que le locuteur allait entendre une nouvelle phrase. Dix secondes ont été planifiées entre chaque phrase, ce qui laissait aux sujets la possibilité de se reprendre s'ils jugeaient leur prestation mauvaise. La bande-son était diffusée à l'aide d'un appareil mp3, et le son diffusé à l'intérieur de la chambre sourde par une enceinte reliée par un câble à l'appareil. Deux modèles de référence ont été créés, l'un avec une voix de femme, l'autre avec une voix d'homme.

Contenu de la base de données

Le corpus ENGLISH se compose de trois phases d'exercices différents en anglais enregistrées selon l'ordre indiqué suivant :

- *Lecture*

Il s'agit d'une tâche de lecture de 4 passages (cf. annexes, section 1) issus du corpus EUROM 1 (P9, Q0, Q9, R3), lus par 63 locuteurs, ce qui représente un total de 252 passages (soit 1260 phrases) pour une durée approximative d'une heure et demie de lecture (soit 4765,3138 secondes). Chacun des passages se compose de 5 phrases liées sémantiquement racontant un évènement, une histoire de la vie quotidienne. La consigne donnée à la fois oralement et à l'écrit était de :

« Lire les passages comme si l'histoire vous était arrivée et que vous la racontiez à un ami ». Les passages étaient donnés à l'avance de manière à ce que les locuteurs se

familiarisent avec les textes et puissent éviter les bafouillages lors de l'enregistrement. Ci-dessous, un exemple de passages proposé :

“I've always found it difficult to sleep on long train journeys in Britain. For one thing, I can never make myself comfortable in the seat. Then the other passengers usually talk so loudly or worse still they snore. In addition, there's the constant clickety-click of the wheels on the track. If I do manage to doze off the ticket inspector comes along and wakes me”.

La tâche de lecture permet de contrôler les variables intra-locuteurs et permet d'obtenir des données parfaitement comparables pour chacun des locuteurs et ce quel que soit son niveau en anglais. Par ailleurs, cette tâche semble être considérée comme la base d'une compétence linguistique sur laquelle un enseignant s'appuie pour contrôler les acquis de ses étudiants. En effet, outre un contrôle de l'acquisition de la syntaxe ou des éléments du niveau segmental, la tâche de lecture oralisée joue un rôle d'apprentissage par un entraînement de l'apprenant à l'articulation, la respiration, le dosage du débit, les pauses mais aussi à l'intonation (Lieberman *et al.*, 1967). L'exercice devient donc une tâche de lecture expressive où le lecteur tente de donner vie à un texte en traduisant des sentiments, en lui donnant une tonalité et donc une intonation particulière.

▪ *Répétition*

Il s'agit d'une répétition de 25 phrases (par 63 locuteurs), selon un modèle natif (féminin ou masculin) enregistré dans le cadre du corpus EUROM 1, issues des 4 passages utilisés pour la tâche de lecture ainsi qu'un passage (R3) ajouté, toujours extrait du corpus EUROM 1 (cf. annexes, section 1). La consigne donnée était de « répéter les phrases suivantes en essayant d'imiter la façon dont elles sont produites ». Cette tâche représente un total de 1575 phrases, soit une durée totale approximative d'une heure et demie (soit 4811, 2246 secondes) pour l'ensemble des phrases répétées. En cas de bafouillage, le sujet avait pour consigne de se reprendre ou recommencer la phrase, le pilotage de la bande son des phrases étant mené par l'examinatrice elle-même (sachant qu'à la base, une durée de

10 secondes avait été fixée entre chaque phrase). Ci-dessous, un exemple de quelques phrases proposées sur les 25 données :

- « [...] 04. In addition, there's the constant clickety-click of the wheels on the track.
05. If I do manage to doze off the ticket inspector comes along and wakes me.
06. What can I have for dinner tonight?
07. I do have some fresh pasta in the fridge.
08. The trouble is I eat that two or three times a week. [...] »

▪ *Monologue*

Cela consiste en un monologue de 2 minutes environ (produit par 63 locuteurs), sur un sujet non imposé (le thème des vacances avait été suggéré à chacun des sujets mais n'était pas imposé afin que le locuteur soit le plus possible à l'aise et confiant dans cet exercice). Ainsi, les locuteurs avaient 'libre antenne' pendant environ deux minutes pour parler d'un sujet de leur choix, non préparé, si ce n'est quelques minutes, le temps de réfléchir à un thème approprié. Cette tâche représente une totalité de 63 monologues, soit une durée totale approximative de parole de trois heures (soit 10799,9619 secondes). Ci-dessous, un extrait d'un monologue produit par une locutrice (F03GB) :

« so for the last three or perhaps since er for four weeks I've been learning this Chopin ballad it's the first one in C minor it's terribly difficult and erm whereas I've been working at it quite a lot there's still a lot more to do what I have to do is break it down into small parts and try to work on the technical aspect of some of it [...] » (cf. annexes, section 1).

Certains auteurs tels que Guaitella (1991) pensent que la prosodie d'un texte lu ne reflète pas la prosodie d'une parole spontanée. Nous avons tenu compte de cette remarque et tenté d'obtenir une partie plus spontanée où le locuteur n'est plus aidé par un support écrit par lequel il n'avait pas à réfléchir à la syntaxe, à la grammaire ou au vocabulaire. Cependant, comme le fait remarquer Freland-Ricard (1995), les corpus dits de parole spontanée sont souvent des extraits de journaux télévisés, d'interviews, ou encore des

entretiens radiophoniques qui utilisent soit un prompteur, soit des fiches de prises de notes préparées à l'avance, le tout étant répété à plusieurs reprises afin de maîtriser l'intervention au moment voulu. La spontanéité n'est plus tout à fait celle que l'on pensait étudier.

Locuteurs

Pour les raisons énoncées précédemment, le choix des locuteurs s'est porté sur un étalonnage varié de deux niveaux assez distincts d'apprenants et un niveau contrôle représenté par les natifs.

ENGLISH compte donc 63 locuteurs, portés volontaires pour enregistrer les trois exercices proposés qui sont répartis en trois groupes : (i) le groupe témoin qui est celui des natifs anglais (GB), (ii) le groupe d'étudiants de 2/3^{ème} année, spécialistes de la langue (FR2), (iii) un groupe de « faux débutants » (Hadley, 2002) avec un niveau BAC en anglais n'ayant pas poursuivi d'études supérieures spécialisées dans l'anglais et dont l'optique serait de se 'remettre' à l'anglais dans un but spécifique ou personnel. Ils pourraient être des utilisateurs potentiels d'un logiciel de langues (FR1). En voici la composition numérale :

-GB : 23 locuteurs, 13 femmes et 10 hommes, britanniques anglais

-FR2 : 20 locuteurs, 10 étudiantes et 10 étudiants anglicistes, fin de 2^{ème}/début 3^{ème} année

-FR1 : 20 locuteurs, 10 femmes et 10 hommes en activité ayant niveau Bac anglais L2

Concernant les locuteurs du groupe GB, la moyenne d'âge est de 31 ans. Tous viennent d'Angleterre, et plus précisément, 6 des Midlands, 4 du Nord et 13 du Sud de l'Angleterre. Pour le groupe FR2, tous sont français de langue maternelle française, étudiants en anglais, spécialistes de la langue cible, ayant suivi le cursus universitaire de phonétique à Aix-en-Provence avec un apprentissage de la prosodie et des caractéristiques segmentales ; ils sont soit en fin de 2^{ème} année, soit en début de 3^{ème} année. Aucun n'a effectué de séjours dans un pays anglophone de plus de quatre semaines consécutives, et aucun ne fait partie d'une famille dont l'un des parents ou membres serait anglophone. Tous ont fait en moyenne 9 à 10 ans d'études d'anglais en milieu scolaire. La moyenne d'âge du groupe se situe entre 19 et 22 ans.

Enfin le groupe FR1 se compose d'adultes français de langue maternelle française, d'un niveau d'études post-baccalauréat, exerçant une activité professionnelle. La moyenne d'âge de ce groupe est de 37,5 ans. Aucun ne possède de connaissances approfondies en phonétique anglaise ; tous ont un niveau bac d'anglais et ont suivi un cursus scolaire d'anglais en tant que L2. Aucun n'a effectué de séjour récent de plus d'un mois dans un pays anglophone.

Toutes ces informations ont été recueillies par le biais d'un questionnaire présenté aux sujets en amont des exercices à réaliser. Un consentement de participation et de droit de diffusion des enregistrements a également été présenté et signé par chacun des participants (cf. annexes, section 1). Toutes les informations relatives aux locuteurs (issues de la fiche signalétique) sont synthétisées et répertoriées par groupe et par sexe dans des tableaux présentés en annexe (cf. annexes, section 1).

Codage du corpus

Afin de pouvoir repérer plus facilement les locuteurs pour la suite de ce travail, nous avons codé leur prestation en fonction des tâches effectuées et de leur groupe d'appartenance :

-pour la tâche de lecture (exemple : F02GBP2)

1. La première lettre indique le sexe du locuteur : **F** pour une femme et **H** pour un homme.
2. En deuxième position, le nombre indique le numéro attribué au locuteur (ici, 02)
3. Les deux lettres suivantes indiquent le groupe d'appartenance du locuteur : **GB**, **FR1** ou **FR2** (ici, GB), le numéro indique le niveau (1= faux débutants ; 2= étudiants anglicistes)
4. Enfin le dernier groupe de caractères du codage indique l'exercice effectué : **P** pour passage et **1**, **2**, **3** ou **4** pour le numéro du passage.

-pour la tâche de répétition (exemple : H05FR1phrase01_Britain)

1. La première lettre indique le sexe du locuteur : **F** pour une femme et **H** pour un homme.
2. En deuxième position, le nombre indique le numéro attribué au locuteur (ici, 05)
3. Les deux lettres suivantes indiquent le groupe d'appartenance du locuteur : **GB**, **FR1** ou **FR2** (ici, FR1), le numéro indique le niveau (1= faux débutants ; 2= étudiants anglicistes)
4. Le mot 'phrase' indique que c'est la tâche de répétition de phrases qui est codée

5. Le nombre suivant est le numéro attribué à la phrase (ici, 01)
6. Le mot est un mot clef de la phrase, il ne correspond pas nécessairement au noyau de la phrase (ici, Britain)

-pour la tâche du monologue (exemple : F10FR2_monologue)

1. Les points 1, 2 et 3 précédents restent inchangés
2. La tâche du monologue est indiquée par le mot « monologue » (comme dans l'exemple cité).

Accessibilité du corpus ENGLISH

Le corpus ENGLISH sera en libre accès, à partir de janvier 2010, sur le Centre de Ressources pour la Description de l'Oral (CRDO) <http://crdo.fr>. Le CRDO (Bel & Blache, 2006) est le CRN (Centres de Ressources Numériques) centré sur les ressources orales, initiative conjointe de la Direction de l'Information Scientifique et du Département scientifique « Homme et Société » du CNRS sous la responsabilité des laboratoires LACITO et LPL. Un accès est ouvert à la communauté scientifique et enseignante dans le but que chacun puisse accéder à cette ressource à des fins scientifiques et pédagogiques.

4.4.3 Transcriptions du corpus

Le travail de transcription et d'annotation a été effectué manuellement avec le logiciel PRAAT³³ (Boersma & Weenick, 2009).

Transcription orthographique

La première grande étape concernant l'annotation des enregistrements (une fois le codage des données achevé), porte sur l'encodage orthographique. ENGLISH étant

³³ www.praat.org

composé de trois types de données, nous nous sommes attelée à deux catégories de transcription orthographique que nous avons réalisée manuellement dans Praat.

Pour les phases 1 et 2 (contextes de lecture et répétition), nous avons effectué une transcription orthographique à partir des scripts d'enregistrements utilisant les textes originaux (issus de Eurom 1). Pour la phase 3 (contexte du monologue) la transcription orthographique a été réalisée par un encodage orthographique à partir du contenu du signal.

Pour effectuer ce codage orthographique (phase 3), nous avons établi une convention de notation selon les principes suivants : la ponctuation n'est pas prise en compte mais les phénomènes typiques de l'oral tels que les faux-départs, les pauses pleines, les répétitions, les mots tronqués sont précisés. En voici quelques exemples :

- les mots tronqués sont suivis d'un tiret ; par exemple lors d'une hésitation : la troncation « bene- » pour « benefited »

- les aspects non linguistiques tels que les rires sont marqués entre crochets, par exemple : [laugh]

- les interjections de type 'hésitation' où le locuteur cherche un mot particulier ou cherche à gagner du temps pour réfléchir à la suite de son discours sont marquées 'er' ou 'erm' pour les locuteurs britanniques et 'euh' pour les locuteurs français ;

- les formes contractées ont été retranscrites telles quelles sans modification, c'est-à-dire sans avoir été ramenées à une forme non contractée, la transcription est donc faite en fonction des réalisations effectives, comme le mot 'cos forme contractée du mot « because » dans la phrase « I think just 'cos I met erm a friend ».

Toutes les transcriptions orthographiques des monologues du groupe des natifs anglais ont été vérifiées et attestées par une personne native.

Transcription phonémique

La transcription phonémique a été réalisée manuellement et de manière canonique pour l'ensemble des locuteurs. Cette transcription concerne uniquement la phase de lecture des passages (exercice que nous utilisons pour l'analyse expérimentale de cette recherche).

Il s'agit d'une transcription phonémique (ou transcription phonétique large) en SAMPA³⁴ (Speech Assessment Methods Phonetic Alphabet) ; cette phonémisation manuelle s'effectue sous Praat.

La question du choix de l'alphabet phonétique adopté s'est posée à nous. Même si, avec l'utilisation du phonétiseur « aligner », le SAMPA nous était imposée, nous avons souhaité approfondir la question. Jusqu'ici, l'Alphabet Phonétique Internationale (API) a beaucoup plus souvent été utilisé que le SAMPA mais cela pose un certain nombre de problèmes. Tout d'abord, il requerrait jusqu'à présent, l'utilisation d'une police spécifique ; ensuite, cet alphabet n'est pas toujours reconnu lors de l'utilisation de certains logiciels et notamment dans le domaine du traitement du signal. En revanche, nous exposons deux points positifs qui contribuent à l'utilisation de l'alphabet SAMPA et qui nous confortent dans ce « choix forcé » : le SAMPA étant constitué uniquement de symboles ASCII, il ne requiert aucune installation préalable et s'utilise sur toutes les plates-formes, un simple clavier d'ordinateur est suffisant.

Un autre avantage est la possibilité de convertir très facilement une transcription SAMPA en API.

Ces deux étapes de transcriptions orthographiques et phonémiques achevées, nous procédons à l'alignement manuel de la transcription orthographique préalablement effectuée.

Alignement acoustique manuel

Afin d'obtenir une segmentation manuelle la plus homogène possible sur l'ensemble du corpus utilisé, des critères traditionnels ont été adoptés (Peterson & Lehiste, 1960). Les mesures ont été obtenues à l'aide de l'inspection visuelle du spectrogramme et de la forme de l'onde.

³⁴ Description disponible sur le site <http://www.phon.uc.ac.uk/home/sampa/home.htm>

A l'heure actuelle, chaque passage de lecture a été segmenté en unités intonatives (UI), en mots (mot), en syllabes (syllabe), en unités rythmiques (NRU), basées sur le modèle de Jassem (1952), et en segments (CVC) codés C (pour consonnes) et V (pour voyelles), pour l'ensemble des locuteurs de sexe féminin. Les passages de lecture des groupes de locuteurs de sexe masculin ont été alignés en unités intonatives (UI) et en phonèmes (CVC) selon le codage C/V énoncé ci-dessus. Le niveau utilisé pour les analyses expérimentales présentées dans le chapitre 5 repose sur le niveau de la segmentation en phonèmes (CVC).

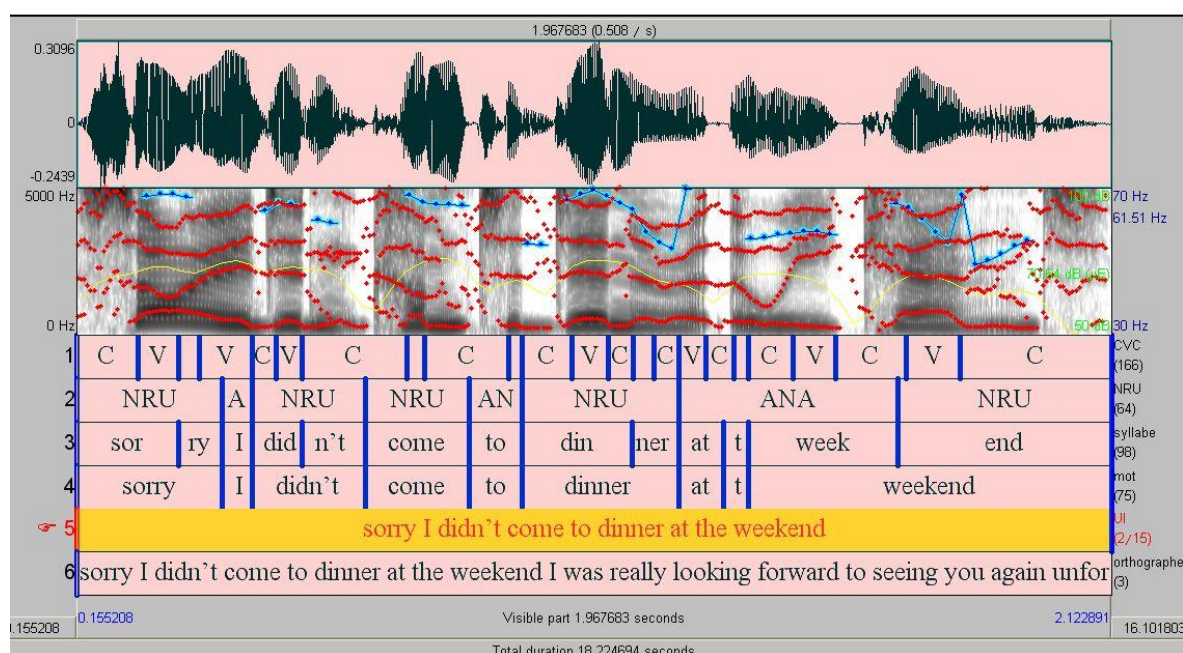


Figure 14. Exemple de TextGrid d'annotation extrait d'ENGLISH effectué dans PRAAT

Concernant le niveau des unités syllabiques, qui constitue le premier niveau supérieur à celui du phonème, la syllabe a été déterminée en fonction du Principe d'Attaque Maximale (Pulgram, 1970) selon lequel la plus grande séquence de consonnes obéissant aux règles phonotactiques de l'anglais est systématiquement préférée en tant qu'attaque syllabique. Ainsi un mot comme « inspector » /Insˈpekt@/ (passage 1) analysé au niveau de la syllabe accentuée, aura pour noyau /e/ et regroupe en attaque la séquence maximale /sp/, la séquence supérieure /nsp/ n'est pas licite.

Le découpage en unités rythmiques est effectué selon le modèle rythmique proposé par Jassem (1952) car nous pensons qu'il présente une structuration rythmique intéressante ; par ailleurs des travaux (Bouzon, 2004) ont montré que le modèle de Jassem

rendait mieux compte de la structuration rythmique de l'anglais que le modèle classique d'Abercrombie/Halliday (1967). La structuration rythmique (selon Jassem) est divisée en « unités rythmiques étroites » (NRU) comprenant la syllabe accentuée du mot suivie de toutes les syllabes du mot et en anacrouses (ANA) qui comprennent toutes les syllabes inaccentuées non incluses dans la NRU. Le niveau des unités rythmiques constitue un niveau supplémentaire d'analyse possible pour diverses perspectives de recherches dont nous ne discutons pas plus ici, ne faisant pas l'objet de ce travail de thèse.

Au niveau des critères de segmentation, le début de chaque phrase a été aligné avant le premier pic d'amplitude dépassant le bruit de fond, alors que sa fin a été établie à la dernière pulsation glottale de f_0 et non à la fin de la courbe d'amplitude car une région de voix soufflée était souvent présente à la fin d'un énoncé. Les variations soudaines de l'amplitude et de l'intensité ainsi que les transitions formantiques ont été utilisées comme critères de segmentation. Les frontières entre consonnes et voyelles ont été marquées à l'apparition de la première période vocalique prenant en compte la stabilité des formants $F1$ & $F2$.

Les pauses, qu'elles soient pleines ou silencieuses, sont notées par le symbole « # ». Les mots ou portions de texte sur lesquels le locuteur a bafouillé, mais qui ont été repris par le sujet, ont été isolés par des frontières à droite et à gauche et ont été marqués également par le symbole « # ».

4.5 Perspectives

Plusieurs perspectives majeures sont envisagées dans le cadre de l'amélioration d'ENGLISH. Nous retenons particulièrement deux progressions au niveau (i) de la transcription phonémique, (ii) au niveau de l'alignement.

Progression de l'annotation phonémique

La transcription phonémique est effectuée à ce jour sur la partie lecture. Il serait important de continuer à la réaliser sur l'ensemble du corpus. Par ailleurs, la transcription proposée est une transcription canonique, il serait donc intéressant de la reprendre et d'envisager une transcription individuelle et propre à chacun des locuteurs.

Progression de l'alignement

Le développement de l'alignement est envisagé tout d'abord sur la partie lecture au niveau des locuteurs de sexe masculin, ce qui permettra de compléter la segmentation actuelle (UI et CVC) par un alignement au niveau du mot, de la syllabe et de l'unité rythmique. De plus, nous envisageons l'alignement des autres tâches proposées dans ANGLISH et particulièrement, celle du monologue de manière à pouvoir appliquer les critères d'évaluation proposés sur de la parole spontanée. Cela laisse envisager de nombreuses heures de segmentation.

Synthèse

Ce chapitre nous a permis de présenter la base de données ANGLISH, qui constitue la réalisation du premier objectif fixé pour notre travail de thèse. Nous avons présenté le protocole d'expérimentation ainsi que les diverses potentialités du corpus.

Actuellement, ANGLISH comporte environ 5H30 de parole en anglais, langue maternelle (L1) par des natifs anglais britanniques et langue cible (L2) par des francophones de deux niveaux différents en anglais. Les productions enregistrées se répartissent sur trois tâches différentes (lecture, répétition, monologue) et comptent 63 locuteurs. ANGLISH constitue une ressource importante pour une analyse comparative de l'anglais produit par des francophones et des anglophones.

Chaque tâche a été transcrite orthographiquement ; plusieurs rangées d'alignement sont proposées pour la tâche de lecture depuis le phonème jusqu'à l'unité intonative en passant par les mots, syllabes et unités rythmiques. Concernant l'exploitation plus particulière d'ANGGLISH, dans le cadre de l'élaboration de critères rythmiques basés sur l'analyse des productions des francophones, nous retenons l'étape d'alignement en CVC dans la mesure où elle a permis l'exploitation de durées temporelles auxquelles nous avons eu recours dans l'étude décrite dans le chapitre suivant.

La base de données ANGLISH est disponible gratuitement sur le site du CRDO <http://crdo.fr>, permettant ainsi une large diffusion, ainsi que, nous l'espérons, un large retour qui contribuera à l'enrichissement de la base de données.

Présentement, nous allons nous tourner vers l'exploitation d'ANGGLISH afin de tester les critères rythmiques évaluatifs possibles des productions de locuteurs francophones, au travers des hypothèses formulées dans le chapitre suivant, ce qui constitue notre second objectif.

CHAPITRE 5. ANALYSES EXPERIMENTALES



Dans ce chapitre, nous proposons tout d'abord un récapitulatif de la problématique soulevée dans le chapitre 3 concernant les métriques rythmiques et une présentation succincte de l'ensemble des différentes hypothèses qui en découlent. Dans une deuxième partie, nous cherchons à corroborer ou infirmer ces hypothèses par le biais d'expériences que nous exposons ci-après.

5.1 Problématique et objectifs

Nous avons soulevé dans notre introduction et constaté dans le chapitre 3, le manque de critères prosodiques évaluatifs dans l'apprentissage d'une langue seconde. A l'heure actuelle, où nous assistons à un véritable essor de la création de nouvelles technologies éducatives, l'absence d'une évaluation objective se fait ressentir. En effet, de manière générale, la seule évaluation possible est un feedback audio-visuel, pas toujours accessible pour l'utilisateur. Pourtant, nous avons vu que ces nouvelles technologies pourraient entre autres, (i) accroître la qualité de l'enseignement phonétique (en effet, il est vrai que l'approche communicative proposée actuellement basait essentiellement la communication sur l'apprentissage des contenus), et (ii) constituer un apport intéressant pour l'apprentissage de la prosodie : nous avons mis en avant que les éléments prosodiques constituaient la base de la communication orale (Di Cristo, 2002), mais aussi qu'ils étaient les éléments les plus difficiles à acquérir pour l'apprenant, pour les diverses raisons exposées dans le chapitre 2. De ces constats, nous pensons qu'il serait intéressant de chercher à compléter ces modules d'évaluation.

Par ailleurs, nous avons vu au travers de la littérature, que l'apprentissage du rythme anglais était une tâche difficile pour les francophones (Adams, 1979). Cette difficulté reposait, entre autres, sur le fait que ces deux langues présentaient des structures rythmiques très différentes (Delattre, 1966 ;Wenk, 1985 ;Bolinger, 1981):

- réduction vocalique,
- structure syllabique,
- accentuation lexicale,
- classification oppositionnelle (etc.).

Ces différences, nous l'avons vu, ont pu être mises en évidence par l'élaboration de métriques rythmiques, dont nous proposons ici un bref récapitulatif :

	% V ³⁵	ΔC, ΔV		rPVI (c,v)		nPVI (c,v)		cvC,cvV	
	Ramus & al. (1999)			Grabe & Low (2002)				Wagner & Dellwo (2004)/ White & Mattys (2007)	
	% durée des segments vocaliques	écart type de la durée des segments cons. & voc.		indice brut de variabilité entre la durée des intervalles cons. & voc. successifs		indice normalisé de variabilité entre la durée des intervalles cons. et voc. successifs		coefficient de variation de la durée des segments cons. & voc.	
FR (L1)	43.6	4.39	3.78	50.4	--	--	43.5	44	50
GB (L1)	40.1	5.35	4.64	64.1	--	--	57.2	47	64

Tableau 5. Récapitulation des différentes métriques utilisées pour notre étude

Parce-que ces métriques permettent de distinguer des langues (Ramus & al., 1999 ; Grabe & Low, 2002 ; Dellwo & Wagner, 2003 ; Lin & Wang, 2005 ; White & Mattys, 2007), des dialectes (Fernandez-Cruz, 2000 ; Ferragne & Pellegrino, 2004), nous émettons l'hypothèse qu'elles permettent de distinguer la production rythmique d'apprenants, des productions orales de natifs. En effet, plusieurs études se sont penchées sur une distinction L1/L2 (Whitworth, 2002 ; Carter, 2005 ; White & Mattys, 2007 ; Grenon & White, 2008) et ont montré qu'il était possible de distinguer la langue maternelle de la langue cible, notamment en utilisant le paramètre de coefficient de variation. De fait, si ces métriques assurent une telle distinction, elles pourraient rendre compte de l'influence de la L1 (le français) sur la langue cible (l'anglais) et permettraient ainsi de différencier la production orale d'anglophones et de francophones parlant anglais. Le but sera d'utiliser ces métriques

³⁵ % C est iso-morphique au % V, il n'est donc pas nécessaire de le prendre en considération

comme critères évaluatifs du rythme dans l'apprentissage d'une langue seconde en explorant la meilleure combinaison de paramètres possibles.

Dans ce chapitre, donc, nous proposons de comparer la structure rythmique de locuteurs anglophones et d'apprenants français à travers les 9 paramètres rythmiques présentés ci-dessus ainsi que les différentes combinaisons qui se sont révélées pertinentes dans les travaux antérieurs.

L'étape suivante est donc d'examiner les performances prosodiques de différents niveaux d'apprenants par méthode comparative avec des productions de natifs. Le but, nous le rappelons, est de déterminer les déviations possibles pour en dégager des critères rythmiques objectifs et évaluatifs, qui pourraient être intégrés à des méthodes d'apprentissage, telles que les logiciels multimédia, voire également, d'être examinés dans le cadre d'études sur l'évaluation en pathologie ou encore en parole de synthèse.

5.2 Méthodologie

5.2.1 Base de données

Afin de mener à bien cette étude, nous avons travaillé sur la partie lecture du corpus ANGLISH (présentée dans le précédent chapitre). Il s'agit donc de productions oralisées, de textes lus par des locuteurs anglophones et des apprenants francophones de différents niveaux.

Nous rappelons ici les modalités de la tâche de lecture ainsi que la composition des trois groupes de locuteurs.

Le corpus de lecture ANGLISH est composé de 4 passages (cf. annexes, section 1) issus du corpus EUROM 1 (P9, Q0, Q9, R3), lus par 63 locuteurs, ce qui représente un total de 252 passages (soit 1260 phrases), soit un total d'une durée approximative d'une heure et demie de lecture (soit 4765,3138 secondes).

Chacun des passages se compose de 5 phrases liées sémantiquement, racontant un évènement, une histoire de la vie quotidienne. La consigne donnée à la fois oralement et à l'écrit était de : « *Lire les passages comme si l'histoire vous était arrivée et que vous la racontiez à un ami* ». Les passages étaient donnés à l'avance, de manière à ce que les

locuteurs se familiarisent avec les textes et puissent éviter les bafouillages lors de l'enregistrement.

Concernant les locuteurs, ces derniers sont répartis en trois groupes, constitués en fonction de critères précis. Ils sont au nombre de 63 (cf. annexes, section 1 pour une description détaillée de l'identité de chaque locuteur).

-GB : 23 locuteurs, 13 femmes et 10 hommes, britanniques anglais

-FR2 : 20 locuteurs, 10 étudiantes et 10 étudiants anglicistes, fin de 2^{ème}/début 3^{ème} année

-FR1 : 20 locuteurs, 10 femmes et 10 hommes en activité ayant niveau Bac anglais L2

A partir des fichiers son, nous avons effectué un travail de transcription et d'annotation manuelles avec le logiciel PRAAT (Boersma & Weenick, 2009). Nous ne revenons pas à nouveau sur ce point, puisqu'il est développé dans le chapitre 4, mais nous rappelons brièvement les « rangées » obtenues et annotées dans un fichier TextGrid, suite à ce travail³⁶ :

-la rangée orthographique, non ponctuée, nommée « orthographe »

-la rangée SAMPA, qui correspond à une transcription phonémique canonique en sampa, nommé « sampa » (cf. annexes, section 2 pour une liste des symboles et équivalence API et section 1, pour la transcription des 4 passages étudiés)

-la rangée des unités intonatives dont chaque pause est notée '#', nommé UI

-la rangée des phonèmes, nommée CVC (cf. figure 15), dont la segmentation suit le codage suivant :

- Les segments vocaliques sont notés : V
- Les segments consonantiques sont notés : C
- Les pauses inter-unités intonatives sont notées : #
- Les pauses/hésitations intra-unités intonatives sont notées : P

³⁶ Les textgrids des locuteurs de sexe féminin sont composés de 7 rangées de segmentation (orthographe, sampa, UI, mots, syllabes, NRU, CVC)

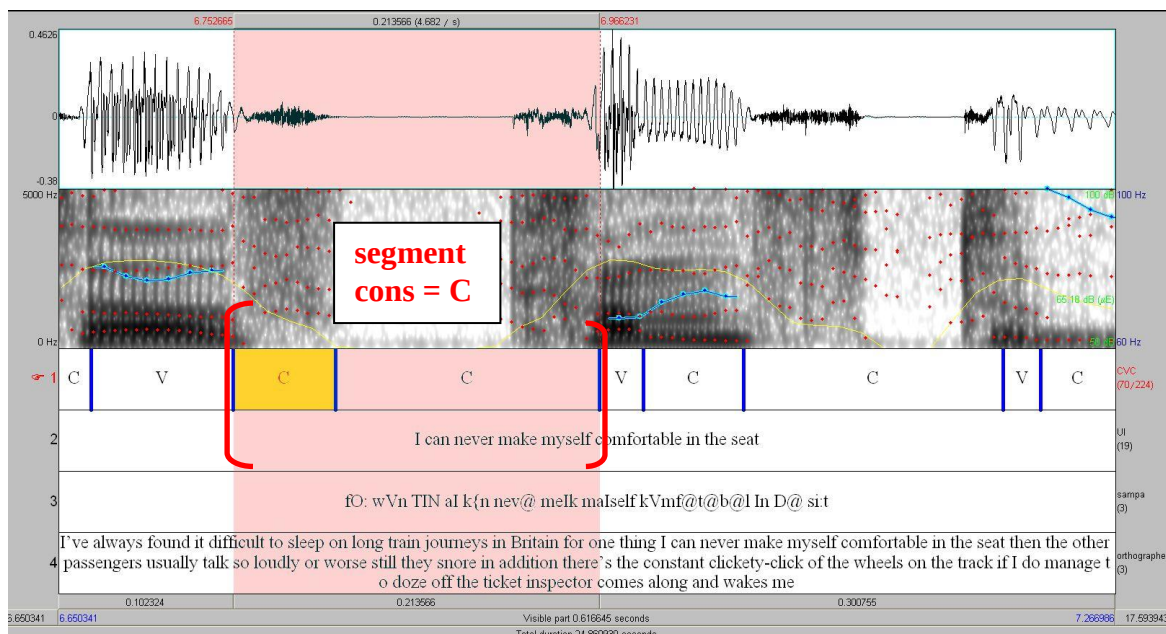


Figure 15. Exemple de segmentation en CVC

Dans notre expérience, c'est à partir de la segmentation en CVC, effectuée sur l'ensemble du corpus de lecture, que nous menons notre étude comparative des différentes métriques proposées dans la littérature, et ce, pour chacun des groupes établis.

5.2.2 Calcul et obtention des métriques

Les métriques rythmiques ont été calculées pour chacune des 5 phrases des 4 passages enregistrés par les 63 locuteurs, soit un total de 1260 données pour chacune des métriques.

A partir d'un script *Praat*, élaboré par Daniel Hirst (cf. annexes, section 2), les métriques sont calculées automatiquement à partir d'un dossier dans lequel sont réunis l'ensemble des fichiers TextGrid. Chaque paramètre est calculé en fonction des intervalles de durées (en millisecondes) des segments consonantiques et vocaliques. Pour chaque locuteur, nous obtenons 9 paramètres rythmiques :

%V	la somme de la durée des intervalles vocaliques divisée par la durée totale des intervalles vocaliques et consonantiques, multipliée par 100
ΔC	écart type des durées des intervalles consonantiques
ΔV	écart type des durées des intervalles vocaliques
rPVI_C	moyenne des différences de durées entre les intervalles consonantiques successifs (brut)
nPVI_C	moyenne des différences de durées entre les intervalles consonantiques successifs divisée par la moyenne de ces mêmes intervalles et multipliée par 100 (normalisé)
rPVI_V	moyenne des différences de durées entre les intervalles vocaliques successifs (brut)
nPVI_V	moyenne des différences de durées entre les intervalles vocaliques successifs divisée par la moyenne de ces mêmes intervalles et multipliée par 100 (normalisé)
cvV	(coefficient de variation=) écart type des durées des intervalles vocaliques divisé par la moyenne des durées des intervalles vocaliques
cvC	(coefficient de variation=) écart type des durées des intervalles consonantiques divisé par la moyenne des durées des intervalles consonantiques

Tableau 6. Méthode de calcul des 9 paramètres rythmiques

Les données de sortie sont obtenues, sous format tabulaire, dans une fenêtre *Praat Info* (représentée ci-après) pour laquelle chaque colonne correspond à :

- « **passage** » : indique le passage étudié pour lequel les métriques sont calculées,
- « **type** » : renseigne si le locuteur est un anglophone (GB), un francophone niveau 1 (FR1) ou 2 (FR2),
- « **speaker** » : identifie le locuteur,
- « **gender** » : informe sur le sexe du locuteur,
- « **IU** » : indique la position de l'unité intonative dans le passage,
- **les colonnes 6 à 14** : donnent les résultats pour chaque paramètre rythmique calculé (détaillé dans le tableau 6).

Praat: Info							
File Edit Search							
passage	type	speaker	gender	IU	percentV	sdC	sdV
P1	FR1	01	F	1	41.066318909686494	0.04982742661134698	0.035768
P1	FR1	01	F	2	38.575967638366166	0.04248710354696088	0.036579
P1	FR1	01	F	3	46.24493292838728	0.0665917214079862	0.052057
P1	FR1	01	F	4	37.7335689939712	0.045280392374372425	0.026203
P1	FR1	01	F	5	44.89487290168739	0.06332900652046533	0.068378
P2	FR1	01	F	1	50.635547378966436	0.04782111431688027	0.044042
P2	FR1	01	F	2	47.786440785778964	0.0573468908395763	0.041628
P2	FR1	01	F	3	42.01198074285268	0.05277636163490621	0.052933
P2	FR1	01	F	4	38.84193784097075	0.09008692944716125	0.041610
P2	FR1	01	F	5	38.826221826616	0.10372534104704861	0.044931
P3	FR1	01	F	1	42.12609802491494	0.07879506082471571	0.050170
P3	FR1	01	F	2	43.155517534022614	0.06330505710577358	0.071268
P3	FR1	01	F	3	39.364794661523796	0.056134232829013836	0.033028
P3	FR1	01	F	4	42.68289347333604	0.06407781401347216	0.039580
P3	FR1	01	F	5	34.77165600248734	0.09266286874357818	0.037794
P4	FR1	01	F	1	42.48663701779169	0.09785401051517519	0.056588
P4	FR1	01	F	2	41.829013457720066	0.04827742639536646	0.038088

Tableau 7. Extrait des données de sortie obtenues par le script "calculate_rhythm_metrics"

5.3 Analyses statistiques

5.3.1 Caractérisation des groupes de 3 niveaux

Nous procédons tout d'abord à une **analyse en composantes principales** (ACP)³⁷, afin d'étudier les ressemblances entre les individus du point de vue de l'ensemble des variables et ainsi apprécier une classification en trois groupes selon un ensemble de paramètres indépendants. L'ACP permet en effet la transformation de nos variables corrélées (ici les 9 paramètres rythmiques) en variables indépendantes, et réduit le nombre de composantes. Elle permet ainsi la caractérisation des énoncés ou de groupes d'énoncés (FR1, FR2 et GB) par les variables (cf. script en annexes, section 2).

A partir de notre objet de composantes principales nouvellement créé, nous effectuons ensuite une **analyse discriminante** (homoscédasticité ou paramétrique linéaire) afin de voir s'il est possible de déterminer, de manière fiable, le groupe d'appartenance des locuteurs (FR1, FR2 ou GB). La figure montre une certaine classification où les locuteurs

³⁷ Les analyses sont effectuées à partir de l'environnement de travail R (R development Core Team, 2008).

du groupe FR1 (en vert) se situent, sur l'axe des abscisses, entre [-4 et -2], ceux du groupe FR2 (en bleu), entre [-2 et 1] et ceux du groupe GB (en rouge) entre [1 et 4]. Au vu de ce graphique, nous pouvons donc dire qu'il semble possible, à partir de critères rythmiques, de distinguer les trois groupes.

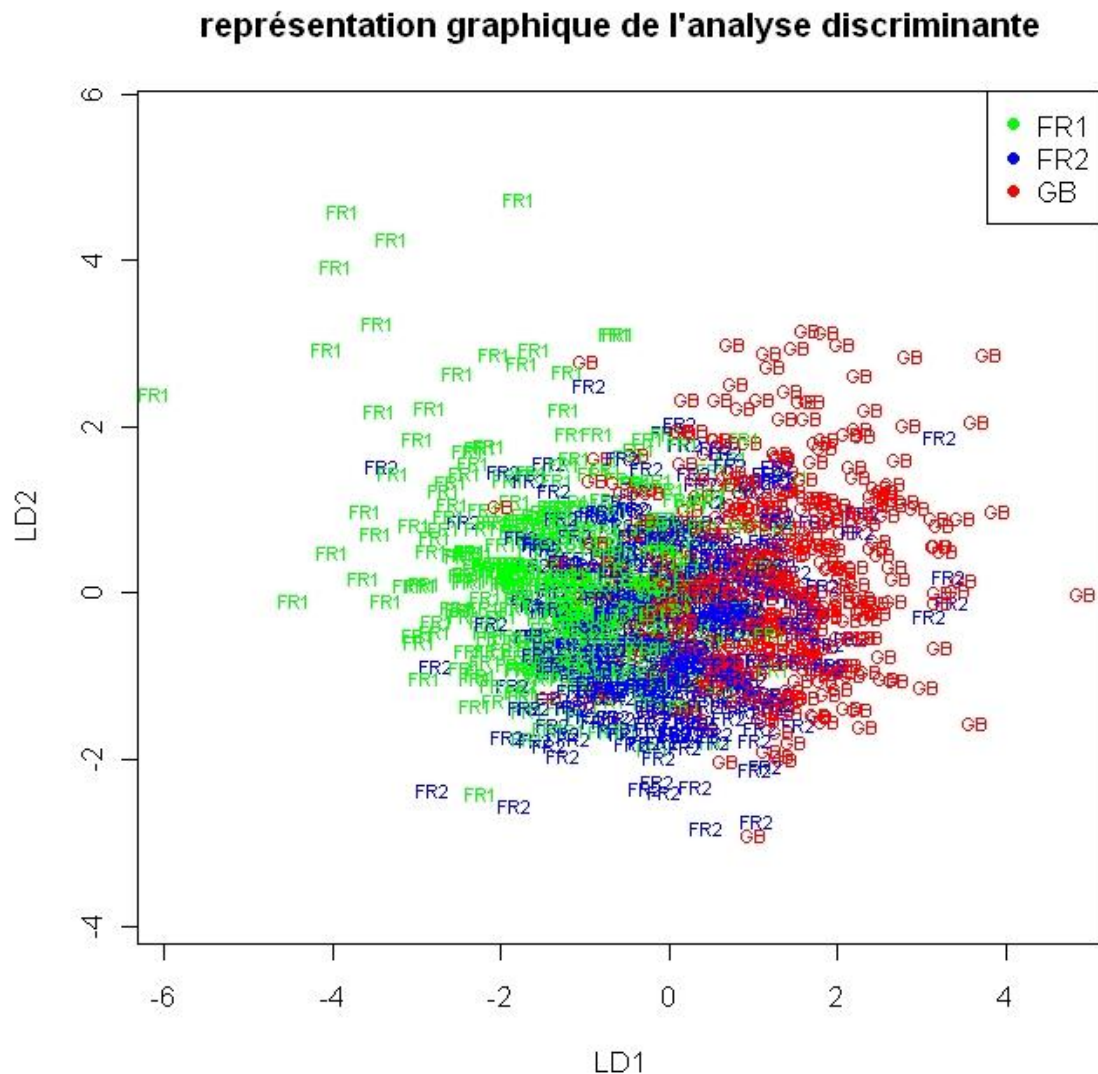


Figure 16. Représentation graphique de l'analyse discriminante des 1260 phrases produites par les 63 locuteurs

Dans cette procédure, chaque phrase est classifiée à l'aide de fonctions discriminantes calculées sur l'ensemble des données. Ici, nous utilisons l'ensemble des variables, c'est-à-dire les 9 paramètres rythmiques.

Classe réelle	Classe prédite		
	FR1	FR2	GB
FR1	274	100	23
FR2	97	192	93
GB	29	108	344

Classe réelle (en%)	Classe prédite		
	FR1	FR2	GB
FR1	69%	25%	6%
FR2	25%	50%	25%
GB	6%	22%	72%

Tableau 8. Présentation des résultats des prédictions des différentes classes (FR1/FR2/GB), (en haut, valeurs brutes ; en bas, valeurs exprimées en pourcentage)

Sur l'ensemble des groupes, les deux groupes opposés (FR1 et GB) se distinguent particulièrement bien ; les prédictions s'élevant à 72% pour le groupe GB sont positives. En effet, seulement 6% des locuteurs de ce groupe sont classés en FR1, ce qui, sur l'ensemble des données, ne représenterait qu'un seul locuteur. On pourrait l'expliquer par le fait que certaines phrases ont peu de consonnes et seraient donc plus proches d'une structure rythmique isosyllabique (donc du français).

Le second groupe (FR1) est également bien discriminé avec 69%. Tout comme les GB, seulement 6% des locuteurs sont mal classés. Le fait que 25% des locuteurs soient classés en FR2 ne semble pas si étonnant dans la mesure où certains locuteurs pourraient avoir un niveau d'anglais légèrement supérieur au niveau faible annoncé des FR1.

La discrimination du groupe FR2 s'effectue sur la base des 50% ; les 50 autres pour cent des apprenants, sont répartis également de part et d'autre, c'est-à-dire, soit en FR1, soit en FR2. Ces résultats sont très intéressants car ils montrent bien le niveau à « mi chemin » des étudiants, entre ceux dont le niveau s'écarte particulièrement des GB ou ceux, au contraire qui y sont très proches. Les prédictions semblent donc assez fidèles à la réalité composite du groupe et justifient qu'elles n'atteignent que 50%.

Il semble donc possible, à partir des 9 critères rythmiques, de prédire avec une relativement bonne fiabilité, le niveau de langue ou de groupe auquel correspond la phrase analysée (69,50% au global). L'analyse ANOVA révèle d'ailleurs une corrélation significative entre les composantes principales et le groupe ($F(2,1257)=69,503$; $pval<2,2 \times 10^{-16}$).

5.3.2 Analyses des métriques rythmiques : distinction L1/L2

A partir des résultats obtenus, nous proposons, en amont d'une analyse statistique, une représentation graphique de la moyenne (représentée par le point jaune) et de l'erreur-type (standard error) calculées pour chacune des variables (c'est-à-dire chaque paramètre métrique) en fonction du groupe (GB, FR1 et FR2), afin d'évaluer de possibles différences significatives entre les niveaux de langues selon des critères rythmiques.

L'erreur-type est l'écart-type de l'erreur de cette estimation ; elle est obtenue par le calcul de l'écart-type divisé par la racine carrée du nombre d'observations : $\sigma/\sqrt{n}$

Ce type de représentation graphique présente un intérêt visuel particulier dans la mesure où, lorsque deux moyennes sont éloignées de plus de 2 erreurs-types, les différences sont significatives. Visuellement les barres se chevauchent lorsque les moyennes ne sont pas significativement distinctes (pour une illustration de chevauchement, cf. figure 21, plan cvC).

Dans cette section, nous présentons de manière chronologique, les résultats du calcul de l'erreur-type pour les métriques 'dites traditionnelles', c'est-à-dire les plus utilisées dans la littérature, à savoir Ramus & al., Grabe & Low, White & Mattys et nous testons également les propositions des meilleures combinaisons avancées par ces auteurs, à savoir %V- Δ C (Ramus & al.) ; nPVI_V,rPVI_C (Grabe & Low) ; %V,cvC (Dellwo) ; %V,cvV (White & Mattys). Les données se présentent comme suit : à gauche, une analyse de dispersion des points en fonction des 3 groupes et de la combinaison avancée, à droite, la représentation graphique de l'erreur-standard selon un agrandissement de la partie du graphe de dispersion.

Nous rappelons que nous cherchons à définir des critères rythmiques évaluatifs et nous pensons que les métriques pourraient jouer un rôle important dans la discrimination

L1/L2, et permettraient de donner une tendance rythmique de la production des apprenants français. Jusqu'ici, aucune étude, à notre connaissance, n'avait mené de comparaison évaluative de l'efficacité de ces métriques dans la discrimination L1/L2, afin d'observer l'influence du rythme de la L1 (français) sur les productions rythmiques de nos apprenants en L2 (anglais).

De manière générale, quelle que soit la combinaison annoncée, nous pensons que si les productions des apprenants FR2 sont rythmiquement correctes, les valeurs obtenues seront proches de celles des natifs. En revanche, nous nous attendons à ce que le groupe FR1, supposément d'un niveau faible, soit clairement distingué des natifs et des FR2, par des résultats beaucoup plus éloignés de ceux des deux autres groupes.

$$\Delta V, \Delta C, \%V \text{ (Ramus \& al., 1999)}$$

*Rappel des calculs de ces différents paramètres :

%V : la somme de la durée des intervalles vocaliques divisée par la durée totale des intervalles vocaliques et consonantiques, multipliée par 100

ΔC : écart type des intervalles de durées des segments consonantiques

ΔV : écart type des intervalles de durées des segments vocaliques

*hypothèses :

La structure syllabique rythmique de l'anglais, nous l'avons dit, est complexe. Suivant les principes théoriques de Ramus *et al.*, nous faisons donc l'hypothèse que le taux de ΔC des anglais sera plus élevé que celui des apprenants de faible niveau, et inversement, que le taux de ΔV , sera plus élevé chez les FR1, et plus faible chez les natifs, le groupe des FR2 se situant à mi-chemin.

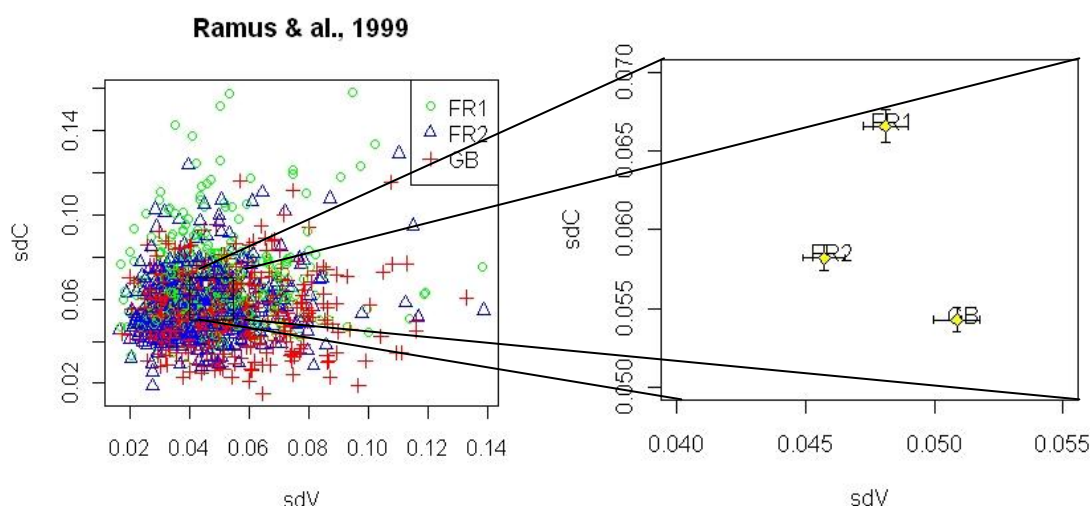


Figure 17. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan (ΔV , ΔC)

Au vu de ce graphique, le ΔC est plus élevé chez les FR1 et plus faible chez les GB, le ΔC des FR2 se situant entre celui des FR1 et des GB. Les résultats infirment donc notre première hypothèse ($F(2,1257)=54,74$; $pval < 2 \cdot 10^{-16}$). Quant au ΔV , ce sont les FR2 qui ont le taux le plus bas et les GB le plus haut graphiquement. Cependant, la différence n'est pas rapportée significative ($F(2,1257)=8,827$; $pval=0,02$).

Bien que nous obtenions des résultats contraires à ceux attendus, le plan (ΔV , ΔC) (figure 17) montre cependant une distinction assez nette des groupes FR1 et FR2, lorsque le paramètre ΔC est considéré séparément, ou en combinaison linéaire avec le paramètre ΔV ($F(2,1257)=15,21$; $pval=2,958 \cdot 10^{-7}$). Dans ce graphique, il est à noter que la production des francophones FR2 s'apparente plus à celle des anglophones. Le paramètre ΔV ne permet pas une distinction des trois groupes ($pval=0,02$). La figure de gauche montre d'ailleurs qu'il est difficile de distinguer notamment, même graphiquement, les groupes FR1 & FR2. Nous pouvons donc dire, à partir de ces premiers résultats graphiques, que ΔC ou la combinaison de ΔC et ΔV pourraient permettre une catégorisation des groupes d'apprenants FR1 et FR2 et du groupe des anglophones, alors que ΔV uniquement ne le permet pas.

Dans leur étude, Ramus & al. montrent également que la combinaison % V- Δ C permet de distinguer les locuteurs français des locuteurs anglais. Les auteurs rapportent un score plus élevé pour les français. Nous formulons donc l'hypothèse que, à la fois le %V sera le plus élevé pour FR1 que pour GB (FR2 se situant toujours à mi-chemin entre les deux) et que la combinaison %V- Δ C est une bonne combinaison permettant la discrimination de nos trois groupes.

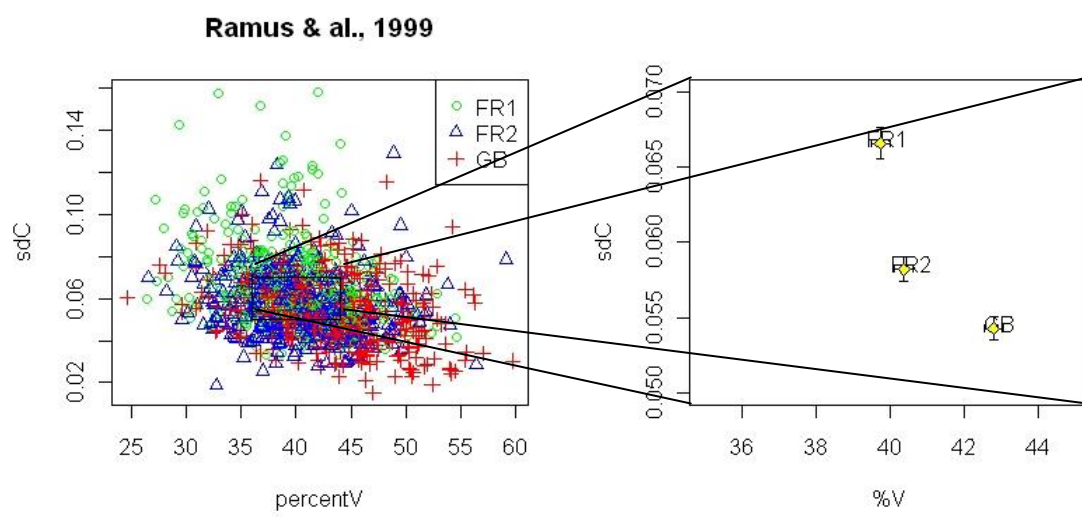


Figure 18. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan (%V, Δ C)

Les représentations graphiques ci-dessus infirment notre première hypothèse ($F(2,1257)=25,87$; $pval= 9,746 \times 10^{-12}$). Cependant, à partir du graphe de dispersion et à partir de la représentation des erreurs-standards, il est quand même possible de distinguer les natifs des apprenants, si l'on considère le paramètre %V uniquement. Le niveau des apprenants ne peut être en revanche distingué. D'ailleurs, la corrélation entre FR2 et %V est non significative ($pval=0,0846$). Nous pouvons donc conclure, au vu de ces représentations graphiques, que %V permet de distinguer les anglophones des francophones, que Δ C permet plutôt la discrimination des niveaux des apprenants, ce qui est également le cas pour la combinaison linéaire %V- Δ C.

- *PVI (Grabe & Low, 2002)*

*rappel des calculs des paramètres :

rPVI_C/rPVI_V : indice brut de variabilité des durées des intervalles consonantiques / vocaliques successifs

nPVI_C/nPVI_V : indice normalisé de variabilité des durées des intervalles consonantiques / vocaliques successifs

*hypothèses :

Selon l'étude de Grabe & Low, le PVI normalisé permet de « corriger » les fluctuations de tempo. Cette mesure apparaît donc également intéressante ici. Les auteurs ont rapporté que les taux de nPVI_V et rPVI_C étaient élevés pour les Anglais et faibles pour les Français. Nous formulons donc l'hypothèse selon laquelle les taux de PVI_V et de PVI_C des natifs seront plus élevés que ceux des FR2, les taux des FR1 étant les plus faibles. En effet, si les réductions vocaliques sont correctement produites, alors les différences de variations vocaliques seront plus importantes (caractéristiques des langues isoaccentuelles).

Pour résumer, plus la production rythmique de l'apprenant est proche de celle d'un natif, plus le taux de nPVI_V se rapproche de celui des natifs et obtient une valeur haute.

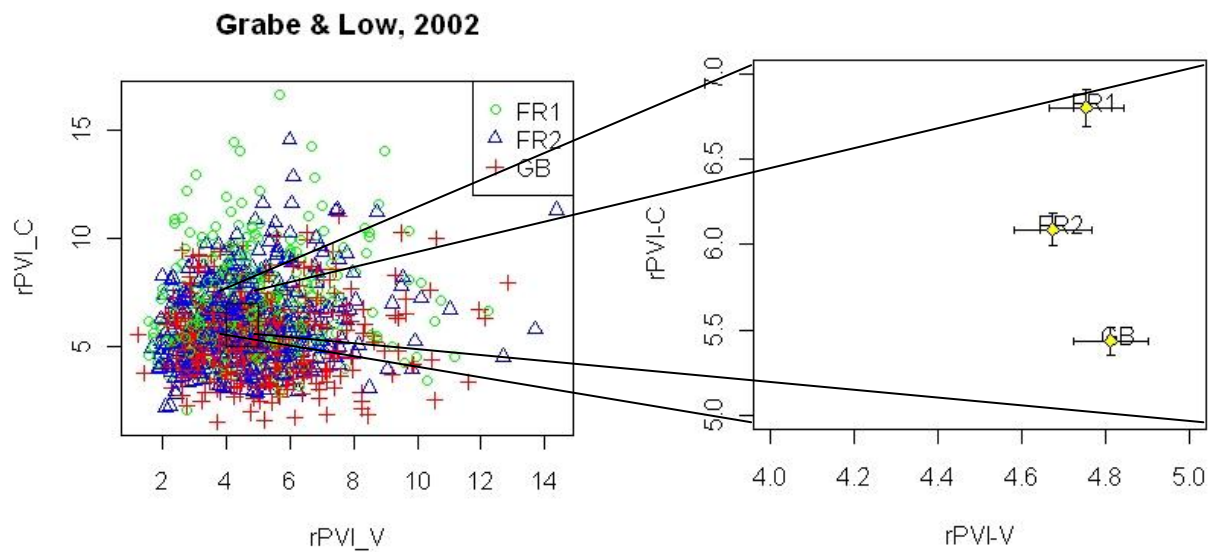


Figure 19. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan (rPVI_V, rPVI_C)

Sur le plan (rPVI_V, rPVI_C), si l'on considère le rPVI_V, il n'y a pas de différence significative entre les différents groupes ($F(2,1257)=0,5938$; $pval= 0,5524$). Si l'on considère en revanche le rPVI_C ($F(2,1257)=54,75$; $pval<2,2 \times 10^{-16}$) ou la combinaison du rPVI_V avec le rPVI_C ($F(2,1257)=16,89$; $pval= 5,797 \times 10^{-8}$), on observe un taux plus élevé chez les FR1 que chez les GB, des résultats qui infirment ceux de la littérature. Si ces représentations graphiques ne corroborent pas notre hypothèse, il apparaît cependant que le rPVI_C ou la combinaison rPVI_V - rPVI_C pourraient être de bons critères de discrimination des groupes FR1, FR2 et GB.

Les prochains graphes présentent les métriques du PVI calculé de manière normalisée.

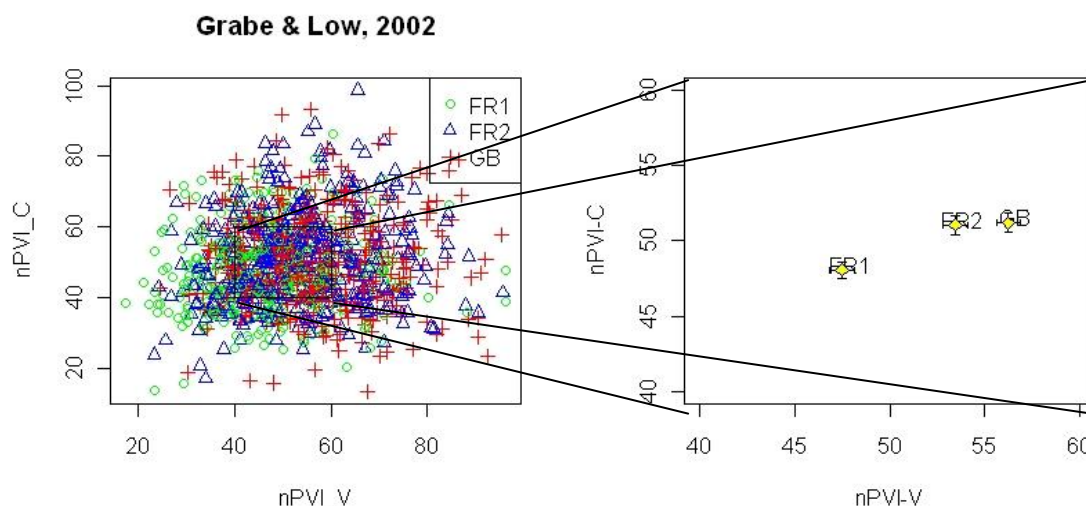


Figure 20. Visualisation des points de dispersion (gauche) et de erreurs-types (droite) selon le plan (nPVI_V,nPVI_C)

Le plan (nPVI_V, nPVI_C) corrobore ici notre première hypothèse. Si l'on considère nPVI_V, on observe que le taux est le plus élevé pour les GB, suivi de celui des FR2, le taux des FR1 étant le plus bas ($F(2,1257)=50,13$; $pval < 2 \cdot 10^{-16}$). Si l'on considère à présent le nPVI_C, il n'y a pas de différence significative entre les groupes FR2 et GB ($F(2,1257)=8,326$; $pval = 0,00025$). Cependant, l'hypothèse est toujours valide, dans le sens où les FR1 ont toujours le taux le plus bas. Similairement au plan (ΔV , ΔC), on observe donc à partir du paramètre nPVI_V et de la combinaison linéaire des paramètres nPVI_V et nPVI_C ($F(2,1257)=41,53$; $pval < 2 \cdot 10^{-16}$) une nette séparation des groupes FR1 et FR2, le groupe FR2 s'approchant graphiquement du groupe GB. A partir du paramètre nPVI_C, on observe uniquement une séparation entre les apprenants de premier niveau et les deux autres groupes. Nous pouvons donc conclure que les paramètres pris séparément ou leur combinaison pourraient permettre la classification des groupes en fonction du niveau (FR1 vs FR2) mais non une distinction en fonction du critère natif vs. non natif.

Selon Grabe & Low, la combinaison qui permet une meilleure discrimination entre les langues, est celle du nPVI_V, rPVI_C, combinaison dont nous commentons les résultats sur nos données.

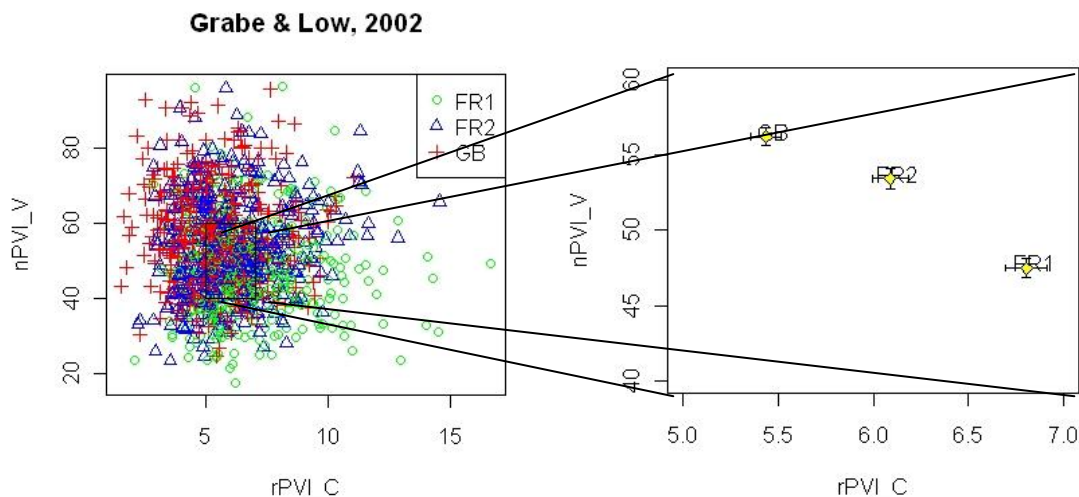


Figure 21. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan (rPVI_C,nPVI_V)

Au vu du graphique de dispersion, nous retrouvons les résultats présentés ci-dessus. Le taux nPVI_V est plus élevé pour les natifs que pour les apprenants ; à l'inverse, le rPVI_C est plus élevé pour les apprenants. Le rPVI_C permet une distinction nette des trois groupes alors que le nPVI_V permet plutôt une distinction du niveau des apprenants. Par contre, la combinaison rPVI_C - nPVI_V ne permet pas une distinction des trois groupes ($F(1, 1257)=4,103$; $pval=0,01675$).

- cvC, cvV (Dellwo, 2006 ; White & Mattys, 2007).

*rappel des calculs des paramètres :

cvC/cvV : coefficient de variation (écart type sur la moyenne) des durées des intervalles consonantiques / vocaliques.

*hypotheses :

Selon les différentes études dirigées par Dellwo (2006) puis White & Mattys (2007), cvC/cvV apparaissent également comme des mesures fiables, puisque le tempo est normalisé dans leurs calculs. Nous pouvons donc formuler l'hypothèse que ces mesures permettent de distinguer les groupes FR1, FR2 et GB. White & Mattys rapportent, entre autres, de leur étude, que le cvV est plus élevé chez les Anglais que chez les Espagnols.

L'espagnol étant une langue isosyllabique, nous pouvons penser que nous retrouverons des résultats similaires chez nos sujets français, à savoir, que nous nous attendons à ce que cvV soit moins élevé chez FR1 et FR2 que chez GB.

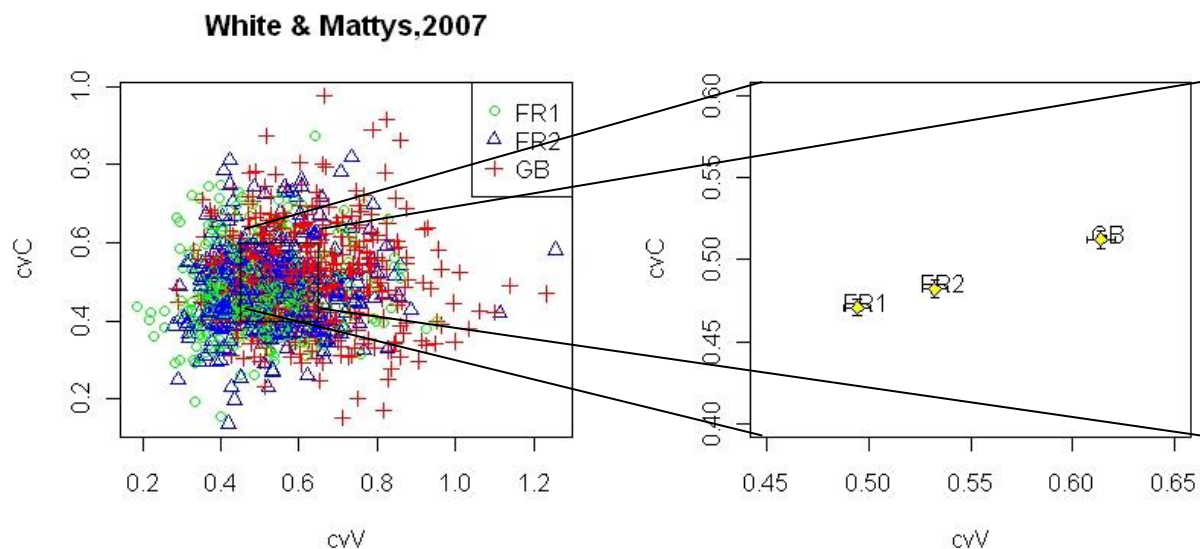


Figure 22. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan (cvV,cvC)

Au vu de ces graphiques, l'hypothèse formulée est validée. Le cvV pour GB est en effet plus élevé que celui de FR1 et FR2 ($F(1, 1257)=87,18$; $pval < 2,2 \times 10^{-16}$). Le cvC est également supérieur pour les natifs que pour les apprenants ($F(1, 1257)=17,21$; $pval = 4,248 \times 10^{-8}$). De plus, lorsque le paramètre cvV est considéré uniquement ou en combinaison avec le paramètre cvC ($F(1, 1257)=86,43$; $pval < 2,2 \times 10^{-16}$), on observe une distinction nette entre les trois groupes, le groupe des FR2 étant plus proche des FR1 qu'il ne l'est du groupe GB. Ces critères rythmiques pourraient donc jouer le rôle de critères évaluatifs puisqu'ils rendent compte linéairement des différents niveaux. Le paramètre rythmique cvC, en revanche, ne permet pas une distinction des groupes FR1 et FR2, les croix se chevauchant. Sur le graphe de dispersion, on retrouve en effet la tendance décrite, c'est-à-dire que l'on peut distinguer les natifs (en rouge) des FR, mais il est plus difficile de distinguer les deux niveaux de FR.

Selon les différents travaux de White & Mattys (2007,2008), il a été mis en évidence que la combinaison %V-cvV permettait une meilleure discrimination des langues, des productions de pathologie et des voix normales, ou encore une distinction L1/L2. Nous vérifions ce que cette combinaison montre sur nos propres données.

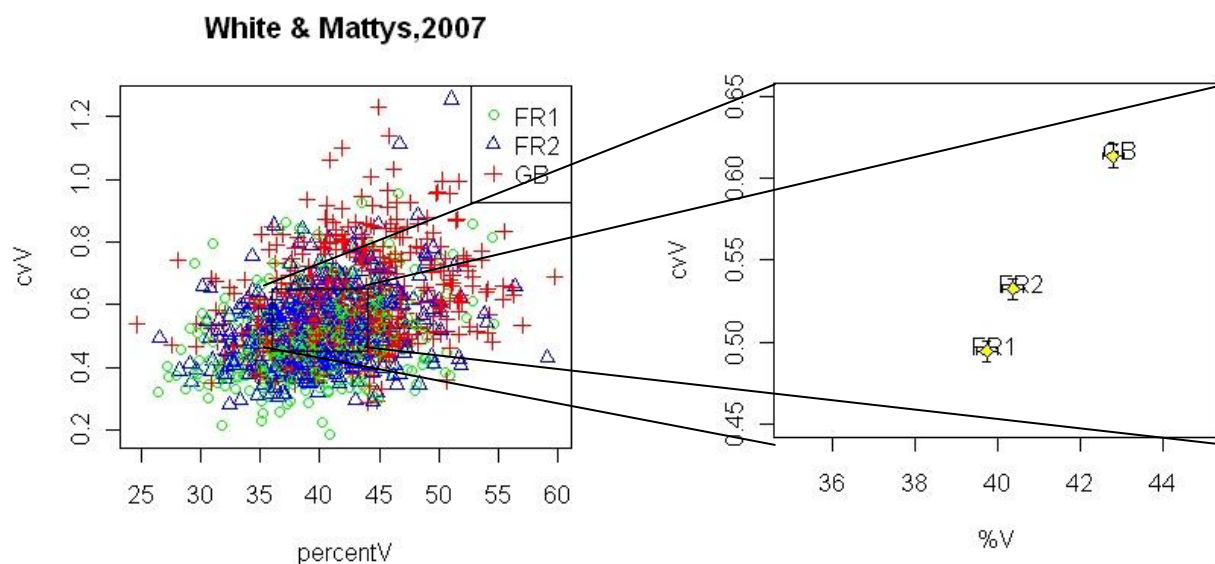


Figure 23. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan (%V,cvV)

Au vu des deux graphes, la distinction se fait au niveau GB/FR. Que ce soit %V ou cvV, les deux groupes FR sont à peine discriminés et sont très proches tandis que les GB s'éloignent assez bien des sujets francophones. Nous pouvons donc conclure qu'une telle combinaison serait intéressante pour une distinction entre natifs et apprenants et non, pour une distinction entre les niveaux des apprenants ($F(1,1257)=102,4$; $pval < 2,2 \cdot 10^{-16}$). Cette combinaison pourrait donc être complémentaire de la combinaison nPVI_V et nPVI_C qui permet plutôt une distinction des niveaux d'apprenants.

Nous proposons d'observer ici également la combinaison de cvC et de %V.

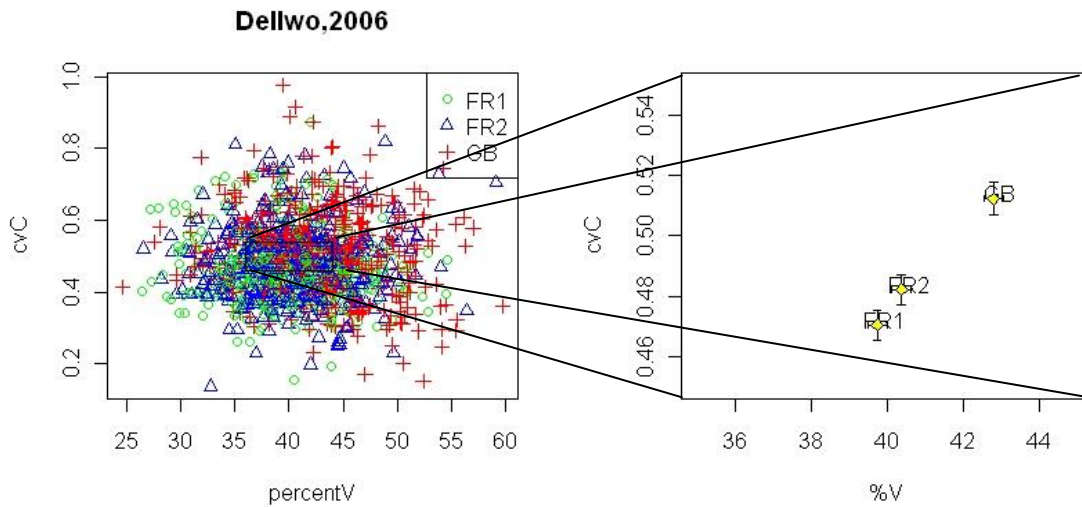


Figure 24. Visualisation des points de dispersion (gauche) et des erreurs-types (droite) selon le plan (%V, cvC)

Le plan (%V, cvC) est similaire à celui du cvV, cvC. Le plan (%V, cvC) montre clairement une distinction entre les groupes des francophones (FR1 et FR2) et le groupe d'anglophones, aussi bien lorsque les paramètres rythmiques %V et cvC sont considérés séparément, qu'en combinaison linéaire³⁸ ($F(1,1257)=46,46$; $pval < 2,2 \cdot 10^{-16}$). La distinction entre les groupes FR1 et FR2 est moins flagrante, les croix n'étant pas loin de se chevaucher. Ces critères rythmiques semblent donc, à première vue, rendre compte des différents niveaux des apprenants français (FR2 étant le plus proche de GB, FR1 le plus éloigné) et permettent la distinction des productions oralisées des francophones et celles des anglophones.

Nous pouvons d'ores et déjà affirmer que la combinaison de plusieurs paramètres rythmiques permet la classification des groupes FR1, FR2 et GB, la considération d'un seul paramètre pouvant s'avérer parfois non discriminante (par exemple, dans le cas de ΔV ,

³⁸ Nous proposons une combinaison linéaire des différents paramètres suite aux études de (Herry & Hirst, 2002 ; Dellwo, 2006)] qui ont montré que la combinaison de plusieurs facteurs rythmiques améliore la discrimination inter- et intra-langues ainsi qu'une discrimination L1-L2.

cvC, rPVI_V et nPVI_C). Nous pouvons également conclure que certaines combinaisons permettent la distinction des niveaux de langue, alors que d'autres permettent celle entre natifs et apprenants.

Afin de confirmer les résultats graphiques que nous venons de commenter, et d'évaluer quel(s) paramètre(s) serai(en)t le(s) critère(s) évaluatif(s) le(s) plus pertinent(s), nous menons des analyses statistiques.

5.3.3 Essai de paramétrisation rythmique : recherche de combinaisons par SVM

Nous proposons, à présent, d'évaluer les paramètres rythmiques, afin d'établir quel est celui ou ceux (par combinaison) qui sont à considérer pour une évaluation du rythme de l'apprenant français. A partir de 9 paramètres rythmiques, plus de 5000 combinaisons sont possibles. Nous allons donc en calculer plus de 255 (sur la base de 1 à 4 paramètres) afin de dégager les meilleures combinaisons et de ce fait, les meilleurs résultats pour une tâche de discrimination L1/L2.

Les combinaisons traitées ici ont été calculées à partir d'un script élaboré par Daniel Hirst (cf. annexes, section 2). Les données de sortie sont récupérées dans un fichier .txt. Les tableaux présentés ci-dessous, résultant d'une analyse discriminante³⁹, rapportent les pourcentages globaux obtenus (c'est-à-dire les taux de prédiction) pour chacune des possibilités combinatoires testées (le détail des résultats obtenus pour chaque groupe est donné en annexes, section 2). Les résultats (en %) ont été ordonnés de façon décroissante dans les tableaux proposés.

La méthodologie suivie se base sur la proposition d'analyse faite par Whitaker (1997). L'auteur remet en cause l'utilisation de l'analyse pas à pas (stepwise methods) - recommandée par tous les paquets statistiques- qui engendre un certain nombre d'erreurs (degré de liberté incorrect, échantillonnage, estimation du sous-ensemble optimal des

³⁹ Support Vector Machines, R.

prédicateurs). Ainsi, au lieu d'effectuer une analyse pas à pas, nous avons donc choisi une analyse complète permettant d'étudier toutes les combinaisons possibles.

▪ *Analyse à un facteur*

Nous tentons de savoir si un seul facteur donne d'assez bons résultats de fiabilité pour être utilisé comme critère évaluatif. Ci-dessous le tableau récapitulatif des valeurs obtenues pour chacun des 9 paramètres.

VARIABLE	SCORE OBTENU
cvV	45,79%
nPVI_V	43,57%
ΔV	43,09%
% V	42,61%
ΔC	42,61%
cvC	42,38%
rPVI_C	42,14%
nPVI_C	38,88%
rPVI_V	37,69%

Tableau 9. Présentation des données de sortie d'une analyse SVM - une discrimination à un paramètre

De manière globale, une discrimination à un facteur donne une fiabilité de classement d'une moyenne de 42% (effectuée sur les paramètres les plus pertinents). Le paramètre qui se dégage et se distingue des 8 autres est celui du cvV avec 45,79% (facteur préconisé par White & Mattys, cependant en combinaison avec %V). Le nPVI_V et le ΔC suivent de près avec une moyenne de 43%.

Pour la suite des analyses de combinaisons à plusieurs facteurs, la fonction « cross-validation » par dixième est utilisée, c'est-à-dire que l'analyse s'effectue sur les 9/10^{ème} du corpus avec validation sur le dixième qui reste et qui n'a donc pas été utilisé pour faire les prédictions ; ensuite, on prélève un autre dixième et on recommence, et ainsi de suite pour tout le corpus. L'utilisation de la méthode des machines à vecteurs de support (support vector machine « SVM ») permet de pouvoir opérer à diverses combinaisons, et ne prend

pas obligatoirement l'ensemble des facteurs. Cette analyse présente plusieurs avantages (cf. Baayen (2008), SVM vs. ADL § 5.2.3).

▪ *Analyse combinatoire à deux facteurs*

Lorsqu'on prend en considération deux facteurs, on observe que la moyenne des pourcentages les plus élevés se situe à environ 56% sauf pour la combinaison ($\Delta C, cvC$) qui se démarque nettement avec une fiabilité de **62,30%**.

VARIABLES	SCORE OBTENU
$\Delta C, cvC$	62,30%
rPVI_C,nPVI_C	57,30%
cvC,rPVI_C	54,92%
$\Delta V, cvV$	54,04%

Tableau 10. Présentation des résultats des meilleures combinaisons à deux facteurs

Afin d'évaluer le poids de chacun des deux paramètres dans la prédiction des classes, nous représentons graphiquement une analyse de dispersion des points ainsi que l'erreur-type des deux paramètres combinés $\Delta C, cvC$.

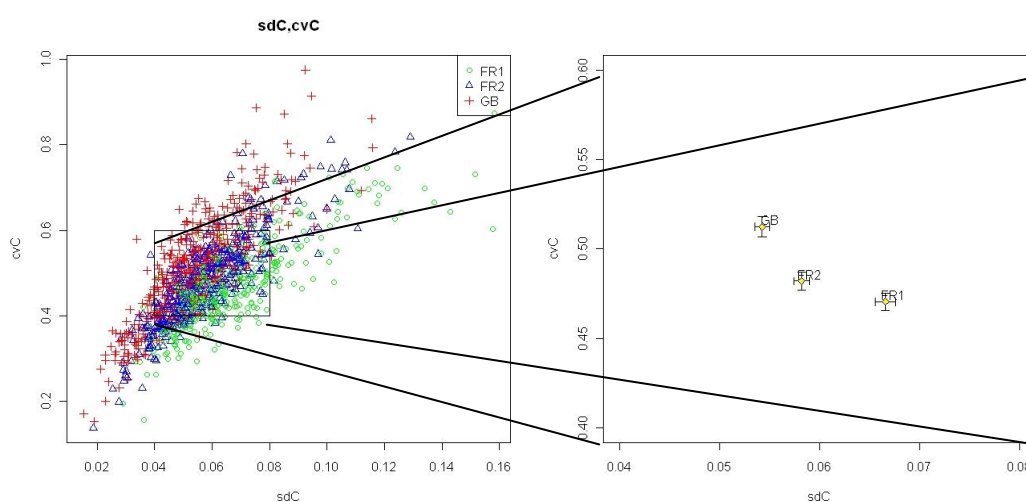


Figure 25. Visualisation des points de dispersions (gauche) et des erreurs-types (droite) selon le plan ($\Delta C, cvC$)

A partir du graphe des points de dispersion, on peut poser la question d'une éventuelle corrélation entre ces deux paramètres. Nous effectuons donc une analyse de variance entre ΔC et cvC qui s'avère confirmer notre observation. Une analyse ANOVA vient confirmer une forte corrélation ($F(1,1258) = 1495$; $p\text{-val} < 2,2e^{-16}$). Or, on s'attendait, *a contrario*, à ce que ces deux paramètres ne soient pas corrélés puisque cette combinaison à deux facteurs ressortait comme la plus discriminante. Cependant, sachant que les paramètres ΔC , cvC , pris séparément, permettent respectivement une prédiction à 42,61% et 42,38%, nous pouvons penser que la prédiction de 62,30%, obtenue par combinaison, serait due aux résidus qui expliqueraient l'amélioration du score.

▪ *Analyse combinatoire à trois facteurs*

La meilleure combinaison à 3 facteurs qui donne une fiabilité de **63,09 %** comprend (**%V, ΔV , cvV**). Cependant, l'ensemble des meilleures combinaisons retenues (cf. tableau 11), suit de près la première proposée car elles se situent à environ 61-62%.

VARIABLES	SCORE OBTENU
%V, ΔV, cvV	63,09%
ΔC, cvC, $rPVI_V$	62,69%
ΔC, ΔV, cvC	62,46%
%V, ΔC, cvC	62,14%
ΔC, cvC, $nPVI_V$	62,14%
ΔC, cvC, cvV	62,06%
ΔC , cvV , $nPVI_C$	61,66%
ΔC , cvC , $rPVI_C$	61,50%
%V, $rPVI_V$, $nPVI_V$	61,34%

Tableau 11. Présentation des résultats des meilleures combinaisons à trois facteurs

Il est à noter ici, que les scores des combinaisons à trois facteurs sont quasi-équivalents aux scores des combinaisons à deux facteurs, avec en moyenne, une fiabilité de 63%. Par ailleurs, nous remarquons que le paramètre %V, important dans une combinaison

à trois facteurs, permet d'augmenter la fiabilité de ($\Delta V, cvV$) de 54,04% à 63,09%. En revanche, lorsqu'il s'agit d'ajouter ce paramètre à la meilleure combinaison binaire sortante, à savoir ($\Delta C, cvC$), cela n'a aucune influence sur la valeur de prédiction qui reste à 62,14% (avec trois facteurs contre 62,30% à deux facteurs).

▪ *Analyse combinatoire à quatre facteurs*

Lorsqu'on passe à une combinaison à quatre facteurs, la fiabilité augmente positivement.

VARIABLES	SCORE OBTENU
%V, $\Delta V, cvV, nPVI_V$	66,74%
%V, $cvV, rPVI_V, nPVI_V$	66,03%
%V, $\Delta V, cvC, cvV$	65,71%
$\Delta C, \Delta V, cvC, nPVI_V$	65,15%
%V, $\Delta V, rPVI_V, nPVI_V$	65,07%
%V, $\Delta C, cvC, nPVI_V$	65%
$\Delta C, \Delta V, cvC, nPVI_C$	65%
$\Delta C, cvC, cvV, rPVI_V$	65%

Tableau 12. Présentation des résultats des meilleures combinaisons à quatre facteurs

Le meilleur taux de fiabilité obtenu se situe à 66,74% et combine **%V, $\Delta V, cvV, nPVI_V$** ; tout comme la combinaison à trois facteurs, ce sont les paramètres rythmiques des intervalles vocaliques qui priment, contrairement à la combinaison binaire qui est constituée de deux paramètres basés sur les intervalles consonantiques. Les autres meilleures combinaisons à quatre facteurs suivent de près la première et se situent à une moyenne de 65%.

Nous constatons donc que les valeurs de fiabilité augmentent avec une combinaison à quatre facteurs avec une marge de 4%. Cependant, par rapport à une combinaison binaire, la différence de fiabilité n'est pas très importante. Lorsque l'on passe de 1 à 2, puis

à 3 facteurs, on gagne en moyenne 10% dans notre prédiction ; si l'on suit ce schéma de progression, nous devrions obtenir une prédiction fiable à 75% avec une combinaison à quatre facteurs. On peut donc conclure que l'augmentation positive que l'on trouve avec quatre facteurs est finalement minime. La question se pose : faut-il prendre en compte 4 facteurs et obtenir une valeur de prédiction fiable légèrement plus importante, ou faut-il se limiter à un nombre moindre de facteurs mais avec une valeur de prédiction toujours relativement bonne? Pour notre part, nous pensons qu'une prédiction à deux facteurs à partir des critères $\Delta C, cvC$ semble plus pertinente car il nous semble que le meilleur modèle à utiliser est celui qui obtient le meilleur score avec le moins de facteurs possibles, mais surtout, cela permet de construire un feedback précis, sur des erreurs ciblées, et de comprendre où se situent les difficultés.

5.4 Discussion et conclusion

5.4.1 Influence de la L1 sur la L2 ?

La première étape était de pouvoir mettre en avant des différences entre les productions des natifs et les productions des francophones en anglais en essayant de voir si le rythme de la L1 était présent dans les productions en L2. Si le français est une langue isosyllabique, contrairement à l'anglais qui est isoaccentuel, les caractéristiques du rythme du français de niveau moyen ou faible, devraient venir se reproduire dans les productions en anglais. Ainsi nous devrions retrouver pour FR1 des résultats assez éloignés des FR2 et encore plus des GB.

Afin de repérer d'éventuels impacts du rythme de la L1 sur la L2, nous avons procédé à une représentation graphique des productions effectuée à partir du calcul des erreurs type (*standard error*) en fonction de deux paramètres rythmiques. Chacun des graphiques est complété par une ANOVA. Nous avons proposé d'étudier les métriques les plus utilisées dans la littérature.

Nos hypothèses fondées sur la littérature, se formulent comme suit :

Plus la production rythmique de l'apprenant est bonne, plus les valeurs obtenues seront proches de celles du natif et tendront ainsi à une production rythmique isoaccentuelle.

-Les taux de ΔC des apprenants devraient être élevés en raison de la structure syllabique complexe de la langue anglaise, tandis que les taux de ΔV devraient être assez bas.

-Le pourcentage de voyelles (%V) devrait être assez élevé pour les locuteurs francophones.

-Les taux de variabilité de PVI devraient être assez élevés si le rythme produit est proche de celui d'un natif.

-les taux de cvC et cvV devraient également être élevés si la production rythmique est correctement réalisée.

Il est possible de distinguer deux types de résultats :

(i). Les résultats corroborent les hypothèses.

A cette corroboration sont attribués les paramètres rythmiques suivants, (seuls et en combinaison) : (nPVI-C,nPVI_V) et (cvV,cvC). Dans les deux cas de combinaisons, les valeurs obtenues étaient celles attendues, à savoir des taux de PVI normalisés et des taux de coefficient de variation élevés pour les natifs, à l'inverse des taux de francophones plutôt bas, correspondant à une production rythmique proche du français. Les taux pour le groupe des FR2 se situent à l'intermédiaire des deux groupes opposés. Ce sont ici les métriques 'normalisées' qui donnent de meilleurs résultats ; la neutralisation des variations de tempo semble donc améliorer les résultats de ces métriques. Ces deux combinaisons semblent refléter au mieux l'influence de la L1 sur la L2.

(ii). Les résultats infirment les hypothèses.

On trouve plusieurs paramètres traités de manière individuelle ou en combinaison, dont les taux obtenus viennent contredire les résultats annoncés dans la littérature: de manière générale, les valeurs donnent des résultats contraires et inversés à ceux des études de référence. Le %V des francophones est relativement bas et plus bas que celui des natifs alors que ce paramètre était prédit plus élevé. Les valeurs de ΔC ainsi que du rPVI des anglais sont bien plus basses que les valeurs obtenues par Ramus et Grabe et inférieures aux valeurs obtenues pour les francophones. Les métriques non normalisées, comme

annoncé dans la littérature, n'indiquent pas de résultats satisfaisants et ne montrent en rien une influence rythmique.

Ces résultats pourraient s'expliquer par le phénomène du rythme des textes utilisés pour l'analyse (Hirst, 2009). En effet, dans ce contexte, le même texte a été utilisé pour les natifs et non natifs, on a donc une neutralisation des différences de structure rythmique. Le rythme du texte est donc neutralisé et cela pourrait expliquer les résultats contraires à ceux énoncés dans la littérature. Ainsi, face à la difficulté phonotactique à laquelle les locuteurs francophones sont confrontés, un effet d'hyper articulation pourrait également confirmer des taux de ΔC , rPVI élevés, car même si la différence de structure consonantique a été neutralisée, le locuteur francophone n'a pas pour habitude de prononcer des séquences de consonnes aussi complexes et pourrait produire une surarticulation.

5.4.2 Discrimination à partir des critères rythmiques

Nous avons établi qu'il y avait des différences de rythme dans la production des francophones et des natifs. Nous avons pu voir, qu'un francophone de niveau faible (FR1) en anglais, aurait tendance à reproduire un rythme isosyllabique dans sa production orale. De fait, s'il existe bien des différences, est-il possible, par le biais des métriques rythmiques, (i) de différencier des natifs/ des apprenants, et (ii) de différencier le niveau des apprenants et ainsi obtenir une gradation FR1/FR2/GB ?

Afin d'établir s'il était possible de différencier plusieurs niveaux et d'en dégager des critères rythmiques, nous avons procédé à plusieurs analyses statistiques.

➤ Analyse discriminante

Il a été montré qu'avec l'ensemble des paramètres rythmiques, il était possible de classer correctement les locuteurs dans leur groupe avec un score global de 69,50%. Cette discrimination a montré une distinction majeure entre FR1 & GB qui sont les deux groupes opposés (natifs vs. apprenants de niveau faible) avec une marge minime d'erreur de 6% pour les deux groupes respectifs. Le groupe FR2 se trouve à mi-chemin entre les opposés, ce qui correspond relativement à la réalité de la composition des groupes. Cela nous

permet de confirmer, par ailleurs, le choix des critères de sélection qui ont été utilisés lors de la constitution des groupes.

➤ Représentation graphique de chaque paramètre

Nous avons testé les différentes métriques proposées dans la littérature, qui représentaient celles les plus utilisées ; nous avons également étudié les différentes combinaisons proposées par les mêmes auteurs. Cela nous a permis de constater que les combinaisons sont meilleures prédictrices qu'un paramètre unique. En effet, un seul paramètre, quel qu'il soit, n'obtient au maximum que 43% et ne dépasse en aucun cas, les 50%.

Par ailleurs, certaines métriques permettent la distinction natifs/apprenants alors que d'autres permettent plutôt de différencier des niveaux au sein des apprenants. Enfin, une troisième catégorie de métriques permet d'établir une gradation entre les 3 niveaux. Voici, un tableau récapitulatif des différentes métriques, en fonction de la tâche de discrimination visée.

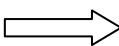
Niveaux de distinction 	apprenant vs natif	FR1 vs FR2	Gradation (FR1-FR2-GB)
1 PARAMETRE	% V	ΔC	ΔC
	cvC	nPVI_C	rPVI_C
		nPVI_V	cvV
COMBINAISON	% V	%V, ΔC	rPVI_V, rPVI_C
	cvV	nPVI_C, nPVI_V	cvV, cvC
			%V, cvC

Tableau 13. Récapitulatif des différentes métriques présentées en fonction de la tâche de discrimination visée

D'après le tableau présenté ci-dessus, il est possible de distinguer 3 types de discrimination en fonction des paramètres obtenus. Ainsi, %V et cvC ainsi que la combinaison %V, cvV donnent une meilleure distinction apprenant/natif. Les paramètres

ΔC , nPVI_C et nPVI_V ainsi que les combinaisons %V, ΔC ; nPVI_C,nPVI_V permettent une meilleure distinction FR1/FR2. Enfin, une gradation, allant de FR1 jusqu'à GB, est possible par l'ensemble des facteurs suivants : ΔC , rPVI_C, cvV en tant que facteurs individuels, ou en combinaisons avec rPVI_V,rPVI_C ; cvV,cvC ; %V,cvC.

➤ SVM (*support vector machine*)

Parce que les combinaisons semblent apporter de meilleurs résultats et que certaines métriques permettraient une distinction graduelle des trois groupes, une SVM par factorisation de diverses combinaisons, allant de 1 à 4 paramètres, est menée afin d'obtenir les meilleurs prédictions possibles. Les meilleures prédictions pour chacune des combinaisons factorielles sont répertoriées dans le tableau ci-dessous :

VARIABLE	SCORE OBTENU
cvV	45,79%
ΔC , cvC	62,30%
%V, ΔV ,cvV	63,09%
%V, ΔV ,cvV,nPVI_V	66,74%

Tableau 14. Récapitulatif des meilleures prédictions pour chacune des combinaisons factorielles

Ces résultats confirment les observations graphiques des représentations de l'erreur-type, selon lesquelles les combinaisons des paramètres, qu'elles soient à deux ou divers facteurs, sont meilleures qu'un paramètre seul. Plus il y a de paramètres, plus la fiabilité de la prédiction augmente. Cependant, on remarque que les prédictions obtenues pour la combinaison - ΔC ,cvC- sont quasi équivalentes à celles des prédictions avec 2 ou 3 paramètres. Cette combinaison est *a priori* surprenante puisqu'on aurait pu penser que ces deux facteurs seraient fortement corrélés.

Si l'on compare ces données avec les combinaisons proposées dans la littérature, aucune des combinaisons binaires avancées ne figure dans la liste des meilleures combinaisons. Selon l'analyse SVM, ces combinaisons obtiennent les résultats de fiabilité suivants :

COMBINAISON PROPOSEE	SCORE OBTENU (SVM)
% V, ΔC	46,58%
nPVI_V, rPVI_C	48,65%
%V, cvV	49,68%
%V, cvC	46,34%

Tableau 15. Présentation des prédictions obtenues par l'analyse SVM des métriques exposées dans la littérature

Ces résultats sont loin du taux de fiabilité qu'obtient $\Delta C, cvC$ avec plus de 63%.

Par ailleurs, nous avons conclu qu'un modèle simple, mais apportant un taux de prédiction pratiquement aussi élevé qu'un modèle pluri-métrique était préférable. Pour notre part, nous pensons donc que la combinaison $\Delta C, cvC$ apparaît être la combinaison la plus discriminante et la plus fiable, dans une tâche de discrimination L1/L2, anglais/français. Cela, bien évidemment, en étant consciente que cette mesure, ΔC , est largement critiquée dans la littérature de par son lien avec les fluctuations de tempo.

On peut se demander à présent, si cette combinaison ne pourrait pas être également efficace dans la distinction inter-langues, inter-dialectales et pour les études de l'évaluation en pathologie. En effet, il serait intéressant de confirmer ces résultats, en testant cette combinaison sur d'autres corpus, y compris des corpus L1/L2, afin de confirmer ces premiers résultats.

Il serait donc tout à fait envisageable d'utiliser la combinaison à deux facteurs $\Delta C, cvC$ à des fins évaluatives dans une comparaison de productions rythmiques natifs/francophones.

Conclusion et perspectives

L'objectif de ce travail était de proposer des critères rythmiques évaluatifs de la production orale d'apprenants français en anglais L2, dans le cadre de l'élaboration d'une évaluation objective de la prosodie de l'anglais britannique. L'idée était d'asseoir des critères objectifs évaluatifs du rythme, à partir de métriques rythmiques.

Dès l'introduction, nous avons mis en évidence le manque d'évaluation de l'oral en anglais L2 et tout particulièrement de la prosodie. Nous ne cherchions pas à convaincre ; toute personne s'intéressant à l'apprentissage de la langue doit déjà être persuadée de l'importance des traits prosodiques dans la production langagière. Le fait est que la prosodie pose un certain nombre de problèmes aux apprenants, elle est un facteur majeur dans la compréhensibilité du message à faire passer à son interlocuteur. Nous avons vu que les logiciels multimédia pouvaient être d'une grande aide pour remédier à ces difficultés et apporter un souffle nouveau dans l'enseignement de la phonétique. Cependant, le constat est pertinent : aucune évaluation objective de la prosodie n'est proposée dans ces outils multimédia, un problème qui s'étend d'ailleurs à l'enseignement de la phonétique de manière globale, dans lequel bien souvent, aucune grille évaluative de la prosodie n'est proposée pour évaluer l'oral des étudiants. Cependant, avant de proposer un critère quel qu'il soit, il était nécessaire en premier lieu d'analyser le cadre de travail dans lequel nous nous situons.

Le premier travail de cette recherche était avant tout descriptif. Avant de pouvoir envisager de soumettre un critère rythmique, il fallait décrire l'objet d'étude de ces travaux. Une recherche approfondie sur la notion de rythme nous a permis de comprendre les différences essentielles entre le fonctionnement des deux systèmes prosodiques des langues étudiées (le français et l'anglais), ce qui nous a conduits à prendre conscience d'un certain nombre d'interférences rythmiques potentielles chez l'apprenant francophone. Ce dernier, nous l'avons vu, se retrouve confronté à de nombreuses difficultés, conscientes et/ou inconscientes. Nous mesurons à présent les problèmes auxquels nous devons faire face lorsque nous basculons d'une langue à l'autre, dans la mesure où il faut plonger vers un système prosodique complètement différent.

Malgré tout, un certain nombre de thématiques n'ont pu être abordées, du fait du champ large de recherche que couvre la prosodie. Nous avons donc dû faire un choix, et celui-ci s'est porté tout naturellement sur le rythme. Nous avons pu constater en effet, aussi bien dans la littérature, que dans le cadre de notre métier d'enseignant, qu'il était tout particulièrement difficile à acquérir pour les apprenants. Nous ne réfutons pas pour autant l'importance du rôle joué par les autres composantes prosodiques, également difficiles à acquérir, mais la place importante qu'occupe le rythme dans la production orale nous a confortée dans notre premier choix. Sans aucun doute, cette recherche se doit d'être complétée par l'étude des autres composantes prosodiques. Par ailleurs, nous avons dû également cibler dans ce travail une partie des aspects rythmiques de la langue. Une étude exhaustive pourrait se pencher sur d'autres phénomènes rythmiques, tels que les pauses, les durées syllabiques, ces éléments étant difficilement indissociables des faits prosodiques.

Parallèlement à cela, nous avons porté grand intérêt à détailler et analyser les outils multimédia proposés actuellement, ce qui fut nécessaire pour constater le manque de paramètres évaluatifs, mais aussi attester de l'apport positif que ces méthodes pourraient susciter.

Le premier objectif fixé pour ce travail de recherche était de proposer une base de données comparatives d'anglais L1/L2, sur laquelle il serait possible de travailler à l'élaboration de critères prosodiques mais également, qui viendrait pallier un manque étonnant de corpus disponibles d'anglais L2. C'est ce que nous nous sommes attachée à faire et c'est ce que nous présentons dans le chapitre 4 de cette thèse. La réalisation de cette base de données a été le fruit d'une longue réflexion et expérimentation qui nous a amenée à la création d'une base de données conséquente nommée ENGLISH. Une des volontés face à cette tâche était de rendre ces enregistrements disponibles à la communauté, ce qui est bien le cas puisque ENGLISH sera disponible librement sur le site du CRDO <http://crdo.fr>, permettant ainsi une large diffusion, ainsi que, nous l'espérons, un large retour qui contribuera à l'enrichissement de la base de données. La partie lecture de ce corpus nous a permis d'établir nos analyses expérimentales. L'élaboration de cette base de données a constitué un point essentiel à cette recherche et nous donne la possibilité d'isoler des paramètres rythmiques importants pour la mise en place d'une évaluation.

Le second objectif de ce travail, que nous traitons dans le chapitre 5, était de mener à bien une proposition de critères pour établir une évaluation objective de l'oral des

apprenants francophones. L'idée est de quantifier un paramètre rythmique afin de proposer un critère fiable d'évaluation. Nous avons choisi d'utiliser les métriques rythmiques dans la mesure où la littérature a mis en avant le rôle de discrimination qu'elles pouvaient jouer entre les langues ou les dialectes. Une première étape réside dans l'évaluation des différentes métriques les plus utilisées proposées dans la littérature, par une représentation des différentes combinaisons selon le calcul de l'erreur-type (méthode utilisée par Ramus *et al.*, 1999). Cela nous a permis de constater dans un premier temps d'un point de vue graphique, qu'il y avait bien une influence du rythme de la L1 sur la L2 mettant en évidence des différences rythmiques dans les productions. Plus la production de l'apprenant était rythmiquement proche d'un rythme français, plus il s'éloignait des valeurs des natifs. Par ailleurs, il a été mis en avant que des combinaisons de paramètres étaient plus efficaces qu'un paramètre isolé. Par la suite, une analyse discriminante calculée sur l'ensemble des neuf paramètres proposés, prédit de manière significative à 69,50% de fiabilité, l'appartenance des locuteurs à leur groupe de niveau (FR1, FR2, GB). Afin de rendre compte des différences rythmiques par un critère évaluatif précis, nous avons mis en évidence que :

- (i) une combinaison binaire était plus favorable à une discrimination dite fiable, à l'instar de combinaisons pluri-paramétriques,
- (ii) certaines combinaisons pourraient servir à distinguer natifs/apprenants, d'autres à faire une distinction FR1/FR2, et une troisième catégorie discriminant le niveau des 3 groupes,
- (iii) la combinaison $\Delta C, cvC$ était la combinaison la plus fiable et que ces deux paramètres rythmiques pourraient être pris en compte comme critères évaluatifs dans le but de mettre en avant la tendance rythmique des productions des apprenants francophones.

Il est possible de discriminer la L1 de la L2, et de distinguer les productions rythmiques des apprenants. Nous souhaitons insister sur le fait que la combinaison métrique utilisable comme critère évaluatif, donnera une tendance rythmique des productions, à savoir plus ou moins isoaccentuelle, plus ou moins proche des productions de natifs. Il est évident que cette tendance doit être renforcée par d'autres critères rythmiques précis, afin d'optimiser les résultats d'une évaluation proposée à l'apprenant.

Ce travail représente un point de départ. Il convient d'approfondir la problématique et compléter cette recherche qui ouvre la voie à de nombreuses perspectives de travail. En effet, beaucoup reste à faire dans ce domaine, d'une part pour définir les points de rupture entre la prosodie de l'anglais et celle des apprenants francophones, et d'autre part, pour élaborer un système évaluatif objectif, basé sur de nombreux critères pertinents, dans le but d'améliorer la production des apprenants et faire en sorte que parler une langue étrangère, et notamment ici, l'anglais, ne soit plus une « corvée » ou une « honte » pour un francophone.

Les résultats des analyses expérimentales permettent de proposer des perspectives et des propositions didactiques.

... Perspectives

Tout d'abord, nous pensons qu'il serait intéressant d'exploiter l'exercice de répétition du corpus dans le but de mener une étude comparative entre la production de lecture et la tâche de répétition. Cette analyse nous permettrait de mieux cibler les erreurs de production en essayant de noter si, une « auto-correction » des erreurs produites en lecture, a été opérée par le locuteur lors de la phase répétition. Cela pourrait permettre l'amélioration de l'élaboration d'un feedback plus précis.

Il serait également intéressant de penser à la mise en place d'un test de perception qui permettrait l'évaluation perceptive des productions francophones en les classant sur une échelle graduée. Nous pourrions ainsi mettre en place une échelle à plusieurs niveaux, donc au-delà des trois niveaux proposés dans ce travail, à partir de laquelle un apprenant pourrait se voir classé, et ainsi apprécier son évolution dans l'apprentissage de la L2. Un tel test permettrait également de valider la répartition des locuteurs et s'inscrirait en complément d'une évaluation subjective à partir de critères de « niveaux académiques », tels que proposés dans cette thèse. Le groupe défini répondrait plus au critère d'homogénéité recherché pour la répartition des locuteurs. Le test de perception permettrait « l'affinage » des groupes puisqu'il amène à l'évaluation objective des variabilités intra-locuteurs possibles au sein d'un même groupe.

Concernant les aspects mêmes de ce travail, nous pensons qu'il serait intéressant d'exploiter la partie spontanée du corpus, selon une même méthodologie, c'est-à-dire par la

mise en place de critères rythmiques (que nous pouvons imaginer différents pour la parole spontanée) afin d'évaluer l'apprenant dans son apprentissage de la L2. S'il est possible d'établir de tels critères, à partir de la parole spontanée, cela permettrait d'évaluer le niveau de l'apprenant dans une production qui lui appartient et qui est non figée sur la répétition de la production orale du natif. De tels critères et une telle évaluation viendraient ainsi compléter la base méthodologique élaborée dans ce travail et les perspectives présentées ci-dessus.

Par ailleurs, nous pourrions tester d'autres métriques présentées dans la littérature mais jusqu'ici très peu usitées et notamment utiliser les algorithmes développés par Loukina & al. (2009) pour automatiser des prédictions de classement.

Pour optimiser les résultats obtenus avec l'analyse discriminante, nous pourrions effectuer une discrimination français/anglais, mais aussi FR1/FR2. Des premiers résultats montrent déjà des perspectives très intéressantes avec 90% de fiabilité pour une discrimination FR/ANG. Par ailleurs, afin de compléter l'élaboration d'un feedback, il s'agirait d'analyser les productions par passage entier, et non plus par phrase, mais aussi, par les 4 passages, c'est-à-dire par locuteur. Jusqu'ici, nous avons élaboré une aide au locuteur en évaluant son énoncé et en lui donnant une tendance rythmique ; par la suite, nous pourrions donner une évaluation des locuteurs eux-mêmes.

L'élaboration de critères, autres que rythmiques, comme nous l'avons mentionné plus haut, apparaît également nécessaire. De tels critères, élaborés à partir des composantes prosodiques que nous connaissons, viendraient compléter ce travail et l'enrichir tout naturellement. Une fois les critères prosodiques bien établis, i.e. décidés à la fois, à partir d'une lecture exhaustive des travaux menés dans le domaine et à partir de la réflexion d'enseignants sur les difficultés d'apprentissage de la L2, des seuils pourraient être définis et ainsi renseigner l'apprenant à la fois sur son niveau de langue, en terme de « prosodie » (nous ne proposons pas une évaluation de ses compétences linguistiques mais de sa prononciation), et sur ses progrès dans l'apprentissage de cette langue. La détermination de seuils pourrait se faire notamment par la collaboration de chercheurs, d'enseignants et d'informaticiens.

A long terme, nous pouvons penser qu'une telle méthodologie permettra l'élaboration d'un système d'évaluation automatique de la prosodie, applicable au champ de la L2 mais également en pathologie ou en parole de synthèse.

Références

- Abercrombie, D. (1967) *Elements of General Phonetics*, Edimbourg, Grande Bretagne: Edinburgh University Press, 203p.
- Adams, C. (1979) *English Speech Rhythm and the Foreign Learner*, Walter de Gruyter & Co, 231p.
- Allen, G.D. (1972) The location of Stress Beats in Rhythmic English, An Experimental Study, 2, Off-print, *Language and Speech*, 15 (2), pp.179-195.
- Allen, G.D. (1975) Speech rhythm: its relation to performance universals and articulatory timing, *Journal of Phonetics*, 33, pp.75-86.
- Alliaume, J. (1989) Prosodie et apprentissage des langues, *Cahiers de L'APLIUT*, Association des Professeurs de langue des IUT, 32 (3), Actes de Congrès, pp.96-104.
- Alliaume J. ; Hesling, I. (1998) *Le langage et les deux cerveaux*, Editions CERALEC, Fort de France, 127p.
- Anderson-Hsieh, J. ; Johnson.R ; Koehler, K. (1992) The Relationship between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure, *Language Learning*, 42 (4), pp. 529-555.
- Asher, R.E. (1985) *Tamil*, London: Croom Helm.
- Astésano, C. (2001) *Rythme et accentuation en français. Invariance et variabilité stylistique*, Coll. Langue et Parole, Paris, France : L'Harmattan, 336 p.
- Aubry, Y. (2004) *Logiciel du traitement de la parole et d'aide à l'enseignement et à l'apprentissage de la prosodie : application au breton*, travaux de D.R.T. d'ingénierie, université du Maine.
- Auran, C. (2004) *Prosodie et anaphore dans le discours en anglais et en français: cohésion et attribution référentielle*, thèse de doctorat, université de Provence.
- Auran, C. ; Bouzon, C. ; Hirst, D. (2004) The aix-MARSEC project: an evolutive database of spoken british English, In *Speech Prosody 3:04*, pp. 561-564.
- Baayen, R.H. (2008) *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*, Cambridge, 368 p.

- Barkat-Defradas, M. ; Hamdi, R. ; Pellegrino, F. (2004) De la caractérisation linguistique à l'identification automatique des dialectes arabes, *MIDL*, Paris, 29-30 novembre 2004.
- Barry, W.J. ; Andreeva, B. ; Russo, M. ; Dimitrova, S. ; Kostadinova, T. (2003) Do rhythm measures tell us anything about language type, *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, pp.2693-2696.
- Beck, A. ; Décurré, N. ; Gaillard, P. ; Kabbaj, N. ; Péchou, A. ; Stenton, A. ; Tazi, S. ; Toma, A. ; Tricot, A. ; Vaillant-Sirdey, C. (2005) SWAMS, un système auteur de synchronisation et d'annotation pour un apprentissage multimodal des phénomènes accentuels en langue vivante 2, *EIHA* 05, pp. 1-14.
- Beckman, M. (1992) Evidence For Speech Rhythms Across Languages, In Tohkura, Y. ; Vatikiotis-Bateson, E. ; Sagisaka, Y. (eds), *Speech Perception, Production And Linguistic Structure*, pp. 457-463.
- Bel, B. & Blache, P. (2006) Le Centre de Ressources pour la Description de l'Oral (CRDO), *Travaux Interdisciplinaires du Laboratoire Parole et Langage (TIPA)*, 25, pp. 13-18.
- Benguerel & D'Arcy. (1986) Time-Warping and the Perception of Rhythm in Speech. Off-print, *Journal of Phonetics*, 14, pp. 231-246.
- Bertinetto, P.M. (1989) Reflections on the dichotomy stress vs. syllable-timing, *Revue de Phonétique Appliquée*, 91-92-93, pp. 99-130.
- Best, C. (1995) A Direct Realist View of Cross-Language Speech Perception, in STRANGE, W. (ed), *Speech Perception and Linguistic Experience*, Issues of Cross-Language Research, Baltimore, IN, USA: New York Press, Inc., pp. 171-204.
- Best, C. ; McRoberts, G. ; Goodell, E. (2001) Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system, *Journal of Acoustic Society of America*, 109, pp.775-794.
- Bloch, B. ; Trager, G.L. (1942) *Outline of linguistic analysis*, Linguistic Society of America (pub.), Waverly Press.
- Bloch, B. (1950) Studies in Colloquial Japanese IV: Phonemics, *Language*, 26, pp. 86-125.
- Blumstein & Cooper (1974) Hemispheric processing of intonation, *Cortex*, 10, pp. 146-158
- Boersma, P. & Weenick, D. (2009) Praat: a system for doing phonetics by computer, version 5.1.04. Téléchargeable à partir de www.praat.org.

- Bolinger, D. (1965) Pitch accent and sentence rhythm, *Forms of English: Accent, morpheme, order*, Cambridge, MA: Harvard University Press, pp. 139-180.
- Bolinger, D. (1972) *Intonation*, Selected Readings, Coll. Penguin Modern Linguistics Readings, Harmondsworth, Grande Bretagne: Penguin Books, 464 p.
- Bolinger, D. (1981) *Two Kinds of Vowels, Two Kinds of Rhythm*, Bloomington, IN, USA : IULC Publications, 68 p.
- Bongaerts, T. (1999) Ultimate attainment in L2 pronunciation: the Case of very advanced late L2 learners, in: *Birdsong*, D. (ed.), *Second Language Acquisition and the Critical Period Hypothesis*, Lawrence Erlbaum, Mahway, NJ, pp. 133-159.
- Bonniol, J.J (1976) La fidélité en question, objet d'étude docimologique, *Connexions*, 19, pp. 3-78.
- Bonniol, J.J ; Vial, M. (1997) *Les modèles de l'évaluation, Textes fondateurs avec commentaires*, De Boert, 368p.
- Bosch, L. ; Sebastian-Galles, N. (1997) The role of prosody in infants' native-language discrimination abilities: the case of two phonologically close languages, In *EUROSPEECH-1997*, pp. 231-234.
- Boulton, A. (2008) Esprit de corpus : promouvoir l'exploitation de corpus en apprentissage des langues, revue électronique *Texte et Corpus*, 3, Actes des Journées de la linguistique de Corpus 2007, pp. 37-46.
- Bouzon, C. (2004) *Rythme et structuration prosodique en anglais britannique contemporain*, Thèse de doctorat, Université de Provence.
- Burns, T.C. ; Yoshida, K. ; Hill, K. ; Werker, J.F. (2007) The development of phonetic representation in bilingual and monolingual infants, *Applied Psycholinguistics*, Psychological and Linguistic Studies across Languages and Learners, 28 (3), Cambridge, Grande Bretagne : Cambridge University Press, pp. 455-474.
- Caelen-Haumont, G. (1983) Le rythme dans la parole: une revue des études portant sur le français, Actes du 4ème colloque *Langages et Signification*, *ALBI*, I, Le Rythme, pp. 233-245.
- Calbris, G. ; Montredon, J. (1975) *Approche Rythmique, Intonative et expressive du français langue étrangère*, vol.1, Clé International, Paris, 77p.
- Carter, P. (2005) Quantifying rhythmic differences between Spanish, English and Hispanic English, Theoretical and Experimental approaches to romantic

- linguistics, *Current Issues in Linguistics Theory* 272, John Benjamins P.C, pp. 64-75.
- Carr, P. (1999) *English phonetics and phonology*, Oxford (UK), Malden, 169p.
- Catford, J.C. (1977) *Fundamental Problems in Phonetics*, Edimbourg, Grande Bretagne : Edinburgh University Press, 278p.
- Cazade, A. (1999) De l'usage des courbes sonores et autre supports graphiques pour aider l'apprenant en langues, *ALSIC*, vol. 2 (2), pp. 3-32.
- Caverni, J.P., Fabre, J.P., Noizet, G., (1975) Dépendance des évaluations scolaires par rapport à des évaluations antérieures, *le travail humain*, tome 38 (2), pp. 213-222.
- Champagne-Muzar, C. ; Bourdages, J.S. (1998) *Le point sur la phonétique*, CLE International, 101p.
- Chan, D., Fourcin, A., Gibbon, D., Grandström, B., Huckvale, M., Kokkinakis, G., Kvale, K., Lamel, L., Lindberg, B., Moreno, A., Mouropoulos, J., Senia, F., Transcoso, I., Velt, C. & Zeiliger, J. (1995) "EUROM – A Spoken Language Resource for the EU." *In Proceedings of Eurospeech '95*, Madrid, 1995.
- Cheong, H.S. (2007) *The role of listener affiliated socio-cultural factors in perceiving native accented versus foreign accented speech*, thèse de doctorat, the Ohio state university.
- Cho, Y.S. (2009) *Etude sur la prosodie de la langue coréenne*, thèse de doctorat, université de Provence.
- Chomsky, N. ; Halle, M. (1968) *The Sound Pattern of English*, Longman Higher Education, 470p.
- Classe, A. (1939) *The rhythm of English prose*, B. Blackwell.
- Corder, S.P. (1967) The significance of learner's errors, *IRAL*, 5(4), pp. 162-169.
- Corder, S.P. (1973) *Introducing Applied Linguistics*, Harmondsworth, Penguin, 400p.
- Corder, S.P. (1980) Que signifient les erreurs des apprenants ?, *Langages*, 15 (57), Paris, pp.9-15 (traduction d'un article de 1967).
- Couper-Kuhlen, E. (1993) *English speech rhythm: Form and function in everyday verbal interaction*, John Benjamins Publishing co, 360p.
- Cruttenden, A. (1986) *Intonation*, Coll. Cambridge Textbooks in Linguistics. Cambridge, Grande Bretagne : Cambridge University Press, 214p.

- Cruttenden, (2001) *Gimson's pronunciation of English*, sixth edition, Arnold, 360 p.
- Crystal, D. (1969) *Prosodic Systems and Intonation in English*, Cambridge University Press, London, 381p.
- Crystal, D. (2003) *English as a Global Language*, Cambridge University Press, 2nd Edition, 228p.
- Cucchiaroni, C. ; Strik, H. ; Boves, L. (2000) Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms, *Speech Communication* n° 30, pp.109-119.
- Cummins, F., Fort, R. (1998) Rhythmic constraints on English, in Walley, A. ; Flege, J., avril 1998, *Journal of Phonetics* 262, pp. 145-171.
- Cuq, J.P. (2003) *Dictionnaire de didactique du français langue seconde et langue étrangère*, Paris, France : Clé internationale, 303p.
- Dauer, R. (1983) Stress-timing and syllable-timing reanalyzed, *Journal of Phonetics*, 11 (1), pp. 51-62.
- Dauer, R (1987) Phonetic and phonological components of language rhythm, *Proceedings of the 11th International Congress of Phonetic Sciences*, 5, pp. 445-450.
- Davis, B.L. & Macneilage, P.F. (1995) The articulatory basis of babbling, *Journal of Phonetics*, 11, pp.51-62.
- De Bot, K. (1983) Visual feedback of intonation I: effectiveness and induced practice behavior, *Language and Speech*, 26 (4), pp.331-350.
- De Cordemoy, G. (1668) *Discours physique de la Parole*
http://fr.wikipedia.org/wiki/G%C3%A9raud_de_Cordemoy.
- De Saussure, F. (1968) *Cours de linguistique générale*, Collection études et Documents, Payot, Paris, 3^{ème} édition, 331p.
- Delais-Roussarie, E. (2002) *Constituer des corpus oraux : méthodes et outils. Carnets de grammaire*, Rapports interne de l'équipe de recherche syntaxe et sémantique, 10, Toulouse-le Mirail.
- Delattre, P. (1966) A Comparison of Syllable Length Conditioning among Languages. Off-print, *IRAL*, 4 (1), sept. 1966, pp. 183-198.
- Dell, F. (1984) L'accentuation dans les phrases en français, in *Formes sonores du langage*, F. Dell ; D.J. Hirst ; Vergnaud, J.R. (eds.), Hermann, pp. 65-122.

- Dellwo, V. (2006) Rhythm and Speech rate: A variation coefficient for ΔC , *Language and Language-Processing*, pp. 231-241.
- Dellwo, V. (2008) The role of speech rate in perceiving speech rhythm, In *Speech Prosody* 4:08, pp. 375-378.
- Delmonte, R. (1998) Prosodic modeling for automatic language tutors, In *STiLL-1998*, pp. 57-60.
- Delorme, A. ; Fluckiger, M. (2003) *Perception et réalité: une introduction à la psychologie des perceptions*, De Boeck, 516p.
- 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. (1975) *Présentation d'un test de niveau destiné à évaluer la prononciation des anglophones*, Université de Provence, R.P.A., 33-34, pp.9-35.
- Di Cristo, A. ; Hirst, D. (1993) Rythme syllabique, rythme mélodique, et représentation hiérarchique de la prosodie du français (Prosodie), *Institut de Phonétique d'Aix et UA 261 du CNRS*. Aix-en-Provence, France: 1993-94, pp. 9-24.
- Di Cristo, A. ; Hirst, D. (1997) Un procédé d'alignement automatique de transcriptions phonétiques sans apprentissage préalable, *4e congrès d'Acoustique*, pp. 14-18.
- Di Cristo, A. ; Hirst, D. (1998) *Intonation Systems. A Survey of Twenty Languages*. Cambridge, Grande Bretagne : Cambridge University Press, 1998, 487p.
- Di Cristo, A. (2000) Interpréter la prosodie, Actes des 23ème *Journées d'Etude sur la Parole*, Aussois, 19-23 juin, pp. 13-23.
- Di Cristo, A. (2002) De la métrique et du rythme de la parole ordinaire : l'exemple du français, Bordas, E. (ed.): *le rythme de la prose*, Revue de sémiolinguistique des textes et discours, 16, pp. 25-43.
- Di Cristo, A. (2004) La prosodie au carrefour de la phonétique, de la phonologie et de l'articulation formes-fonctions, *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, 23, pp. 67-211.
- Dickerson, W.B. (1989) *Stress in the Speech Stream : the Rhythm of Spoken English*, Urbana and Chicago, University of Illinois Press.
- Dodane, C. (1999) Apprentissage précoce de langue étrangère : interférences de traitement entre langue et musique, 3^e congrès international du rythme : le

rôle du rythme pour le développement humain, Institut Jacques-Dalcroze, Genève.

- Dodane, C. ; Konopczynski, G. (2001) *Utilisation des technologies multimédias dans la correction prosodique de l'adulte (L2)*, communication orale, Ecole Doctorale Européenne.
- Dubois et al. (1973) *Dictionnaire de linguistique*, Paris, France : Larousse, 516p.
- Durand, J. ; Delais-Roussarie, E. (2003) *Corpus et variation en phonologie du français*, PUM, 372p.
- Elbers, L. (1982) Operating principles in repetitive babbling: A cognitive continuity approach, *Cognition*, 12, pp.45–63.
- Erikson, L.J. (1991) Development of the Filtered U Algorithm for Active Noise Control, *Journal of the Acoustical Society of America*, 89 (1), pp. 257-265.
- Eskenazi, N. (1999) Using a Computer in Foreign Language Pronunciation Training: What Advantage?, *CALICO Journal*, 16 (3) , pp.447-471.
- Fant, G. (1991) Units of temporal organization: stress groups versus syllables and words, *Proceedings of the XIIth ICPHS*, 1, université de Provence, Aix-en-Provence, pp.241-261.
- Faure, G. (1962) *Recherches sur les caractères et le rôle des éléments musicaux dans la prononciation anglaise. Essai de description phonologique*, Coll. Etudes anglaises, Didier, Paris, 380p.
- Faure, G. ; Di Cristo, A. (1977) *Le français par le dialogue*, Hachette, Collection outils, Paris, 80p.
- Faure, G. ; Hirst, D. ; Chafcouloff (1980) Rhythm In English: Isochronism, Pitch and Perceived Stress, In WAUGH, L. R. ; SCHOONEVELD, C.H. van (eds), *The Melody Of Language. Intonation And Prosody*, pp. 71-80.
- Felps, D. ; Bortfeld, H. ; Gutierrez-Osuna, R. (2009) Foreign accent conversion in computer-assisted pronunciation training, *Speech Communication*, 51, pp. 920-932.
- Fernandez-Cruz, R. (2000) *Analyse phonologique et phonétique du portugais brésilien parlé par les communautés noires d'Amazonie*, thèse de doctorat, Université de Provence.
- Ferragne, E. ; Pellegrino, F. (2004) Rhythm in Read British English : Interdialect Variability, *Proceedings of Interspeech 04 ICSLP*, Jeju Island, Korea, pp.1573-1576.

- Field, J. (2005) Intelligibility and the listener: the role of lexical stress, *TESOL Quarterly*, 39 (3), pp.399-423.
- Fitzpatrick, M. (2002) Theories of Child Language Acquisition, *Child Language Acquisition* [Online], [2004, February 4].
- Flege, J.E. (1995) Second-language speech learning: Findings and problems, in Strange, W. (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, Timonium, MD: York Press, pp. 233-273.
- Flege, J.E. ; Hillenbrand, J. (1984), Limits on phonetic accuracy in foreign language speech production (French & English), *Journal of the Acoustical Society of America*, 76 (3), pp. 708-721.
- Fraisse, P. (1956) Les structures rythmiques, *Studia Psychologica*, ERASMES, S.A, Paris-Bruxelles, 125p.
- Fraisse, P. (1974) *Psychologie Du Rythme*, Paris, France : PUF, Presses Universitaires de France, 244 p.
- Freland-Ricard, M. (1995) *Analyse multilingue des erreurs prosodiques chez l'apprenant étranger-essai de paramétrisation*, thèse de doctorat, Université de Provence.
- Fromkin, V. ; Rodman, R. (1998) *An introduction to Language*, Harcourt, sixth ed., 565p.
- Galazzi-Matasci, E. ; Pedoya, E. (1983) Et la pédagogie de la prononciation ?, *Le Français dans le Monde*, 180, pp. 39-44.
- Germain, A. ; Martin, P. (2000) Présentation d'un logiciel de visualisation pour l'apprentissage de l'oral en langue seconde, *Alsic*, 3 (1), pp.61-76.
- Gibbon, D. ; Moore, R. ; Winski, R. (1998) *Spoken Language System and Corpus Design*, Berlin: Mouton de Gruyter, 216p.
- Ginésy, M. (1995) *Mémento de phonétique anglaise*, Nathan Université, 288p.
- Ginestet, F. (2002) L'enseignement/apprentissage de l'intonation du français. Une approche de l'intonation implicative avec des apprenants chinois, *Travaux didactique du français langue étrangère*, 47, pp. 87-104.
- Goldman, J.P. ; Auchlin, A. ; Simon, A.C. ; Avanzi, M. (2007) Phonostylographe : un outil de description prosodique. Comparaison du style radiophonique et lu, *Nouveaux cahiers de linguistique française*, 28, pp. 219-237.
- Granger et al. (2009) *The LONGDALE Project Longitudinal Database of learner English*, <http://cecl.fltr.ucl.ac.bc/LONGDALE.html>

- Goodglass, H. ; Calderon, M. (1977) Parallel processing of verbal and musical stimuli in right and left hemispheres, *Neuropsychologia*, 15 (3), pp. 397-407.
- Grabe, E. ; Low, E.L. (2002) Durational variability in speech and the rhythm class hypothesis, In Gussenhoven, C. ; Warner, N. (éds.), *Laboratory Phonology 7*. Coll. Phonology & Phonetics, 4-1. Berlin, Allemagne : Mouton De Gruyter, pp. 515-546.
- Graddol, D. (2000) *The future of English? A guide to forecasting the popularity of the English language in the 21st century*, The British Council, The English company (UK) Ltd, 65p.
- Grenon, I. ; White, L. (2008) Acquiring Rhythm: a Comparison of L1 and L2 speakers of Canadian English and Japanese, In H. Chan, H. Jacob & E. Kipia (Eds.), *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Language Development*, Somerville, Massachusetts: Cascadia Press, pp.155-166.
- Grodzinsky, Y. (1990) *Theoretical perspectives on language deficits*, Cambridge Mass ; London: MIT press, 192p.
- Grosjean, F. (1983) How long is the sentence? Prediction and prosody in the on-line processing of language, *Linguistics*, 21, pp. 501-529.
- Grosjean, F. ; Collins, M. (1979) Breathing, Pausing, Reading, *Phonetica*, 36, pp. 98-114.
- Guaitella, I. (1991) *Rythme et parole: comparaison critique du rythme de la lecture oralisée et de la parole spontanée*, Thèse de doctorat, Université de Provence.
- Guberina, P. (1939) *Valeur logique et valeur stylistique des propositions complexes*, thèse d'état, université de la Sorbonne.
- Guberina, P. (1974) 'Structuration et dépassement des structures, *Actes du 3ème colloque international SGAV pour l'enseignement des langues*, Didier, pp. 41-55.
- Gut, U. ; Urua, E. ; Adouakou, S. ; Gibbon, D. (2001) Rhythm in West African tone languages: a study of Ibibio, Anyi and Ega, *Typology of African Prosodic Systems Workshop*, Bielfield University, Germany.
- Gut, U. (2003) Prosody in second language speech production: the role of the native language, *Fremdsprachen Lehren und Lernen*, 32, pp. 133-151.
- Habert, B. ; Nazarenko, A. ; Salem, A. (1997) *Les linguistiques de corpus*, Paris : Armand Colin, 240p.

- Hadley, G. (2002) Sensing the winds of change: an introduction to data-driven learning, *RELC journal*, 32 (2), pp. 301-319.
- Han, M.S. (1962) The feature of duration in Japanese, *Onsei no kenkyuu*, 10, pp. 65-80.
- Halliday, M. (1967) *Intonation and Grammar in British English*, Coll. Janua Linguarum, Series Practica, 48. La Haye, Pays-Bas : Mouton & Co., 62p.
- Handel, S. (1983) The contextual nature of rhythmic interpretation, in *Perception and psychophysics*, 34 (2), pp. 103-120.
- Handel, S. (1989) *Listening. An Introduction to the Perception of Auditory Events*. Coll. Bradford Books. Cambridge, MA, USA : The MIT Press, 1989, 597p.
- Hansen, T.K. (2006) Computer assisted pronunciation training: the four 'K's' of feedback, in 4th International conference on *multimedia and information and communication technologies in education*, Seville, Espagne, pp. 342-346.
- Hardison, D.M. (2004) Generalization of Computer-Assisted Prosody Training: Quantitative and Qualitative Findings, *Language Learning & Technology*, 8 (1), pp. 34-52.
- Herry, N. (1997) *Prediction and Transcription of Intonation Patterns in English*, mémoire de maîtrise, université de Provence.
- Herry, N. (2001) *Evaluation objective et subjective de la prosodie anglaise parlée par des français : Apport de l'enseignement assisté par ordinateur*, Thèse de doctorat, Université de Provence.
- Herry, N. ; Hirst, D. (2002) Subjective and objective evaluation of the prosody of English spoken by French speakers: the contribution of computer assisted learning, *Proceedings of Speech Prosody 1:02*, Aix-en-Provence.
- Hesling, I. (2000) *Le rôle de l'hémisphère cérébral droit dans la production de la perception de la parole*, Thèse de doctorat, UTM, <http://www.univ-tlse2.fr/rech/sout-theses/2000-01-21-ang-hesling.html>.
- Hincks, R. (2003) Speech technologies for pronunciation feedback and evaluation, *ReCALL* 15, Cambridge University Press, pp. 3-20.
- Hincks, R. ; Edlund, J. (2009) Promoting increased variation with transient visual feedback, KTH Royal Institute of Technology, *Language Learning and Technology*, 13 (3), pp. 32-50.
- Hirst, D. J. (1977) *Intonative Features. A syntactic Approach to English Intonation*, Collection Janua Linguarum, Series Minor, Mouton, La Haye, Pays Bas, 139p.

- Hirst, D. & Di Cristo, A. (ed.). (1998). Intonation in British English, in Hirst, D.J. ; Di Cristo, A. (eds), *Intonation Systems: A survey of Twenty Languages*, Cambridge: Cambridge University Press, pp. 56-77.
- Hirst, D. (2000) The Prosody paradox: Universal and language-specific parameters of speech prosody, In proceedings of Workshop prosody 2000: *Speech Recognition and Synthesis*, (ESCA/ISCA), (Kraków, Poland).
- Hirst, D. (2009) The rhythm of text and the rhythm of utterances: from metrics to models, in Proceedings of *Interspeech*, Annual Conference of the International Speech Communication Association, (10: 2009 septembre 6-10 : Brighton).
- Hockett, C.F. (1958) *A Course in Modern Linguistics*. New York, NY, USA: MacMillan & Co. Ltd., 621p.
- Hoequist, C.J. (1983) Durational correlates of linguistic rhythm categories, *Phonetica*, 40, pp. 19-31.
- Horgues, C. (2005) Contribution à l'étude de l'accent français en anglais. Quelques caractéristiques prosodiques de l'anglais parlé par des apprenants francophones et leur évaluation perceptive par des juges natifs, Actes des VIIIèmes RJC ED268 '*Langage et langues*', Paris III, 21 mai 2005.
- Imoto, K. ; Tsubota, Y. ; Raux, A. ; Kawahara, T. ; Dantsuji, M. (2002) Modeling and Automatic Detection of English Sentence Stress for Computer-Assisted English Prosody Learning System, 7th International Conference on *Spoken Language Processing-ICSLP*, Denver, Colorado, USA.
- Jakobson, R. (1941) *Kindersprache, Aphasie und allgemeine Lautgesetze*, Språkvetenskapliga Sällskapet I Uppsala Förhandlingar, Uppsala: Uppsala Universitetet. English translation: *Child Language, Aphasia and Phonological Universals*, The Hague & Paris: Mouton, 1968.
- James, E. (1972) *L'acquisition des faits prosodiques à l'aide d'un visualiseur de mélodie*, Thèse de 3^{ème} cycle, Université de Provence.
- James, E. (1976) The Acquisition of Prosodic Features of Speech Using a Speech Visualizer, *International Review of Applied Linguistics in Language Teaching*, 14 (3), pp. 227-243.
- James, E. (1977) The Acquisition of a Second-Language intonation Using a Visualizer, *Canadian Modern Language Review*, 33 (4), pp. 503-506.
- Jassem, W. (1952) Stress in Modern English, *Bulletin de la société polonaise de linguistique*, fascicule XI, Poland / Croatie.

- Jilka, M. (2000) The Contribution of Intonation to the Perception of Foreign Accent, Doctoral Dissertation, Arbeiten des Instituts für Maschinelle Sprachverarbeitung (AIMS), 6 (3), University of Stuttgart.
- Jun, S.A. (2005) *Prosodic Typology: The Phonology of Intonation and Phrasing*, New York: Oxford University Press, 466p.
- Kacmarek, L. (1953) *Kształtowanie Sie Nowy Dziecka*, Poznan, towarzystwo Przyjaciół Nauk.
- Kent, R.D. (1976) Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies, *Journal of speech and hearing Research*, 19, pp. 421-444.
- Kent, R.D. ; Mitchell, P.R. ; Sancier, M. (1991) Evidence and Role of Rhythmic Organization in Early Vocal Development in Human Infants, in *The Development of Timing Control and Temporal Organization in Coordinated Action*, J.Fagard & P.H. Wolff editors, 81, pp.135-149.
- Klein, W. (1989) *L'acquisition de langue étrangère*, Paris, Armand Colin, 251p.
- Kochanski, G. ; Grabe, E. ; Coleman, J. (2004) The difference between a question and a statement: a survey of English dialects. *The Journal of the Acoustical Society of America* 115(5), pp. 23-98
- Kohler, K.J. (1982) F0 In The Production Of Lenis And Fortis Plosives, Off-print, *Phonetica*, 39, pp. 199-218.
- Kohler, K.J. (1991) The Phonetics Phonology Issue in the Study of Articulatory Reduction, *Phonetica*, pp. 180-192.
- Kommissarchik, J. ; Kommissarchik, E. (2000) Better Accent Tutor, Analysis and visualization of speech prosody, *Proceedings of InSTIL 2000*, Dundee, Scotland, pp. 86-89.
- Konopczynski, G. (1986) *Du prélangage au langage : acquisition de la structure prosodique*, Thèse d'Etat, Université de Strasbourg.
- Konopczynski, G. (1979) Le statut de la prosodie dans les recherches sur l'acquisition du langage : Réflexions sur les corpus, in Léon, P. ; Rossi, M., *Problèmes De Prosodie*, vol. 1: Approches Théoriques, 1979, pp. 57-70.
- Konopczynski, G. (1990) *Le langage émergent : caractéristiques rythmiques*, Hamburg :Buske (Beitrag zur Phonetik und Linguistik ; Bd.60).
- Konopczynski, G. (1999) *L'acquisition du système prosodique de la langue maternelle et ses implications pour l'apprentissage d'une L2*. in Delcloque, Philippe, p 62-70.

- Krashen, S. (1981) *Second Language acquisition and learning*, Oxford, Pergamon Press, 150 p., téléchargeable à partir de:
http://www.sdkrashen.com/SL_Acquisition_and_Learning/index.html.
- Kuhl, P.K. (1998) Effects of language experience on speech perception, *Journal of Acoustical Society of America*, 103, pp. 29-31.
- Kuhl, P.K. (2000) A new view of language acquisition, *Proceedings of the National Academy of Science*, 97, pp.11850-11857.
- Ladefoged, P. (1975) *A Course in Phonetics*. New York, NY, USA : Harcourt Brace Jovanovich, Inc., 296p.
- Landercy, A. ; Renard, R. (1977) *Eléments de Phonétique*, Didier, Bruxelles, 214p.
- Lane, H. ; Buiten, R. (1965) A preliminary manual for the speech auto-instructional device, Behavior Analysis Laboratory, *Progress Report* n°5, nov. 1, University of Michigan.
- LAVAC (version 4.03.i-1999). LAVAC (Laboratoire Vidéo Actif Comparatif) a été conçu par Tony Toma. Éditeur C3 : Montpellier. Consulté en novembre 1999 : <http://www.alizes.fr/cp3i>.
- Lecanuet, J. P. (1995) *L'audition auditive prénatale*, In : I. Deliège & J. A. Sloboda Naissance et développement du sens musical, Paris : PUF.
- Lado, R. (1957) *Linguistics across Cultures*, Ann Arbor, MI: University of Michigan Press, 141p.
- Laurendeau, J. (1990) *Maurice Martenot, luthier de l'électronique*, Croissy-Beaubourg, Dervy-Livres.
- Le Robert (2000) *Dictionnaire historique de la langue française*, Maury-eurolivres, Manchecourt, p.28.
- Lehiste, I. (1977) Isochrony reconsidered, *Journal of Phonetics*, 5 (3), pp. 253-263.
- Léon P. & Martin P. (1971) Linguistique appliquée et enseignement de l'intonation, *Etudes de Linguistique appliquée*, n°3, p. 36-45.
- Leopold, W.F. (1948) The Development of Descriptive Linguistics in America 1925-1956, German Ch., In JOOS, Martin (ed), *Reading in Linguistics*, vol. 1, pp. 215-216.
- Lepetit, D. (1992) *Intonation française: enseignement et apprentissage*, coll.*Etudes Phonétiques*, Studies in Phonetics, 1, Toronto, Canada: Canadian Scholars' Press, 113p.

- Lester, B. (1985) Introduction: there's more to crying than meets the ear, In *Infant Crying: Theoretical and Research perspectives*, New York, B. Lester & C.F.Z. Boukydis editors, Plenum Press.
- Lewis, M.M. (1936) *Infant Speech: a study of the beginnings of language*, the International Journal of Psychoanalysis, 18, pp. 324-325.
- Liberman, A.M. ; Cooper, F.S ; Shankweiler, D.P. ; Studdert-Kennedy, Y.M. (1967) Perception of Speech code, *Psychological Review*, 74 (6), pp. 13-50.
- Liberman, M. ; Prince, A. (1977) On Stress and Linguistic Rhythm, *Linguistic Inquiry* n°8, pp. 249–336.
- Lin, H. ; Wang, Q. (2005) Vowel quantity and consonant variance: a comparison between Chinese and English, In proceedings of *Between Stress and Tone*, Leiden, June 2005.
- Lin, H. & Wang, Q. (2007) Mandarin Rhythm: An Acoustic Study, *Journal of Chinese Language & Computing*, 17 (3), pp. 127-140.
- Llorca, R. (1998) Relations rythmiques entre geste et parole: Application à la pédagogie et au spectacle, in Guiatella *et al.* eds, *Orage 98*, Paris, L'Harmattan.
- Lloyd James, A. (1940) *Speech Signals in Telephony*, London
- Lopez, G. ; Prodeau, M. ; Herry-Benit, N. ; Tortel, A. « From assessing levels to analysing the productions », (à paraître)
- Loukina, A. ; Koschanski, G. ; Shih, C. ; Keane, E. ; Watson, I. (2009) Rhythm measures with language independent segmentation, *Proceedings of Interspeech conference 2009*, Brighton, England.
- Low, E. ; Grabe, E. ; Nolan, F. (2000) Quantitative characterizations of speech rhythm: 'Syllable-timing' in Singapore English, *Language and Speech*, vol. 43, pp. 377-401.
- Luangthongkum, T. (1977) *Rhythm in Thai*, doctoral dissertation, University of Edinburgh.
- Lyster, R. (2001) Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms, *Language Learning* 51 (1), pp.265-301.
- Mackey, W.F. (1965) *Language Teaching Analysis*, Longman, London, 554p.
- Magen, H.S. (1998) The Perception of Foreign Accented Speech, In WALLEY, Amanda C., FLEGE, James E., *Journal of Phonetics*, pp.381-400.

- Mak B., Siu M. ; Ng M. ; Tam Y. ; Chan Y. ; Chan K. ; Leung K. ; Ho S. ; Chong F. ; Wong J. ; Lo J. (2003) PLASER: pronunciation learning via automatic speech recognition, *Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing* - Tome 1, pp. 23 -29.
- Major, R.C. (2001) *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*, Mahwah (NJ.): Lawrence Erlbaum Associates, 211p.
- Martin, P. (2005) Winpitch LTL, un logiciel multimédia d'enseignement de la prosodie, *Alsic*, 8 (2), pp.95-108.
- Marieb, W.F. (1993) *Anatomie et physiologie humaine*, De Boeck-Wesmael S.A., seconde édition, pp.382-389.
- Markus, D. ; Bond, D. ; Stockmal, V. (2003) Measures of Native and non-native Rhythm in a Quantity Sensitive Language, University of Latvia, Riga ; Université d'Ohio, pp.1-11.
- Marty, F. (1983) Les enseignants de langues face à l'Enseignement Assisté par Ordinateur, *Le Journal de la Formation Continue et de l'EAO*, 169, 10/12/1983.
- Mauranen, A. (2004) Speech Corpora in the classroom, in G.Aston *et al.* (eds.), *Corpora and Language Learners*, Amsterdam, John Benjamins, pp. 95-211.
- Mehler, J. ;Juszyk,P. ; Lambertz, G. ; Halstead, N. ;Bertoncini, J. ;Amiel-Tison, C. (1988) A precursor of language acquisition in young infants, *Cognition*, vol. 29, pp. 143-178, dans Hallé, P ; (1999) Acquisition du langage: spécialisation des enfants dans leur langue maternelle, 3^e Congrès International du Rythme : *le rôle du rythme pour le développement humain*, Institut Jacques-dalcroze, Genève.
- Mennen, I. (2006) Phonetic and phonological influences in non-native intonation: an overview for language teachers, *Working Papers Queen Margaret University College* 9.
- Meumann, E. (1894) Untersuchungen zur Psychologie u. Aesthetik d. Rhythmus, *Philos.Stud.*, vol. 10, pp. 249-322.
- Miller, J.L. Grosjean, F. & Lomanto, C. (1984) Articulation rate and its variability in spontaneous speech: a reanalysis and some implications, *Phonetica*, 41 (4), p. 215
- Millot, J.L. (1985) *Contribution à l'étude des interactions entre le nouveau-né et son environnement social. Aspects somesthésiques et acoustiques*, Thèse d'Etat, Université de Besançon.

- Minematsu, N. ; Fujisawa, Y. ; Nakagawa, S. (1999) Automatic detection of stressed syllables in English words using HMM and its application to prosodic evaluation of pronunciation proficiency, *Trans. IEICE*, J82-D-II (11), pp.1865-1876.
- Montredon J., *et al.*, (1975) *C'est le printemps 1*, Paris, Clé international.
- Munro, M.J. ; Derwing, T.M. (1995) «Foreign-Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners». *Language Learning*, vol. 45 (2), pp. 73-97.
- Munro, M.J. ; Derwing, T. (1998) Evidence in Favor of a Broad Framework for Pronunciation Instruction, *Language Learning*, 48 (2), pp. 393-410.
- Munro, M.J. ; Derwing, T.M. (2001) Modeling perceptions of the accentedness and comprehensibility of L2 speech: the role of speaking rate, *Studies in Second Language Acquisition*, 23, pp.451-468.
- Murray, G.L. (1999) Autonomy and language learning in a stimulated environment, *System* 27(3), pp.295-308.
- Myers, G. & Myers, M. (1984) *Les Bases De La Communication Interpersonnelle. Une Approche Théorique Et Pratique*, Montréal, Canada : McGraw-Hill, 470 p. (Original : "The Dynamics of Human Communication". Traduit par P. Racine).
- Nava, E. ; Goldstein, L. ; Saltzman, E. ; Nam, H. ; Proctor, M. (2008) Modeling prosodic rhythm: evidence from L2 speech, *Journal of Acoustical Society of America*, 124 (4), pp. 2577-2577.
- Nazzi, T. ; Jusczyk, P.W. ; Johnson, E.K. (2000) Language discrimination by English learning 5-month-olds: effect of rhythm and familiarity, *Journal of Memory and Language*, n°43, pp.1-19.
- Neri, A. ; Cucchiaroni, C. ; Strik, H. ; Boves, L. (2002) The pedagogy-technology interface in computer assisted pronunciation training, *Computer Assisted Language Learning*, 15 (5), pp. 441-467.
- Nespor, I. (1990) On the rhythm parameter in phonology, In *Logical issues in Language Acquisition* (I.Roca, editor), Dordrecht: Foris, pp. 157-175.
- Neumeyer, L. ; Franco, H. ; Abrash, V. ; Julia, L.. Ronen, O. ; Bratt, H. ; Bing, J. ; Digalakis, V. ; Rypa, M. (1998): "Webgrader(TM): a multilingual pronunciation practice tool", In *STiLL Speech Technology in Language Learning*, pp. 61-64.
- Nicaise, A. ; Grey, M. (1998) *L'intonation de l'anglais*, Paris : Nathan université, 128p.

- Nikolov, R., Dommergues, J.-Y. et Ryst, E. (2007) SaRP : un outil de représentations graphiques multi-points et multi-séries des formants vocaliques, *Scientific Works*, vol. 45, Book 1, Philology, Plovdiv University "Païssii Hilendarski", Bulgaria.
- Norris et al. (1995) Competition and segmentation in spoken word recognition, *Journal of Experimental Psychology : Learning, Memory and Cognition*, 21, pp. 1209-1228.
- O'Connor, J.D., Tooley, O. (1961) *The Perceptibility Of Certain Word Boundaries*, In ABERCROMBIE, David ; FRY, D.B. ; MACCARTHY, P.A.D. ; SCOTT, N.C. et al. (eds), In Honour Of Daniel Jones. Papers Contributed On The Occasion Of His Eightieth Birthday: 12 September 1961, pp. 171-176.
- Oller, D.K. (1978) Infant vocalization and the development of speech, *Allied Health and Behavioral Sciences* 1, pp.523-549.
- Pettito, L.A. & Marantette, P.F. (1991) Babbling in the Manual Mode: Evidence for the Ontogeny of Language, *Science* n° 251, pp. 1493-1496.
- Piaget, J. (1950) *Introduction à l'épistémologie génétique*, Payot, Paris.
- Pike, K. (1945) *The Intonation of American English*. Ann Arbor, MI, USA: The University of Michigan Press, 203p.
- Platon, *Les Lois*, livre II, ed. Flammarion, 2006, traduction de L. Brisson & JF. Pradeau, pp. 114-159.
- Platt, J.T. & Weber, H. (1980) *English in Singapore and Malaysia: Status, Features and Functions*. Kuala Lumpur: Oxford University Press.
- Poquin, M. (2009) *Phonological issues in the production of prosody by francophone and sinophone learners of English as a second language*, Thèse de Doctorat, Université de Québec, Montréal
- Pritchard, R. (1985) The Teaching of French Intonation to Native Speakers of English, *International Review of Applied Linguistics*, vol. 23 (2), p. 117-147.
- Pulgram, E. (1970) *Syllable, word, nexus, cursus*, La Haye, Pays-Bas, Mouton & Co., 174p.
- Querleu, O.D., Renard, X. (1981) Perception auditive et récativité fœtale aux stimulations sonores, *Journal Gynecol, Obstet, Biol. Repr.* n° 10, pp. 307-314.
- Ramus, F. (1997) Le Rôle du Rythme pour la Discrimination des Langues. *Actes des JIOSC 97*, pp 225-229.

- Ramus, F. ; Mehler, J. (1999) Language identification with suprasegmental cues: a study based on speech synthesis, *Journal of the Acoustic Society of America*, 105 (1), pp. 512-521.
- Rasier, L. ; Hiligsmann, P. (2007) Prosodic transfer from L1 to L2. Theoretical and methodological issues, *Nouveaux cahiers de linguistique française* 28, pp.41-66.
- Revis, J. (2004) *L'analyse perceptive des dysphonies : approche phonétique de l'évaluation vocale*, thèse de doctorat, Université de la Méditerranée, Marseille.
- Richards, J.C. (1971) A Non-Contrastive Approach to Error Analysis, *English Language Teaching* 25, pp. 204-219.
- Roach, P. (1982) On the Distinction between "Stress Timed" and "syllable Timed" Languages. In CRYSTAL, David (ed.), *Linguistic Controversies. Essays In Linguistic Theory And Practice*, In Honour Of F.r. Palmer, 1982, pp. 73-79
- Roach, P. (1991) *English Phonetics and Phonology*, Cambridge University Press, 262p.
- Rondal, J.A. (1983) *L'interaction adulte-enfant et la construction du langage*, Coll. Psychologie et Sciences Humaines. Bruxelles, Belgique : Mardaga, 197p. (Dédié à Art Staats).
- Rost, M. (2002) *Teaching and Researching Listening*, Harlow:Pearson, 328p.
- Russo, M. ; Barry, J. (2008) Isochrony reconsidered. Objectifying relations between rhythm measures and speech tempo, In: *Proceedings of Speech Prosody 4*: 08, pp. 419-422.
- Santacrose, M. (2002) Informatique, numérique et multimédia : quelques mythes récurrents en didactique des langues, *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)*, 21, pp. 121-151.
- Santerre, L. (1989) *Peut-on juger de la production par la perception? (ou faut-il en croire ses oreilles?)*, Mélanges de phonétique générale et expérimentale offerts à Péla Simon, vol. 2, Publications de l'Institut de phonétique de Strasbourg, pp. 735-755.
- Schmidt, E.N. (1939) *Über den Aufbau rhythmischer Gestalten*, N. *Psychology studies*, n°14, in Fraisse (1956).
- Selinker, L. (1972) Interlangue, *IRAL*, n°10, pp. 209-231.

- Selkirk, E. (1984) On the Major Class Features and Syllable Theory. In *LIEBER, Rochelle, Consonant Gradation In Fula: An Autosegmental Approach*, pp. 107-136.
- Smith, D.M. (1973) Creolization and language ontogeny: a preliminary paradigm for comparing language socialization and language acculturation, *New ways of analyzing variation in British English*, Washington, Georgetown University.
- Smith, A. (1976) The timing of French, with reflections on syllable timing, *Work in Progress* 9, Edinburgh University Department of Linguistics, pp.97-108.
- Stampe, D. (1969) The acoustic of the phonetic representation, in *Papers from the Fifth Regional Meeting of the Chicago Linguistics Society*, Chicago, IL: Chicago Linguistics Society, pp. 433-444.
- Steele, J. (1779) *Prosodia Rationalis, or An Essay towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols*, 2nd edition amended and enlarged. London: J. Nichols [Kessinger Publishing, 2008], 268p.
- Stenton, A. ; Péchou, A. ; Vaillant-Sirdey, C. ; Tricot, A. (2005) *Effet du double codage synchrone de l'accentuation en L2 selon les modalités de restitution du sujet*, communication au 1^{er} colloque international de didactique cognitive, Toulouse, 26-28 janvier 2005.
- Stockmal, V. ; Markus, D. ; Bond, D. (2005) Measures of Native and Non-Native Rhythm in a Quantity Language, *Language and Speech*, 48 (1), pp. 55-63.
- Taniguchi, M. ; Abberton, E. (1999) Effect of interactive visual feedback on the improvement of English intonation of Japanese EFL learners, *Speech, Hearing and Language: Work in progress*, n°11, pp. 76-89.
- Tanner, M.W. ; Landon, M.M. (2009) The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation and overall comprehensibility, Brigham Young University, *Language Learning and Technology*, 13 (3), pp. 51-65.
- Tarone, E. ; Frauenfelder, U. ; Selinker, L. (1978), *The Phonology of Interlanguage*, in Richards, J.C., ed., pp. 15-33.
- Teixiera, C. ; Franco, H. ; Shriberg, E. ; Precoda, K. ; Sonmez, K. (2000) Prosodic Features for Automatic Text-Independent Evaluation of Degree of Nateness for Language Learners, *ICSLP-6th International Conference on Spoken Language Processing*, Beijing, China, October 2000.
- Thelen, E. (1979) Rhythmical behavior in infancy: an ethological perspective, *Development Psychology*, n°17, pp. 237-257.

- Tongue, R.K. (1979) *The English of Singapore and Malaysia*, Singapore: Eastern Universities Press.
- Toivanen, J. (2003) Tone choice in the English intonation of proficient non-native speakers, Media Team, Unmea University, *PHONUM* 9, pp.195-168.
- Tomatis, A. (1971) *Education et dyslexie*, Paris, France : Editions ESF, 201p.
- Tortel A. (2004) Evaluation of intonation and rhythm of French and Native speakers, mémoire de master I, Université de Provence.
- Tortel, A. (en cours) DIALANG: évaluation de la proposition de diagnostic
- Trofinovitch, P. ; Baker, W. ; Mack, M. (2001) Context -and Experienced- Based Effects on the Learning of Vowels in a Second Language, *Studies in the Linguistic Sciences*, 3 (2), pp. 167-186.
- Troubetzkoy, N. S. (1949) *Principes de Phonologie*, Klincksieck, Paris, 396p.
- Vaissière, J. (1989) On automatic extraction of prosodic information for automatic speech recognition system, *Eurospeech* 98, European Conference on Speech communication and technology, Paris, vol. 1, pp.202-205.
- Van Santen, J.P.H. ; Tucker Prud'hommeaux, E. ; Black, L.M. (2009) Automated assessment of prosody production, *Speech Communication*, 51 (11), pp.1082-1097.
- Vardanian, R.M. (1964) Teaching English through oscilloscope displays, *Language Learning*, 3/4, pp. 109-117.
- Wagner, P. ; Dellwo, V. (2004) Introducing YARD (Yet Another Rhythm Determination) And Re-introducing Isochrony to Rhythm Research, In: *Proceedings of Speech Prosody* 2:04, Nara, Japan.
- Wang, H. ; Waple, C.J. ; Kawahara, T. (2009) Computer Assisted Language system based on dynamic question generation and error prediction for automatic speech recognition, *Speech Communication* 51 (10), pp.995-1005.
- Weber, G. (1997). The World's 10 most influential Languages, *Language Today*, 3, pp.12-18.
- Weinreich, U. (1953) *Languages in contact. Findings and problems*, New York, La Haye, Pays Bas: Mouton & Co., 148p.
- Wenk, B. J. (1985) Speech rhythm and second language acquisition, *Language and Speech*, vol. 28 (2), pp. 157-175.

- Wenk, B.J. ; Wioland (1982) Speech Rhythms In Second Language Acquisition, in Tjaden, K. , avril-juin 1985, pp. 157-176.
- White, L. ; Mattys, S.L. (2007) Calibrating rhythm: First language and second language studies, *Journal of Phonetics*, 35(4), pp 501-522.
- Whitworth, N. (2002) Speech Rhythm Production in Three German-English Bilingual Families, Leeds Working Papers in *Linguistics & Phonetics*, D. Nelson editors, n° 9, pp.176-205.
- Williams, R. and Herrup, K. (2001) The Control of Neuron Number, *The Annual Review of Neuroscience*, n° 11, pp. 423–453.
- Wolff, P. (1969) The natural history of crying and other vocalizations in early infancy, In *Determinants of Infant Behaviour*, vol. 4, B.M. Foss editors, London: Methuen.

Ouvrages ou articles consultés

- Astolfi, J.P. (2006) *L'erreur, un outil pour enseigner*, Paris, ESF, 117p.
- DeBoysson Bardies, B. (1996) *Comment la parole vient aux enfants, de la naissance jusqu'à deux ans*, (ed) Odile Jacob, Paris, France, 289p.
- Faure, G. (1975) *Manuel Pratique d'anglais parlé*, Hachette, 208p.
- Krashen, S. (1985) *The Input Hypothesis: Issues and Implications*, Longman, Harlow Essex, England, 120p.
- Mc Laughlin, B. (1987) *Theory of Second Language Learning*, Ed. Arnold, 184p.
- Malherbe, M. (1982) *Les langages de l'humanité*, Seghers, Paris, 444p.

RESUME

L'absence de système d'évaluation de la prosodie de l'anglais L2 parlé par des apprenants francophones a motivé la réalisation de ce travail de thèse. Dès lors, l'analyse de facteurs rythmiques dans la production d'apprenants français, qui fait l'objet de cette recherche, a un double objectif : analyser l'influence du rythme de la langue maternelle (L1=français) sur la langue cible (L2=anglais), et d'autre part, proposer des critères évaluatifs du rythme des productions des francophones, à partir de différentes métriques rythmiques proposées dans la littérature. Ces travaux ont mené à l'élaboration d'une base de données d'anglais L1/L2 nommée ANGLISH, à partir de laquelle nous avons montré une discrimination des locuteurs natifs/non natifs s'appuyant sur la combinaison de paramètres rythmiques. Les résultats de ces travaux de recherche montrent donc qu'il est possible de distinguer la L1 de la L2, et de donner une tendance rythmique des productions des apprenants francophones.

Mots-clefs: prosodie, rythme, métriques, anglais, L1 & L2, apprenants francophones, évaluation, critères rythmiques, corpus d'apprenants

TITLE: Qualitative evaluation of prosody of French learners: contribution to prosodic characterisation

SUMMARY

This thesis was motivated by the absence of systems of evaluation of the prosody of English spoken by French learners. The main aim of our research was to analyse rhythmic factors in the production of French learners with two objectives: first of all, to analyse the influence of the mother tongue's rhythm (L1=French) on the target language (L2=English), but also to provide evaluative criteria for French speakers based on various rhythmic metrics proposed in the literature. This led to the elaboration of a database of English L1 / L2, ANGLISH, which enabled us to distinguish between native and non-native speakers by using a combination of rhythmic parameters. The results of this research show that it is possible to distinguish between L1 and L2 rhythm, and to foresee the rhythmic tendencies in the productions of French-speaking learners.

Key-words: prosody, rhythm metrics, English, L1 & L2, French learners, evaluation, rhythmic criteria, learners corpus

UFR LACS

Formation doctorale: Cognition, Langage, Education

LABORATOIRE PAROLE ET LANGAGE (LPL)
UMR 6057 CNRS / Université de Provence

5, avenue Pasteur
13100 Aix-en-Provence, France